THE ROLE OF HIGH-ORDER CHROMATIN ORGANIZATION IN GENE REGULATION

EDITED BY: Alexey V. Pindyurin and Veniamin Fishman

frontiers Research Topics

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

THE ROLE OF HIGH-ORDER CHROMATIN ORGANIZATION IN GENE REGULATION

Topic Editors:
Alexey V. Pindyurin, Institute of Molecular and Cellular Biology (RAS), Russia
Veniamin Fishman, Institute of Cytology and Genetics, Russian Academy of Sciences (RAS), Russia

Table of Contents

Check for updates

Editorial: The role of high-order chromatin organization in gene regulation

Veniamin Fishman[1,2]* and Alexey V Pindyurin[3]*

[1]Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia, [2]Novosibirsk State University, Novosibirsk, Russia, [3]Institute of Molecular and Cellular Biology SB RAS, Novosibirsk, Russia

Editorial on the Research Topic
The role of high-order chromatin organization in gene regulation

It is now accepted that high-order chromatin packaging is non-random and essential for genome functioning. However, details of mechanisms underlying chromatin folding and its role in epigenetic regulation remain to be discovered. One of the most actively studied problems in this field is understanding the molecular basis of correlations between gene expression and chromatin architecture. Although causal relationships between these processes are still not well established, information about genome folding can already be employed for functional interpretation of genomic variants. For example, spatial proximity between gene promoters and regulatory elements can be instructive for linking GWAS variants with their putative target genes. An article by Thiecke et al. published in this Research Topic describes how information about genome folding was employed for prioritization of candidate genes underpinning COVID-19 host genetic traits. Similarly, chromatin architecture can help to understand cancer genomics. Studies by Adeel et al. and Baur et al. groups show how to infer connections between genome architecture and disease susceptibility by profiling chromatin contacts in multiple cancer samples. This approach allows authors to identify chromatin alterations specific for cervical and breast cancer development.

Several papers published in this Research Topic allow more comprehensive review of known mechanisms and new hypothesis explaining connections between changes of chromatin architecture. The paper by Boltsis et al. published in this article Research Topic, describes general principles connecting genome architecture with development and disease. Another review by Daghsni and Aldiri discusses this problem in the context of mammalian retina development, and reviews Tchurikov and Kravatsky and Malashicheva and Perepelina highlights the important role of ribosomal genes clusters and lamin proteins in global epigenetic regulation. Finally, a review by Kumar et al. extends the subject to the area of plant genomics.

Articles mentioned above show the importance of studying chromatin architecture in cancer and other human tissues. However, in many experimental systems, application of conventional chromatin profiling techniques is challenging due to a limited amount and/or peculiar properties of the available biological material. Several articles in this Research Topic describe new methodologies aiming to overcome this limitation. Research by Animesh et al. shows how to apply Hi-C method to nasopharyngeal cancer needle biopsy patient samples. In the article by Bylino et al. perform critical analysis of the chromosome conformation capture procedure to provide a useful guide to the 3C procedure. Another research from the same group Bylino et al. describes application of this updated 3C method to study the developmental genes in *Drosophila* larvae. Finally, Aljogol et al. compare computational methods of capture Hi-C-data analysis and highlight their advantages and disadvantages.

Whereas some of the aforementioned methods aim to modify the original Hi-C protocol to make it suitable for low-input samples, a very important direction of research is single-cell analysis of genome folding. Complementary to these new molecular methods, recently developed high-resolution microscopy approaches extend our understanding of genome architecture in individual cells. Cardozo Gizzi discusses how new molecular technologies developed in the rapidly evolving single-cell genomics field change our view of chromatin architecture. And two articles Maslova and Krasikova and Zakirov et al., discuss how the structures revealed by molecular and modern microscopy analysis correspond to each other.

Overall, this article Research Topic shows how using new computational and molecular tools can extend our understanding of mechanisms underlying chromatin folding and transcription dynamics, and how this knowledge can be used in solving problems in fundamental biology, agriculture, and molecular medicine.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Structural Variations of the 3D Genome Architecture in Cervical Cancer Development

Muhammad Muzammal Adeel[1,2], Hao Jiang[2], Yibeltal Arega[2], Kai Cao[2], Da Lin[3,4,5], Canhui Cao[6,7], Gang Cao[3,4,5], Peng Wu[6,7]* and Guoliang Li[1,2]*

[1] National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, China, [2] Agricultural Bioinformatics Key Laboratory of Hubei Province, Hubei Engineering Technology Research Center of Agricultural Big Data, 3D Genomics Research Center, College of Informatics, Huazhong Agricultural University, Wuhan, China, [3] State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan, China, [4] College of Veterinary Medicine, Huazhong Agricultural University, Wuhan, China, [5] College of Bio-Medicine and Health, Huazhong Agricultural University, Wuhan, China, [6] Department of Gynecologic Oncology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [7] Cancer Biology Research Center (Key Laboratory of the Ministry of Education), Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Human papillomavirus (HPV) integration is the major contributor to cervical cancer (CC) development by inducing structural variations (SVs) in the human genome. SVs are directly associated with the three-dimensional (3D) genome structure leading to cancer development. The detection of SVs is not a trivial task, and several genome-wide techniques have greatly helped in the identification of SVs in the cancerous genome. However, in cervical cancer, precise prediction of SVs mainly translocations and their effects on 3D-genome and gene expression still need to be explored. Here, we have used high-throughput chromosome conformation capture (Hi-C) data of cervical cancer to detect the SVs, especially the translocations, and validated it through whole-genome sequencing (WGS) data. We found that the cervical cancer 3D-genome architecture rearranges itself as compared to that in the normal tissue, and 24% of the total genome switches their A/B compartments. Moreover, translocation detection from Hi-C data showed the presence of high-resolution $t(4;7)$ (q13.1; q31.32) and $t(1;16)$ (q21.2; q22.1) translocations, which disrupted the expression of the genes located at and nearby positions. Enrichment analysis suggested that the disrupted genes were mainly involved in controlling cervical cancer-related pathways. In summary, we detect the novel SVs through Hi-C data and unfold the association among genome-reorganization, translocations, and gene expression regulation. The results help understand the underlying pathogenicity mechanism of SVs in cervical cancer development and identify the targeted therapeutics against cervical cancer.

Keywords: cervical cancer, gene expression, Hi-C, SVs, translocation detection, topologically associating domains

INTRODUCTION

Cervical cancer (CC) is the fourth most common cancer affecting women worldwide. With an estimated 570,000 cases and 311,000 deaths in 2018, this disease accounts for 3.3% of all cancer-related deaths (Bray et al., 2018), and there is a wide variation in incidence and mortality in various regions. In general, cancer is characterized by uncontrolled growth and cell proliferation

due to several genomic changes such as gene mutations, insertion/deletions, and chromosomal rearrangements (Engreitz et al., 2012). In China, oncogenic HPV infection in women has been reported as 5–20%, depending on location and age (Münger et al., 2004; Dai et al., 2006; Zhao et al., 2012). Several studies have suggested that human papillomavirus (HPV) is the leading cause of cervical cancer and HPV genome integration is the key mechanism. Previously reported studies had suggested that the HPV integration hotspots, molecular pathogenesis, the role of episomal HPV E6/E7 expression, and HPV integration in human genome 3D structure (Fudenberg et al., 2011; Koneva et al., 2018; Cao et al., 2020) play a vital role in cervical cancer development (Garraway and Lander, 2013).

Structural variations (SVs) such as deletions, duplications, insertion, inversions, and translocations are majorly associated with disease development. Chromosome conformation capture techniques such as Hi-C and ChIA-PET have revealed that SVs alter the three-dimensional (3D) genome and gene regulations in the cancer genome (Dixon et al., 2018). SVs, specifically translocations that occur at specific hotspots in the genome, cause a significant impact on the 3D structure and gene expression (Lawrence et al., 2013). The detection of SVs and their effects on chromosomal architecture and gene expression has significantly increased our understandings of tumor development (Dixon et al., 2018). Multiple conventional techniques such as Microarray (Alkan et al., 2011), fluorescence *in situ* hybridization (FISH) (Cui et al., 2016), and PCR are already available to identify SVs (Sanchis-Juan et al., 2018). However, these methods have some drawbacks because they required prior knowledge (Mardis and Wilson, 2009); most of the techniques cannot accurately locate the sequence of breakpoints, making it more challenging to monitor the impact of specific SVs on gene structure (Hu et al., 2020). Nowadays, several studies have been designed to apply the most advanced high-throughput techniques such as whole-genome sequencing (WGS), RNA-seq, and chromosome conformation capture (Hi-C) data to study the SVs effectively (Dixon et al., 2018).

Despite the massive ongoing progress in cancer studies, there is still plenty of room to devise comprehensive research that uses an integrative approach to study SVs and their consequences in the cervical cancer genome.

Here, we have used normal and cervical cancer tissue data high-throughput chromosome conformation capture (Hi-C), transcriptome (RNA-seq), and WGS to identify SVs, specifically translocations. We monitored their local and global effects on the chromosomal 3D organization and gene expression. The results will help us to get a better insight into the correlation between SVs, specifically translocations and expression of oncogenes in cervical cancer.

MATERIALS AND METHODS

Data Source

Hi-C data generated from Digestion Ligation Only Hi-C (DLO Hi-C) technique (Lin et al., 2018), WGS, and RNA-sequence data for normal and cervical cancer tissues were downloaded from Genome Sequence Archives[1] under accession number CRA001401.

Hi-C Data Processing and Breakpoint Detection

For Hi-C data processing, we used human genome hg19 and HPV-16 genome merged assembly as a reference genome. First, quality control of raw fastq files was performed with FastQC v0.11.8 (Andrews, 2010). The DLO Hi-C tool (Hong et al., 2019) was used to process the Hi-C data generated by the Digestion-ligation-only Hi-C technique (Lin et al., 2018). This tool removes pair end tags (PETs) of self-ligation, re-ligation, and dangling pairs. The contact matrices at different resolutions were normalized using ICE method (Imakaev et al., 2012). Topologically associated domains (TADs) and TAD boundaries at 40 kb resolution were identified using TopDom R-Package at default parameters (Shin et al., 2016). Juicer eigenvector was used to define A/B compartment, and bins with positive values were considered as A compartments, while bins with negative values were defined as B compartment at 500 kb resolution (Durand et al., 2016b). HiTC Bioconductor Package was used for quality control analysis of Hi-C data (Servant et al., 2012).

As we know, Hi-C data represent the contact probabilities between two regions of interacting chromosomes in a matrix form, which enables the detection of translocation. So, we used publically available pipelines such as HiCTrans (Chakraborty and Ay, 2018) and hic_breakfinder[2], which use Hi-C data to find translocations. HiCTrans takes Hi-C contact matrices as an input to find translocation breakpoints based on change point values obtained by calculating the contact probability across each chromosomal contact pair (Chakraborty and Ay, 2018). hic_breakfinder uses mapped file (*.bam file) as an input and human genome assembly-based filtering list of false positives and reports refined translocations at different resolutions (1 Mb, 100 kb, and 10 kb). Moreover, we used an *in-house* build script that uses a valid-pairs file of Hi-C data in *bedpe* format and detects chromosomal breakpoints. The resulted breakpoints of all tools were compared by using *bedtools pairToPair* to find overlapped and unique translocated regions.

Whole-Genome Sequence Data Analysis and Structural Variation Detection

After quality control check of cervical cancer tissue and normal blood WGS data (Experiment ID: CRX040585 and CRX101064), through FastQC (Andrews, 2010) and Trimmomatic (Bolger et al., 2014), refined raw reads were aligned against proxy genome (hg19 + HPV16) using Burrows–Wheeler Alignment (BWA) tool (Li and Durbin, 2009) at default parameters, and the duplicates were marked and removed using Picard[3]. SAMtools was used for alignment quality estimation and sorting bam reads (Li et al., 2009). For SV detection in cervical cancer tissue WGS data (Experiment ID: CRX040585), we used Manta-tumor

---

[1]https://bigd.big.ac.cn/gsa/s/P37lFNi0
[2]https://github.com/dixonlab/hic_breakfinder
[3]http://broadinstitute.github.io/picard/

only (Chen et al., 2016). SV caller at default settings, additional refinement "PASS" parameter was applied, and results were visualized by Integrative Genome Viewer (IGV) (Robinson et al., 2011). ANNOVAR was used to annotate the SVs detected by Manta (Wang et al., 2010). Copy number variation (CNV) analysis was carried out by Control-FREEC software (Boeva et al., 2012). WGS data of cervical cancer tissue (Experiment ID: CRX040585) were used as an input. The ploidy parameter was set to 2 and other parameters were set as "default."

RNA-Sequence Analysis

RNA-Seq data of three normal (Experiment ID: CRX040582, CRX040583, and CRX040584) and two cervical cancer data (Experiment ID: CRX040580 and CRX040581) biological replicates were pre-processed as described (Andrews, 2010; Bolger et al., 2014) and mapped against Y-Chromosome less, HPV-16 and hg19 merged genome using HISAT2 tool (Kim et al., 2015). Gene expression abundance was quantified through featureCounts (Liao et al., 2014), and the gene expression level was calculated in RPKM value. Differentially expressed genes (DEGs) were detected by using DESeq2 R-package (Love et al., 2014). For enrichment analysis of DEGs, we used PANTHER online resource[4] gene ontology (GO) tests, and statistical enrichment tests. To gain an overview of the gene pathway networks, web-based Kyoto Encyclopedia of Genes and Genomes (KEGG) server was recruited[5]. Furthermore, we used EnrichR (Chen et al., 2013) to assess the TF-lof enrichment, ENCODE TF ChIP-Seq enrichment, and Virus–Host Protein–Protein Interactions of selective genes list.

RESULTS

Comparative 3D-Genome Structural Analysis

In order to find the genome-wide structural architecture variations, we compared the cervical cancer and normal tissue Hi-C data. Four replicates of two cervical cancer experiments (Experiment ID: CRX040576 and CRX040577) and one replicate of normal tissue DLO Hi-C data (Experiment ID: CRX040578) were used for analysis (**Supplementary Table 1**). DLO Hi-C tool first filtered out the same (AA, BB) as well as the different (AB, BA) linkers, ~285 million for normal tissues ~146, ~142, ~156, and ~162 million reads were obtained from four cervical cancer tissue replicates (**Supplementary Table 2**). Hi-C results showed the numbers of valid reads of normal sample CRX040578, cervical cancer tissue CRX040576, and cervical cancer tissue CRX040577 as 60,929,741, 28,518,853, and 36,389,304, respectively (**Supplementary Table 3**). Next, we visualized the whole-genome interaction map of normal and cervical cancer tissues to detect the differential arrangements. The higher order genomic organization was observed; apparently, the chromosomal architecture was consistent between normal and cervical cancer tissue heatmaps, but some chromosomes showed differential organizations (**Figures 1A,B**; Cao et al., 2020). We

visualized the whole-genome contact matrices for both samples through juicebox (Durand et al., 2016a) and found that various regions showed the differential interactions frequency between different chromosomes (**Figure 1C**). The *cis*-interaction ratio between both samples was very similar, but the trans-interactions showed a significant increase (Fisher's exact test p-value = 2.2e-10) (**Supplementary Table 3** and **Figure 1D**). Several differential genomic organizations were detected in different chromosomes (**Supplementary Figure 1**). For example, in chromosome 7, higher order rearrangements were observed, and variable regions were found at 10–45 and 75–110 Mb regions (**Figure 2A**). Another variable region was present at 75–110 Mb position. A distinctive interaction pattern (enlarged and highlighted with the black square) appeared in cervical cancer tissues, but it was missing at the normal sample's corresponding chromosomal region. In chromosome 4, two significant arrangements were observed from 0–50 and 55–191 Mb (highlighted with the black squares). The more dense architecture was observed at 55–191 Mb region in chromosome 4 of the cervical cancer sample (**Figure 2B**).

Further, we also checked the A/B compartments in both samples at 500 kb resolution. 76% of the total genome remained conserved, and only 13 and 11% of the genome showed compartment switching from A to B and B to A, respectively (Fisher's exact test p-value < 2.2e-16) (**Figure 2C**). Moreover, we also monitored the A/B compartment switching at a chromosomal level such as in chromosome 7, 4 (**Figures 2A,B**), 1, and 16 (**Supplementary Figure 1**) several regions showed A/B compartment switching. It has been demonstrated that tumor development is associated with the alterations of TADs (Valton and Dekker, 2016). TADs at 40 kb resolutions were detected in each Hi-C experiment data; a total of 6,468, 6,033, and 6,268 TADs were found in normal tissue (CRX040578), cervical cancer tissue 1 (CRX040576), and cervical cancer tissue 2 (CRX040577), respectively. The comparison of TAD boundaries between experimental samples was calculated with *bedtools intersect -f 0.70 –r* parameters and represented in the bar-graph (**Figure 2D**). We identified that 2,260 TADs were shared between all samples, 817 TADs were conserved, 1,998 and 2,342 were unique between cervical cancer tissue 1 (CRX040576) and normal tissue (CRX040578), respectively (Fisher's test p-value = 0.00629), and 1,049 TADs found overlapped, 2,001 and 2,343 TADs were found as unique between cervical cancer tissue 2 (CRX040577) and normal tissues (CRX040578), respectively (Fisher's test p-value = 0.003128). We also detected and compared the number of TADs between both cervical cancer samples (CRX040577 and CRX040576) and found 958 overlapped and 2,001 and 1,998 unique TADs which were statistically non-significant (Fisher's test p-value = 1). Collectively, these results suggested that 3D-genome architecture shows differential behavior from normal to a cancerous condition.

Translocation Identification in Cervical Cancer Hi-C Data

We observed a large inter-chromosomal interaction region during Hi-C data analysis that suggests a translocation event in the cervical cancer sample (Experiment ID: CRX040576 and
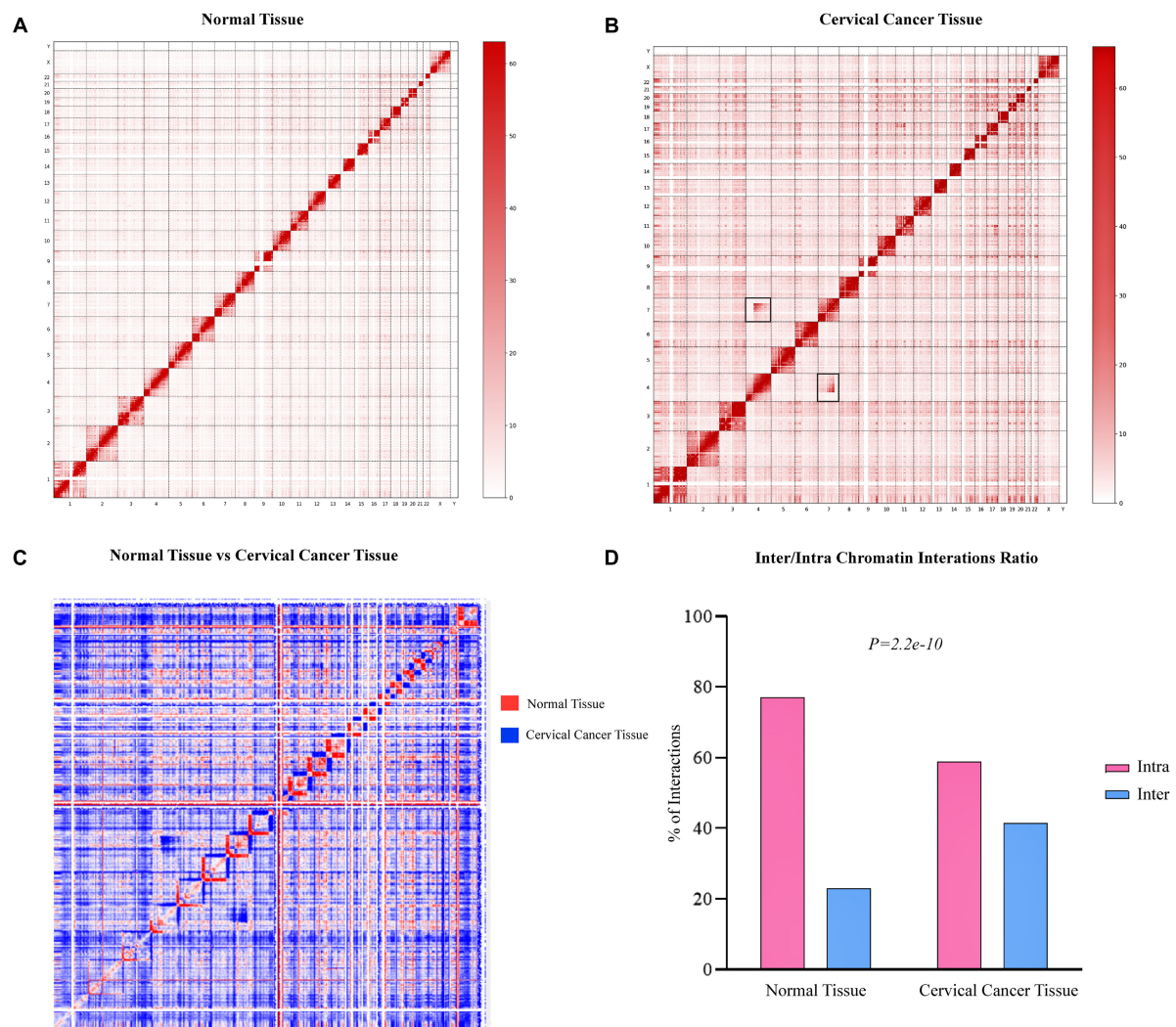
FIGURE 1 | Hi-C data showing genome-wide variations in cervical cancer 3D-genome. Genome-wide Hi-C interaction map at 1 Mb resolution. Heatmap representing normal **(A)** and cervical cancer **(B)** tissues chromosomal interactions, respectively. Black squares indicate the inter-chromosomal rearrangements. **(C)** A heatmap shows the difference of higher order chromatin interactions between normal tissue (red color) and cervical cancer (blue color) tissue Hi-C data. **(D)** Hi-C quality analysis indicating the number of inter- and intra-chromosomal interactions between both samples. Red bars depict intra-chromosomal interactions, while sky-blue bars denote inter-chromosomal interactions.

CRX040577). To predict the translocated area in a cervical cancer tissue sample, we further analyzed the Hi-C data through available tools such as hic_breakfinder (Dixon et al., 2018) and HiC-trans (Chakraborty and Ay, 2018). hic_breakfinder uses bam input and detects translocations by creating sub-matrices of the original matrix that potentially contains chromosomal rearrangements (Dixon et al., 2018). It predicted seven breakpoints in cervical cancer sample 1 (CRR045289) and 3 (CRR045291) each, and eight breakpoints in cervical cancer sample 2 (CRR045290) and sample 4 (CRR045288). A total of seven chromosomal pairs were observed that undergo translocations, such as chr2-chr12, chr3-chr12, chr4-chr7, chr16-chr1, chr6-chr5, chr17-chr11, and chr3-chr6; the breakpoint boundaries for each pair are given in **Supplementary Table 4**. HiC-trans detected several chromosomal pairs with the translocations at different bin sizes 40, 80, and 120 kb;

in cervical cancer Hi-C sample 1 (CRR045289) 25, in cervical cancer Hi-C sample 2 (CRR045290) 35, in cervical cancer Hi-C sample 3 (CRR045291) 49, and in cervical cancer Hi-C sample 4 (CRR045288) it predicted 54 breakpoints (**Supplementary Table 5**). We observed that hic_breakfinder and HiC-trans use different detection approaches by considering different biases that resulted in more false positive detections. To overcome that issue, we build an *In-house script* that detects the obvious breakpoints using interacting pair files as input. It predicted six translocated chromosomal pairs in cervical cancer Hi-C sample 1 (CRR045289) in such a way that two in chr4-chr7 and four breakpoints in chr1-chr16 pair that show translocations (**Supplementary Figure 2**). In other cervical cancer Hi-C samples (CRR045290, CRR045291, and CRR045288), our *script* predicted two breakpoints in chr4-chr7 pair each (**Supplementary Table 6**). The translocation between

FIGURE 2 | Comparative Hi-C data revealing the genome architecture changes at chromosomal level. Differential chromosomal architecture. **(A,B)** Comparison of chromatin interaction heatmaps from chromosomes 7 and 4 between normal (left panel) and cervical cancer (right panel) tissues. Variable regions are highlighted with the black boxes. Interaction maps are shown at 500 kb resolution. A/B compartment switching between normal and cervical cancer tissue was detected at 500 kb resolution using juicer-eigenvector. Values greater than 0 were assigned as A compartment (green color), and values less than 0 were designated as B compartment (Red color). **(C)** A/B compartment changes between normal and cervical cancer tissue data. Bar-graph representing overall compartment switching, orange: conserved compartments, green: A to B compartment conversion, and sky blue: B to A compartment change. **(D)** Bar-graphs show the comparison of the number of TADs found in cervical cancer and normal Hi-C data. Purple: TADs found in normal tissues, red: cervical cancer data in CRX040576 experiment, and blue: TADs found in cervical cancer tissue CRX040577 experiment.

chr7:123,374,769–123,376,789 and chr4:63,481,072–63,483,072 is shown in **Figure 3A**. The translocated region between chr1: 144,816,374–144,826,374 and chr16:70,838,537–70,848,537 is shown in **Supplementary Figure 2**.

We compared the translocation boundaries detected by publically available tools with our *script*; we used *bedtools pairToPair* to find the overlaps and unique translocations. Here, we took one sample as an example in which hic_breakfinder detected eight breakpoints, HiCtrans 25, and our *script* detected six breakpoints. After comparison, we found that eight breakpoint regions were overlapped between hic_breakfinder and HiCtrans. Five breakpoint regions were similar between the results of *in-house script* and hic_breakfinder. Neither had we detected any overlap between HiCtrans and our *script* nor in all other tools (**Figure 3B**).

Structural Variation Detection in Whole-Genome Sequence Data

We analyzed WGS data of the corresponding cervical cancer tissue data of patients (Experiment ID: CRX040585) to find the novel SVs such as deletions, translocations, insertions, and duplications, and to evaluate the precision of translocations detected in Hi-C data. Sequence quality was determined by

FastQC (Andrews, 2010). Sequence contaminations such as overrepresented sequences and low-quality reads were clipped using appropriate sequence clipper in Trimmomatic (Bolger et al., 2014). Total 99.66% reads were used for mapping, and 97.82% read pairs were adequately aligned against the customized reference genome (hg19 + hpv16). The average read depth of the cervical cancer sample was calculated by SAMtools (Li et al., 2009), which was ∼45X. Manta Tumor-Only Analysis was performed to find the structural variants (Chen et al., 2016). Manta predicted that 3,579 reads have maximum depth, 301 reads did not match with default filtration score or aligned to multiple locations around the breakpoints, and 16,712 variants passed the filtration threshold score of Manta.

Further, we split the genome-wide SVs into their respective types, such as 7,858 inter-chromosomal translocations breakpoints (BND), 5,799 deletions (DEL), 1,185 duplications (DUP), and 1,870 inversions (INV). Our primary focus was to check the consistency of translocation results of Hi-C data, such as chr4-chr7 and chr1-chr16, with variations identified by the Manta tool. We found that the positions of chr7-chr4 and chr1-chr16 breakpoints were coherent with inter-chromosomal translocation identified from WGS data (**Figure 3C**). Additionally, we have also inspected the WGS paired read analysis, and found the presence of

FIGURE 3 | Hi-C and WGS results showing the consistency of translocations detected in cervical cancer samples. **(A)** Magnified representation of translocated regions observed by visual observation (left panel), while the right panel (zoomed snapshot) represents the Hi-C breakpoints between chromosome 7 and chromosome 4 detected by *in-house script* at highest resolution. Red peaks show cervical cancer RNA, while green color indicates the normal RNA sequence. Reference genes are highlighted with black color. **(B)** UpSet plot of breakpoints results detected by different methods. In the bottom left panel, blue horizontal bars represent the number of translocations detected by each method; vertical red-colored bars represent the size of breakpoint intersections of each translocation sets. Black dots show the sample set, and the intersection between methods is represented by a vertical black line. This graph is generated by UpSetR package. **(C)** Circos plot representing the chromosomal rearrangements between chr1, chr4, chr7, and chr16 detected by Hi-C and WGS. Deep pink- and green-colored arcs show the overlapped results of Hi-C and WGS, while blue color represents WGS-based translocations. **(D)** The translocation event between *t*(4;7) (q13.1; q31.32) in cervical cancer data visualized by IGV tool, showing WGS paired reads supporting the translocation results of Hi-C data. Breakpoints are labeled by "BND" in the upper panels of both IGV windows. Upper IGV window represents Chr4 reads, and the lower window shows Chr7 reads. Green and red colors depict the first and second read of mate-pairs.

translocated mate-pair reads in chr1-chr16 (**Supplementary Figure 2**) and chr4-chr7 (**Figure 3D**). Moreover, we detected the protein-coding genes at the translocated regions, specifically in chr4:61,875,072–61,895,072 and chr7:123,370,769–123,376,769 and chr1:144,816,374–144,826,374 and chr16:70,838,537–70,848,537 region. WASL gene was found at chromosome 7 (q31.3) (**Figure 3A**), NBPF20 at 1 (q21.1), and HYDIN gene was present at chromosome 16 (q22.2) (**Table 1** and **Supplementary Figure 2**). WGS annotation results suggested that translocation between chr4-chr7 has a very "high" impact on this WASL gene. Copy number variation is another key phenomenon that contributes to cancer development. So, Control-FREEC identified several copy number variations (CNVs) in multiple chromosomes such as chromosome 1, 2, 4, 8–11, 16, 18, and 21.

A total 249 "gain" and 32 "loss" events occurred (**Supplementary Figure 3** and **Supplementary Table 7**). These results collectively showed that translocations identified by Hi-C data are consistent with the WGS data. Additionally, we found several protein-coding genes at the translocated region directly involved in female cancer development.

Effects of Translocations on Gene Expression

Previously, it is reported that SVs play a significant role in changing gene expression, leading to cancer development. To explore the effect of SV mainly the translocations on the expression of surrounding genes, we used the transcriptome

TABLE 1 | Translocation breakpoints and neighboring genes around translocated pairs.

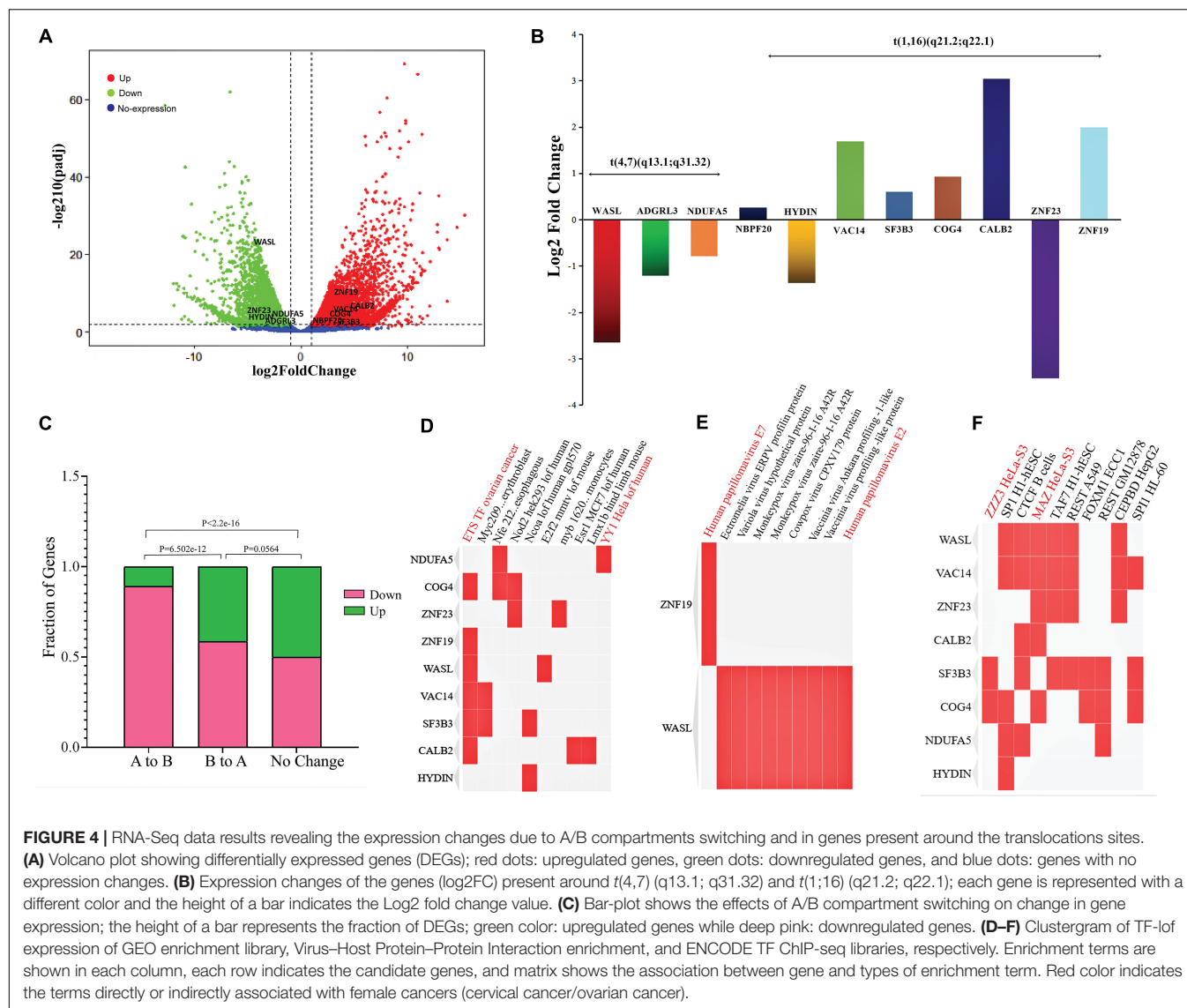| Karyotype	Breakpoint coordinates	Disrupted genes	Neighboring genes
t(4;7) (q13.1; q31.32)	Chr4:61,875,072–61,895,072 Chr7:123,370,769–123,376,769	Intergenic region WASL	ADGRL3 NDUFA5
t(1;16) (q21.2; q22.1)	Chr1:144,816,374–144,826,374 Chr16:70,838,537–70,848,537	NBPF20 HYDIN	VAC14 SF3B3 COG4 CALB2 ZNF23 ZNF19

of cervical cancer and normal cervix tissue. A total of ∼8,000 genes were detected as DEGs (**Supplementary Table 8**) that fulfill the criteria of false discovery rate (FDR) < 0.05 and absolute Log2 fold change (Log2 FC) > 1 (**Figure 4A**). Next, we were curious to determine the effects of translocation on the expression of genes present within translocated regions and in their vicinity. In t(4;7) (q13.1; q31.32), WASL gene (7q31.3) underwent translocation, showed expression disruption, and was significantly downregulated, while at chromosome 4 (4q13.1), an intergenic region was observed. We further extended our analysis in the neighboring (within ± 1 Mb region) genes at chromosome 4, such as ADGRL3 and NDUFA5, and found that both genes were downregulated (**Figure 4B**). We also monitor the gene expression changes in t(1;16) (q21.2; q22.1) and nearby regions. This translocation occurred within NBPF20 and HYDIN genes located at 1q21.2 and 16q22.1, respectively (**Supplementary Figure 2**). The former was upregulated, and the latter was downregulated. The neighboring genes such as ZNF23 showed downregulated expression, while VAC14, SF3B3, COG4, CALB2, and ZNF19 appeared as upregulated genes (**Figure 4B** and **Table 1**).

Since our Hi-C results showed a significant A/B compartment switching, previous studies have already reported the correlation between compartments switching and gene expression changes (Wu et al., 2017). So, here we aimed to check how many genes were affected by compartment switching in translocated chromosomes (1, 4, 7, and 16). In A to B compartment change, a total of 239 and 36 genes were found as down- and upregulated, respectively, while in B to A, 69 genes were downregulated and 49 were detected as upregulated genes. In the no-change category, 688 genes showed upregualted and 691 showed downregulated expression. Fisher's exact test suggested that gene expression changes between A to B and B to A category were fairly significant (Fisher's exact test *p-value* = 6.502e-12). In B to A and no-change category, a significant (Fisher's test *p-value* = 0.0564) number of genes were found to be changing gene expression, while the number of genes changing expression in A to B and conserved genome category was found as highly significant (Fisher's exact test *p-value* < 2.2e-16) (**Figure 4C**).

In GO analysis, we predicted the overall enrichment of DEGs in cellular components and molecular functions. Results suggested that the DEGs were involved in maintaining different cellular components and regulating various molecular functions such as structural constituent of ribosome (18%), immune receptor activity (12%), cytokine binding (12%), structural molecular activity (8%), and signaling receptor binding (8%) (**Supplementary Figure 4**).

Additionally, we extracted the list of genes around the translocation and performed GO using TF-lof expression, ENCODE TF ChIP-seq, and Virus–Host PPI enrichment libraries by enrichR annotation platform. Results showed that in TF-lof expression library enrichment, all genes except ZNF23, NDUFA5, and HYDIN were controlled by previously reported *ETS* transcription factor (up expression) of ovarian cancer (Llauradõ et al., 2012). NDUFA5 gene was significantly enriched with the transcription factor *YY1* (down-expression) of HeLa cell line consistent with the previous studies of human (Rizkallah and Hurt, 2009). ZNF23 showed *myb* TF enrichment in primary monocytes of humans (Huang et al., 2007; **Figure 4D**). Virus–Host PPI enrichment analysis suggested that WASL and ZNF19 genes were highly enriched in interacting with HPV E7 and HPV type 144 proteins, respectively (**Figure 4E**; White et al., 2012). ENCODE TF ChIP-Seq data library enrichment results showed that SF3B3 and COG4 were enriched by transcription factor *ZZZ3* and WASL, ZNF23, VAC14, and CALB2 genes were controlled by *MAZ* TF of HeLa-S3 (cervical cancer) cell line (**Figure 4F**; Dolfini et al., 2016). Several other TFs were also detected that potentially influence the transcription of the genes mentioned above. Pathway enrichment analysis showed the association of these genes in carcinomal pathways such as TP53 pathway (Tommasino et al., 2003), tumor suppressor pathway (Cohen et al., 2003), PDGF pathway (Li et al., 2017), p38 MAPK signaling pathway (Kang and Lee, 2008), G13 signaling pathway (Yuan et al., 2016), and Notch signaling pathway (Rodrigues et al., 2019; **Figure 5**). Overall, we concluded that gene expression significantly changed around translocated region in cervical cancer samples and A/B compartment changes lead to change in gene expression; moreover, the enrichment analysis suggested that the regulation of genes present at translocation regions was controlled by previously reported transcription factors of cervical cancer/ovarian cancer studies. Additionally, we have also predicted the direct role of newly identified genes in cervical cancer-related pathways.

FIGURE 4 | RNA-Seq data results revealing the expression changes due to A/B compartments switching and in genes present around the translocations sites.
(A) Volcano plot showing differentially expressed genes (DEGs); red dots: upregulated genes, green dots: downregulated genes, and blue dots: genes with no expression changes. **(B)** Expression changes of the genes (log2FC) present around *t*(4,7) (q13.1; q31.32) and *t*(1;16) (q21.2; q22.1); each gene is represented with a different color and the height of a bar indicates the Log2 fold change value. **(C)** Bar-plot shows the effects of A/B compartment switching on change in gene expression; the height of a bar represents the fraction of DEGs; green color: upregulated genes while deep pink: downregulated genes. **(D–F)** Clustergram of TF-lof expression of GEO enrichment library, Virus–Host Protein–Protein Interaction enrichment, and ENCODE TF ChIP-seq libraries, respectively. Enrichment terms are shown in each column, each row indicates the candidate genes, and matrix shows the association between gene and types of enrichment term. Red color indicates the terms directly or indirectly associated with female cancers (cervical cancer/ovarian cancer).

DISCUSSION

Cervical cancer is the leading type of female cancer caused by HPV genome integration (Li et al., 2021). HPV integration has complex effects on the phenotype, and the mechanism driving these effects is poorly understood. Some studies have reported that structural rearrangements are also the driving force behind tumorigenesis (Zhang et al., 2018). These SVs cause higher order disorganizations, which we assume play a significant role in changing gene expression and ultimately result in tumor development and progression around the cervix tissues. Several studies have shown that SVs, including duplications, deletions, translocations, insertions, and inversions, can disrupt the 3D-genome specifically the TAD boundaries, in a way that they can produce neo-TADs, fused-TADs, or can cause the deletion of TADs (Valton and Dekker, 2016). In HPV-induced cervical cancer, the remodeling of TADs is associated with the enhancer-hijacking (Cao et al., 2020), which leads to the change in gene expression followed by the deleterious phenotypes such

as developmental disorders and cancer (Melo et al., 2020). Moody and Laimins (2010) have reported that in HPV + state, viral E6 and E7 oncogenes are predominantly attributable to the SVs/instability of the host genome. Previous research has reported that HPV-integrations sites are prone to change in the local structure of host loci and gene expression (Cao et al., 2020). The detection of SVs is still a burning issue in cancer research. This study used Hi-C, RNA-Seq, and WGS data of cervical patients to perform a comparative analysis between cervical cancer and normal tissue. Here, we identified the genome-wide SVs, specifically the translocations away from the HPV-integration point using Hi-C-based translocation detection methods, and to monitor their effects on gene expression around the breakpoints. Hi-C data from four cervical cancer patients were used to detect the chromosomal interactions and rearrangements; comparative results showed that overall 3D-genome architecture appeared to be consistent between normal and cervical cancer tissue data consistent with the findings of the previous study (Cao et al., 2020), although

FIGURE 5 | GO analysis of translocation neighboring genes showing their involvement in different pathways controlling cervical cancer development. Pathway enrichment analysis of translocations neighboring genes. Red-colored bar indicates the pathways that are directly involved in HPV-mediated cervical cancer development. Blue-colored bar shows the pathways involved in other types of cancers.

some chromosomes showed higher order chromatin structure variations, including a significant change in the number of TADs and A/B compartment switching.

Furthermore, we have observed some translocations in the genome-wide interaction map. HiC-Trans (Chakraborty and Ay, 2018) and hic_breakfinder (Dixon et al., 2018) were used to locate the translocations precisely. Both tools produced a higher number of breakpoints, and a higher number suggested the high false discovery rate (FDR) (Wang et al., 2020). To overcome this problem, we have designed *in-house script* that only detected the obvious translocation breakpoints at the highest resolution and less computational cost. Two translocations such as t(4;7) (q13.1; q31.32) and t(1;16) (q21.2; q22.1) were detected. WGS data validated the presence of translocations detected by the *in-house script*.

Additionally, WGS paired read analysis found the presence of translocated reads in chr1-chr16 (**Supplementary Figure 2**) and chr4-chr7. WGS results were coherent with the breakpoints detected by the newly designed script in Hi-C data. The coherence of Hi-C and WGS results confirmed the sensitivity and specificity of our *in-house script* output. Further, annotation results suggested the presence of intergenic regions and coding genes such as WASL, NBPF20, and HYDIN at the breakpoint region. Previous studies have suggested the active role of NBPF20 in gene fusion in cervical cancer

patients (Li et al., 2021). Mutagenesis studies of gynecological cancers have revealed that HYDIN gene undergoes frequent mutations in both ovarian and cervical cancer (Guo et al., 2020). Published research has demonstrated that the WASL (Wiskott–Aldrich Syndrome-Like) gene belongs to the oncogenes category and plays a significant role in tumor progression and metastasis in cervical cancer (Hidalgo-Sastre et al., 2020). CNV analysis depicted that chromosomes 1, 4, 7, and 16 were among the chromosomes which undergo copy number variations consistent with the previous cervical cancer studies (Adey et al., 2013). Collectively, 3D-genome structure, WGS, and CNVs results showed a strong association between the chromosomal architecture and breakpoints.

Chromosomal translocations lead to the disruption of gene expression and cause proto-oncogenes activation (Rabbitts, 1994). Some studies also reported that SVs in general have multiple local and global effects on chromosomal structure, chromatin interactions, and gene expression (Zhang et al., 2018). Here, we aimed to study the potential impact of translocations on gene expression. The transcriptome data of cervical cancer suggested many upregulated and downregulated genes compared to the normal tissue sample. We obtained the neighboring genes around the breakpoints regions in chromosomes 1, 16, 4, and 7 and found the disrupted gene expression. GO analysis depicted

that DEGs were involved in various molecular functions and cellular processes.

Furthermore, we performed various enrichment analyses using different GO libraries such as TF-lof expression of GEO enrichment, Virus–Host Protein–Protein Interaction enrichment, and ENCODE TF ChIP-seq enrichment libraries. TF-lof expression of GEO enrichment results showed that the most of the detected genes were enriched with the previously reported transcription factors YY1 of Hela-cells (cervical-cancer cell line) (Rizkallah and Hurt, 2009) and ETS transcription factor of ovarian cancer (Llauradō et al., 2012). Virus–Host PPI enrichment results suggested the interaction enrichment of disrupted genes such as ZNF19 with HPV-E7 and WASL with HPV-E2 (Cohen et al., 2003). Next, we examined the regulatory network of detected genes through ENCODE TF ChIP-Seq enrichment library and found that these genes were enriched with the previously reported TFs of Hela-S3 (cervical cancer) cell lines.

Pathway analysis results also showed that the genes located in breakpoint regions are strongly associated with various cancer-mediated pathways such as TP53 network, GPCRs Class B Secretin-like pathway, p38 MAPK signaling pathway, apoptosis-related network in ovarian cancer, Notch signaling pathway, PDGF pathway, and IGF1-Akt signaling (Steller et al., 1996; Serrano et al., 2008; Chen, 2015; Wu et al., 2017). Overall, these findings provide strong evidence that breakpoints occurred in the genes that have a strong correlation with HPV-mediated cervical cancer.

Although we have predicted some novel translocations, there are some limitations associated with this study, for example, the limited availability of a dataset used to carry out the analyses. Increasing the size of cohort will help to get better understandings about cervical cancer development mechanism. Another major limitation is the heterogeneous nature of cervical cancer making the study more challenging. Combining all analyses, we unveil that in cervical cancer, multiple genomic alternations such as translocations, CNVs, and 3D reorganization occur that affect the gene expression.

These findings shed light on the importance of studying the effects of SVs on the 3D genome and finding candidate genes in cervical cancer. We believe that this will help us to improve our understandings of the HPV-mediated cervical cancer mechanism and identify the targeted clinical therapeutics against cervical cancer.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**. The *in-house script* is available at https://github.com/Adeel3Dgenomics/In-house-script-for-translocation-detection.

AUTHOR CONTRIBUTIONS

MA and GL: conceptualization, writing—original draft preparation, and writing—review and editing. MA: design.

HJ: software. CC: sample collection. DL and CC: biological experiments. MA, HJ, KC, and YA: formal data analysis. GL, PW, and GC: supervision. GL and PW: funding acquisition. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.706375/full#supplementary-material

Supplementary Figure 1 | Hi-C results showing the comparative analysis for chromosome 1 and chromosome 16 in both normal and cervical cancer samples. Differential chromosomal architecture: **(A)** Chromosome 1 of normal tissue (left) and cervical cancer tissue (right) Hi-C heatmaps. A/B compartments between both samples have been shown in the lower panel of A. **(B)** Chromosome 16 of the normal sample (left) and cervical cancer tissue (right). A/B compartments are shown in the bottom panel. Dark Green; A compartment, Red; B compartment.

Supplementary Figure 2 | Hi-C-based translocation detected in chromosome 1 and 16 was coherent with the WGS mate-pairs analysis. **(A)** Enlarged Hi-C interaction map shows the breakpoint position corresponding to translocation t(1;16) (q21.2; q22.1). NBPF20 and HYDIN genes were found at the translocation region. RNA-seq peaks are also visualized, red-colored peaks indicate cervical cancer RNA, and green color shows normal sample gene expression. **(B)** Visualization of translocated read pairs in translocation event t(1;16) (q21.2; q22.1). Red: translocated read mate-pairs.

Supplementary Figure 3 | Copy number variations detected in cervical cancer sample showing that translocations containing chromosomes also undergo CNVs. Estimated genome-wide copy number variations (CNVs) from cervical cancer whole-genome sequence (WGS) data. Red dots represent "gains," blue dots show "loss" events, and green dots show "normal" events.

Supplementary Figure 4 | Gene ontology analysis of DEGs. Enrichment of genes in various **(A)** cellular components; red bars indicate the cellular component term and length represents the −log10 (FDR). **(B)** Molecular functions; each slice of the circle depicts the molecular function and gene involvement percentage.

Supplementary Table 1 | Data source for cervical cancer. The table shows the experiment ID, sequencing run accession number, sequence type, and total reads.

Supplementary Table 2 | Linkers filtering information. Linkers detected by DLO Hi-C tool from normal tissue and cervical cancer tissues (1, 2, 3, and 4) DLO Hi-C dataset.

Supplementary Table 3 | Hi-C data analysis stats for normal and cervical cancer samples. Hi-C results for each experiment; each table shows the information of

self-ligated and non-ligated pairs, valid reads, intra–inter chromosomal interactions, and long- and short-range chromatin interactions.

Supplementary Table 4 | Translocations detected by hic_breakfinder. Translocations detected by hic_breakfinder in cervical cancer samples, the table showing the translocation boundaries along with the strand information.

Supplementary Table 5 | Translocations identified by HiC_Trans. Translocations detected by HiC_Trans in cervical cancer samples, the table showing the breakpoint boundaries along with the z-score.

Supplementary Table 6 | Translocations identified by *in-house* script. Breakpoints detected by in-house build script in cervical cancer sample Hi-C data, the table showing the breakpoint boundaries information of translocated regions.

Supplementary Table 7 | Copy number variations (CNVs) in cervical cancer WGS. CNVs identified by Control-FREEC in cervical cancer data, number of gains, and loss events are shown.

Supplementary Table 8 | Gene expression data for cervical cancer tissue. DEGs detected in cervical cancer and normal RNA-seq data of cervical cancer patients.

REFERENCES

Adey, A., Burton, J. N., Kitzman, J. O., Hiatt, J. B., Lewis, A. P., Martin, B. K., et al. (2013). The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line. *Nature* 500, 207–211. doi: 10.1038/nature12064

Alkan, C., Coe, B. P., and Eichler, E. E. (2011). Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* 23, 363–376. doi: 10.1038/nrg2958

Andrews, S. (2010). Available online at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappo, J., Schleiermacher, G., et al. (2012). Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* 28, 423–425. doi: 10.1093/bioinformatics/btr670

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., Jemal, A., et al. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492

Cao, C., Hong, P., Huang, X., Lin, D., Gang, C., Wang, L., et al. (2020). HPV-CCDC106 integration alters local chromosome architecture and hijacks an enhancer by three-dimensional genome structure remodeling in cervical cancer. *J. Genet. Genomics* 47, 437–450. doi: 10.1016/j.jgg.2020.05.006

Chakraborty, A., and Ay, F. (2018). Identification of copy number variations and translocations in cancer cells from Hi-C data. *Bioinformatics* 34, 338–345. doi: 10.1093/bioinformatics/btx664

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., et al. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 14:128. doi: 10.1007/s00701-014-2321-4

Chen, J. (2015). Signaling pathways in HPV-associated cancers and therapeutic implications. *Rev. Med. Virol.* 25, 24–53. doi: 10.1002/rmv.1823

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., et al. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* 32, 1220–1222. doi: 10.1093/bioinformatics/btv710

Cohen, Y., Singer, G., Lavie, O., Dong, S. M., Beller, U., Sidransky, D., et al. (2003). The RASSF1A tumor suppressor gene is commonly inactivated in adenocarcinoma of the uterine cervix. *Clin. Cancer Res.* 9, 2981–2984.

Cui, C., Shu, W., and Li, P. (2016). Fluorescence in situ hybridization: cell-based genetic diagnostic and research applications. *Front. Cell Dev. Biol.* 4:89. doi: 10.3389/fcell.2016.00089

Dai, M., Bao, Y. P., Li, N., Clifford, G. M., Vaccarella, S., Snijders, P. J., et al. (2006). Human papillomavirus infection in Shanxi Province, people's republic of China: a population-based study. *Br. J. Cancer* 95, 96–101. doi: 10.1038/sj.bjc.6603208

Dixon, J. R., Xu, J., Dileep, V., Zhan, Y., Song, F., Le, V. T., et al. (2018). Integrative detection and analysis of structural variation in cancer genomes. *Nat. Genet.* 50, 1388–1398. doi: 10.1038/s41588-018-0195-8

Dolfini, D., Zambelli, F., Pedrazzoli, M., Mantovani, R., and Pavesi, G. (2016). A high definition look at the NF-Y regulome reveals genome-wide associations with selected transcription factors. *Nucleic Acids Res.* 44, 4684–4702. doi: 10.1093/nar/gkw096

Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., et al. (2016a). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* 3, 99–101. doi: 10.1016/j.cels.2015.07.012

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S., Huntley, M. H., Lander, E. S., et al. (2016b). Juicer provides a one-click system for analyzing loop-resolution hi-c experiments. *Cell Syst.* 3, 95–98. doi: 10.1016/j.cels.2016.07.002

Engreitz, J. M., Agarwala, V., and Mirny, L. A. (2012). Three-dimensional genome architecture influences partner selection for chromosomal translocations in human disease. *PLoS One* 7:e44196. doi: 10.1371/journal.pone.0044196

Fudenberg, G., Getz, G., Meyerson, M., and Mirny, L. A. (2011). High order chromatin architecture shapes the landscape of chromosomal alterations in cancer. *Nat. Biotechnol.* 29, 1109–1113. doi: 10.1038/nbt.2049

Garraway, L. A., and Lander, E. S. (2013). Lessons from the cancer genome. *Cell* 153, 17–37. doi: 10.1016/j.cell.2013.03.002

Guo, Y., Liu, J., Luo, J., You, X., Weng, H., Wang, M., et al. (2020). Molecular profiling reveals common and specific development processes in different types of gynecologic cancers. *Front. Oncol.* 10:584793. doi: 10.3389/fonc.2020.584793

Hidalgo-Sastre, A., Desztics, J., Dantes, Z., Schulte, K., Ensarioglu, H. K., Bassey-Archibong, B., et al. (2020). Loss of Wasl improves pancreatic cancer outcome. *JCI Insight* 5:e127275. doi: 10.1172/JCI.INSIGHT.127275

Hong, P., Jiang, H., Xu, W., Lin, D., Xu, Q., Cao, G., et al. (2019). DLO Hi-C tool for digestion-ligation-only Hi-C chromosome conformation capture data analysis. *Genes* 11:289. doi: 10.1101/764332

Hu, L., Liang, F., Cheng, D., Zhang, Z., Yu, G., Zha, J., et al. (2020). Location of balanced chromosome-translocation breakpoints by long-read sequencing on the oxford nanopore platform. *Front. Genet.* 10:1313. doi: 10.3389/fgene.2019.01313

Huang, C., Jia, Y., Yang, S., Chen, B., Sun, H., Shen, F., et al. (2007). Characterization of ZNF23, a KRAB-containing protein that is downregulated in human cancers and inhibits cell cycle progression. *Exp. Cell Res.* 313, 254–263. doi: 10.1016/j.yexcr.2006.10.009

Imakaev, M., Fundenberg, G., McCord, R. P., Naumova, N., Goloborodko, A., Lajoie, B. R., et al. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* 9, 999–1003. doi: 10.1038/nmeth.2148

Kang, Y. H., and Lee, S. J. (2008). The role of p38 MAPK and JNK in arsenic trioxide-induced mitochondrial cell death in human cervical cancer cells. *J. Cell Physiol.* 217, 23–33. doi: 10.1002/jcp.21470

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements Daehwan HHS public access. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Koneva, L. A., Zhang, Y., Virani, S., Hall, P. B., McHugh, J. B., Chepeha, D. B., et al. (2018). HPV integration in HNSCC correlates with survival outcomes, immune response signatures, and candidate drivers. *Mol. Cancer Res.* 16, 90–102. doi: 10.1158/1541-7786.MCR-17-0153

Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499, 214–218. doi: 10.1038/nature12213

Li, H., Shi, B., Li, Y., and Yin, F. (2017). Polydatin inhibits cell proliferation and induces apoptosis in laryngeal cancer and HeLa cells via suppression of the PDGF/AKT signaling pathway. *J. Biochem. Mol. Toxicol.* 31:e21900. doi: 10.1002/jbt.21900

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, W., Lei, W., Chao, X., Song, X., Bi, Y., Wu, H., et al. (2021). Genomic alterations caused by HPV integration in a cohort of Chinese endocervical adenocarcinomas. *Cancer Gene Ther.* 1–12. doi: 10.1038/s41417-020-00283-4

Liao, Y., Smyth, G. K., and Shi, W. (2014). FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Lin, D., Hong, P., Zhang, S., Xu, W., Jamal, M., Yan, K., et al. (2018). Digestion-ligation-only Hi-C is an efficient and cost-effective method for chromosome conformation capture. *Nat. Genet.* 50, 754–763. doi: 10.1038/s41588-018-0111-2

Llauradó, M., Abal, M., Castellví, J., Cabrera, S., Gil-Moreno, A., Pérez-Benavente, A., et al. (2012). ETV5 transcription factor is overexpressed in ovarian cancer and regulates cell adhesion in ovarian cancer cells. *Int. J. Cancer* 130, 1532–1543. doi: 10.1002/ijc.26148

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550. doi: 10.1186/s13059-014-0550-8

Mardis, E. R., and Wilson, R. K. (2009). Cancer genome sequencing: a review. *Hum. Mol. Genet.* 18, 163–168. doi: 10.1093/hmg/ddp396

Melo, U. S., Schöpflin, R., Acuna-Hidalgo, R., Mensah, M. A., Fischer-Zirnsak, B., Holtgrewe, M., et al. (2020). Hi-C identifies complex genomic rearrangements and TAD-shuffling in developmental diseases. *Am. J. Hum. Genet.* 106, 872–884. doi: 10.1016/j.ajhg.2020.04.016

Moody, C. A., and Laimins, L. A. (2010). Human papillomavirus oncoproteins: pathways to transformation. *Nat. Rev. Cancer* 10, 550–560. doi: 10.1038/nrc2886

Münger, K., Baldwin, A., Edwards, K. M., Hayakawa, H., Nguyen, C. L., Owens, M., et al. (2004). Mechanisms of human papillomavirus-induced oncogenesis. *J. Virol.* 78, 11451–11460. doi: 10.1128/JVI.78.21.11451-11460.2004

Rabbitts, T. H. (1994). Chromosomal translocations in cancer. *Nature* 372, 143–149. doi: 10.1016/j.bbcan.2008.07.005

Rizkallah, R., and Hurt, M. M. (2009). Regulation of the transcription factor YY1 in mitosis through phosphorylation of its DNA-binding domain. *Mol. Biol. Cell* 20, 4524–4530. doi: 10.1091/mbc.E09

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. doi: 10.1038/nbt.1754

Rodrigues, C., Joy, L. R., Sachithanandan, S. P., and Krishna, S. (2019). Notch signalling in cervical cancer. *Exp. Cell Res.* 385:111682. doi: 10.1016/j.yexcr.2019.111682

Sanchis-Juan, A., Stephens, J., French, C. E., Gleadall, N., Mégy, K., Penkett, C., et al. (2018). Complex structural variants resolved by short-read and long-read whole genome sequencing in mendelian disorders. *bioRxiv* [Preprint]. doi: 10.1101/281683

Serrano, M.-L., Sánchez-Gómez, M., Bravo, M.-M., Yakar, S., and LeRoith, D. (2008). Differential expression of IGF-I and insulin receptor isoforms in HPV positive and negative human cervical cancer cell lines. *Horm. Metab. Res.* 40, 661–667. doi: 10.1055/s-0028-1082080

Servant, N., Lajoie, B. R., Nora, E. P., Giorgetti, L., Chen, C. J., Heard, E., et al. (2012). HiTC: exploration of high-throughput "C" experiments. *Bioinformatics* 28, 2843–2844. doi: 10.1093/bioinformatics/bts521

Shin, H., Shi, Y., Dai, C., Tjong, H., Gong, K., Alber, F., et al. (2016). TopDom: an efficient and deterministic method for identifying topological domains in genomes. *Nucleic Acids Res.* 44:e70. doi: 10.1093/nar/gkv1505

Steller, M. A., Delgado, C. H., Bartels, C. J., Woodworth, C. D., and Zou, Z. (1996). Overexpression of the insulin-like growth factor-1 receptor and autocrine stimulation in human cervical cancer cells. *Cancer Res.* 56, 1761–1765.

Tommasino, M., Accardi, R., Caldeira, S., Dong, W., Malanchi, I., Smet, A., et al. (2003). The role of TP53 in cervical carcinogenesis. *Hum. Mutat.* 21, 307–312. doi: 10.1002/humu.10178

Valton, A.-L., and Dekker, J. (2016). TAD disruption as oncogenic driver. *Curr. Opin. Genet. Dev.* 36, 34–40. doi: 10.1016/j.gde.2016.03.008

Wang, K., Li, M., and Hakonarson, H. (2010). Annovar: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38:e164. doi: 10.1093/nar/gkq603

Wang, S., Lee, S., Chu, C., Jain, D., Kerpedjiev, P., Nelson, G. M., et al. (2020). HiNT: a computational method for detecting copy number variations and translocations from Hi-C data. *Genome Biol.* 21:73. doi: 10.1186/s13059-020-01986-5

White, E. A., Kramer, R. E., Tan, M. J. A., Hayes, S. D., Harper, J. W., Howley, P. M., et al. (2012). Comprehensive analysis of host cellular interactions with human papillomavirus E6 proteins identifies new E6 binding partners and reflects viral diversity. *J. Virol.* 86, 13174–13186. doi: 10.1128/jvi.02172-12

Wu, P., Li, T., Li, R., Jia, L., Zhu, P., Liu, Y., et al. (2017). 3D genome of multiple myeloma reveals spatial genome disorganization associated with copy number variations. *Nat. Commun.* 8:1937. doi: 10.1038/s41467-017-01793-w

Yuan, B., Cui, J., Wang, W., and Deng, K. (2016). Gα12/13 signaling promotes cervical cancer invasion through the RhoA/ROCK-JNK signaling axis. *Biochem. Biophys. Res. Commun.* 473, 1240–1246. doi: 10.1016/j.bbrc.2016.04.048

Zhang, X., Zhang, Y., Zhu, X., Purmann, C., Haney, M. S., Ward, T., et al. (2018). Local and global chromatin interactions are altered by large genomic deletions associated with human brain development. *Nat. Commun.* 9:5356. doi: 10.1038/s41467-018-07766-x

Zhao, F. H., Lewkowitz, A. K., Hu, S. Y., Chen, F., Li, L. Y., Zhang, Q. M., et al. (2012). Prevalence of human papillomavirus and cervical intraepithelial neoplasia in China: a pooled analysis of 17 population-based studies. *Int. J. Cancer* 131, 2929–2938. doi: 10.1002/ijc.27571

Chromatin Conformation in Development and Disease

Ilias Boltsis[1], Frank Grosveld[1], Guillaume Giraud[1,2]* and Petros Kolovos[3]*

[1] Department of Cell Biology, Erasmus Medical Centre, Rotterdam, Netherlands, [2] Cancer Research Center of Lyon – INSERM U1052, Lyon, France, [3] Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupolis, Greece

Chromatin domains and loops are important elements of chromatin structure and dynamics, but much remains to be learned about their exact biological role and nature. Topological associated domains and functional loops are key to gene expression and hold the answer to many questions regarding developmental decisions and diseases. Here, we discuss new findings, which have linked chromatin conformation with development, differentiation and diseases and hypothesized on various models while integrating all recent findings on how chromatin architecture affects gene expression during development, evolution and disease.

Keywords: chromatin conformation, TAD, development, differentiation, disease, cancer, gene regulation, regulatory element

INTRODUCTION

All eukaryotic species share the ability to reproduce and transmit their genetic information to their offspring. Mammals originate from single cells, with all the hereditary information stored in the DNA. The 2 meters of chromatin, consisting of DNA plus associated proteins must be compacted to fit in a nucleus with a diameter that varies between 2 and $10\,\mu$m.

The chromatin fiber is a highly dynamic polymer undergoing cycles of de-compaction and re-compaction during the cell cycle and proliferation/differentiation of the cells (Woodcock and Ghosh, 2010). Compaction impacts on chromatin accessibility to transcription factors (TF) and RNA polymerases (RNAPs) and is one of the parameters that fine-tunes the regulation of gene transcription. Thus, different cell fates require a different three-dimensional genome architecture that is closely related to gene expression and cellular function (Dixon et al., 2015). The nuclear genome appears to be organized non-randomly, through a variety of chromatin loops and rosettes and suggests that transcription is also architecturally organized (Lanctôt et al., 2007). Recent data suggest that alterations in chromatin architecture could be causal in diseases and cancer (Spielmann et al., 2018). Here, we describe recent findings about the relation between chromatin conformation and gene regulation in development and diseases and propose a model for chromatin architecture and the formation of loops during development.

HIGH-ORDER CHROMATIN STRUCTURES

Chromosomal Territories

Although the sequence of many genomes has been elucidated, the study of its 3D organization is subject to increasing endeavors using a variety of techniques, most prominent of which are 3C related technologies and high-resolution microscopy. Chromatin is divided into a dark and a light electron-dense region, representing heterochromatin and euchromatin, respectively and gene

activity is related to the position of the genes in the 3D chromatin architecture (Shachar and Misteli, 2017). The sub-nuclear space occupied by a chromosome is called "chromosomal territory" (CT) (**Figures 1A,B**; Cremer and Cremer, 2001; Dixon et al., 2016). On a smaller scale, the genome contains two levels of topological organization, one at a megabase level (A/B compartments) and one at sub-megabase level (topologically associated domains, TADs) (Dixon et al., 2016).

While chromosomes generally reside in distinct territories, CTs sometimes overlap (Branco and Pombo, 2006). These overlapping areas have been suggested to have a functional role in gene regulation suggesting that co-transcription of multi-gene complexes is hierarchical and may require intra and inter-chromosomal interactions (Fanucchi et al., 2013).

Chromosomal Compartmentalization and Its Dynamic Nature

Inside CTs, chromosomes are thought to be divided in two compartments. The large multi-Mb euchromatic A-compartments occupy the internal regions of the nucleus with generally actively transcribed genes, while the heterochromatic B-compartments occupy the periphery of nuclei containing inactive genes (**Figure 1C**; Lieberman-Aiden et al., 2009; Denker and de Laat, 2016; Szabo et al., 2019). However, in some cases, the positions of A and B compartments inside the nucleus are inverted, indicating the dynamic relationship between heterochromatin and euchromatin (Falk et al., 2019).

DNA regions interact more frequently with regions in the same compartment rather than with regions in other compartments (**Figure 1D**; Robson et al., 2019). Every cell type expresses a different set of genes and therefore the content of A/B compartments is cell-type specific. A/B compartments are highly dynamic and change according to the requirements of the cell (Corces et al., 2016; Javierre et al., 2016; Schmitt et al., 2016; Azagra et al., 2020), and the availability of transcription factors and chromatin-modifying enzymes (Therizols et al., 2014; van Steensel and Belmont, 2017), although ~40% has little variability among different human tissues and cell types (Schmitt et al., 2016). Various studies suggest that genes reposition from the periphery to the nuclear interior and vice versa during cell differentiation to activate or repress genes (van Steensel and Furlong, 2019). Such a compartment switch from B to A is observed during T-cell differentiation, where *BCL11B* is activated and the entire locus moves from the periphery to the center of the nucleus (Isoda et al., 2017). Another example is the rearrangement of the *Igh* locus in mice from the periphery to the center of the nucleus during B cell maturation (Kosak et al., 2002). The opposite switch also occurs, e.g., during neuroblast formation in D. melanogaster where the hunchback (*hb*) gene moves to the nuclear lamina (Kohwi et al., 2013). Interestingly, 36% of A/B compartments of human genome switched from an open to closed state and/or *vice versa* during differentiation, while maintaining their TAD boundaries (Dixon et al., 2015). The number of B compartments increases during differentiation from embryonic stem cells to differentiated cells (Xie et al., 2013). Nevertheless, the expression patterns of the majority of genes
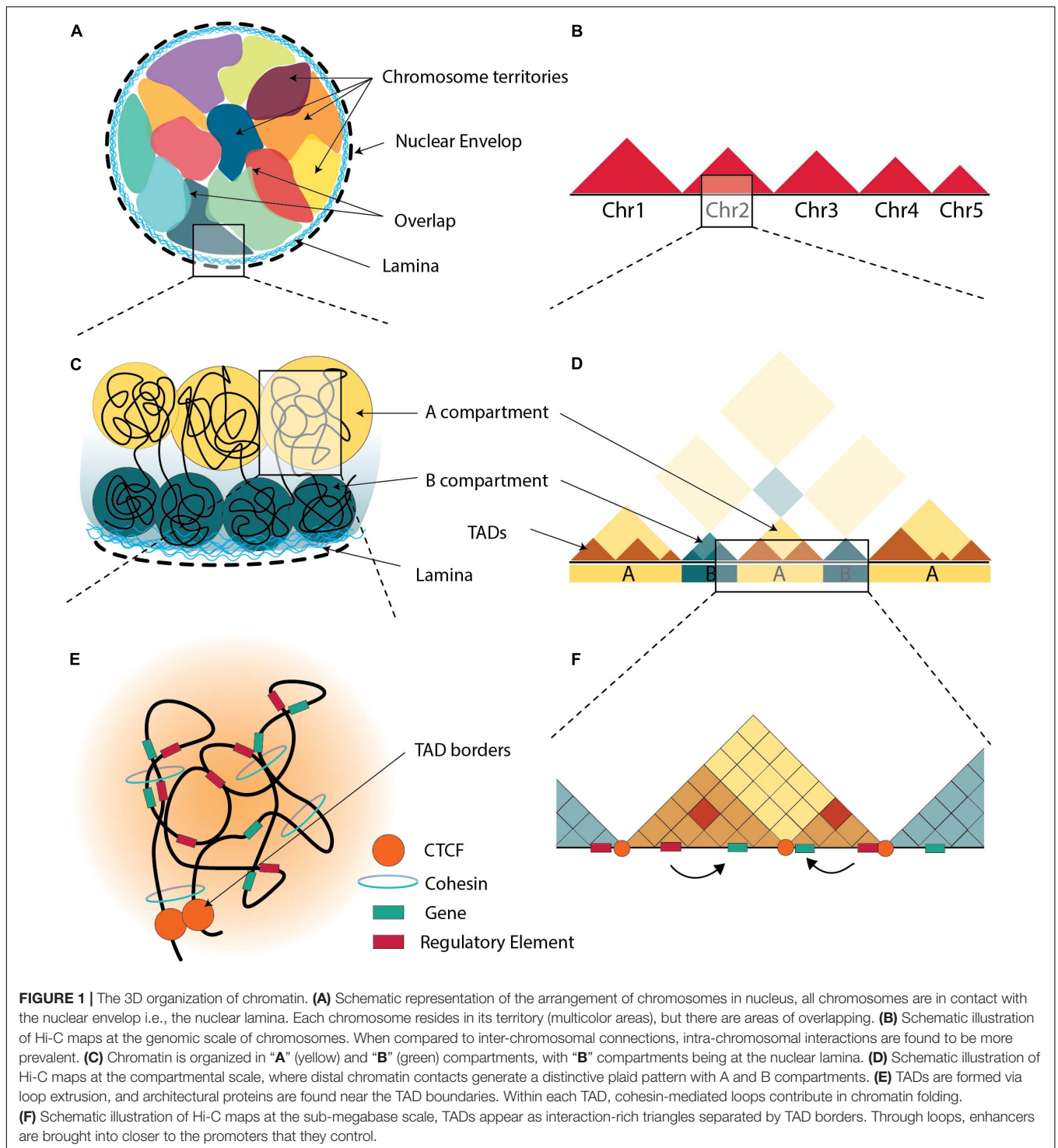
did not change (Dixon et al., 2015). Thus, compartmentalization seems to be dependent on the levels of transcription in a genomic region, and not the expression patterns of each gene (Zheng and Xie, 2019). Recently, an intermediate compartment termed 'I' was identified in maturing B-cells, which contains mainly poised promoters and Polycomb-repressed chromatin states (Vilarrasa-Blasi et al., 2021).

Nature, Topology and Role of Topologically Associated Domains

The second sub-megabase level of topological organization comprises compartments which are organized in self-associating domains and are divided by linker regions. These compartments are called "topologically associated domains" (TADs) (**Figure 1E**; Dixon et al., 2012). This organization facilitates physical contacts between genes and their regulatory elements (Nora et al., 2012; Sexton et al., 2012) and range between 0.2 to 1 Mbp (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012). Contacts between regulatory elements are more frequent inside a particular TAD rather than between two different TADs (**Figure 1F**; Nora et al., 2012).

TADs are highly conserved upon stem cell differentiation, reprogramming, stimulation and in different cell types (Bonev and Cavalli, 2016; Andrey and Mundlos, 2017; Flyamer et al., 2017; Sauerwald and Kingsford, 2018; Zheng and Xie, 2019). Many differentiated cell types contain hundreds of TADs similar to those of human ESCs (Dixon et al., 2015; Schmitt et al., 2016). Thus, TADs are regarded as the basic unit of the folded genome and are considered as structural elements of chromosomal organization (Cremer and Cremer, 2001; Dekker and Heard, 2015; Sexton and Cavalli, 2015). TADs may also not appear as stable structures in single cells, but rather as a mix of chromatin conformations present in a population of cells (Nora et al., 2012; Flyamer et al., 2017; Zheng and Xie, 2019). A multiplexed, super-resolution imaging method identified TAD-like structures in single cells, although these were not stable (Bintu et al., 2018). Similar observations were made even between individual alleles (Finn et al., 2019). Interestingly, a number of studies have indicated that TADs could also be conserved between species (Rudan et al., 2015; Harmston et al., 2017; Krefting et al., 2018), while others come to the conclusion that that TADs certainly have some functional conservation but that specific TAD structures and their location may not be conserved (Eres and Gilad, 2021).This difference in conclusions suggests that the observed various sorts of conservation could be the result of study designs and/or different analytical choices.

As discussed in recent reviews (Sexton and Cavalli, 2015; van Steensel and Furlong, 2019), TADs could affect gene expression in various ways. TADs play an important role in regulation of gene expression by either acting as barriers or by facilitating or preventing loop interactions, because two points (regulatory elements) tethered on a string interact more frequently (**Figure 1F**; Dillon et al., 1997; Lieberman-Aiden et al., 2009; Symmons et al., 2014; Bonev and Cavalli, 2016; Bompadre and Andrey, 2019; Robson et al., 2019; Schoenfelder and Fraser, 2019; Sun et al., 2019). Importantly TADs appear to be lost

FIGURE 1 | The 3D organization of chromatin. **(A)** Schematic representation of the arrangement of chromosomes in nucleus, all chromosomes are in contact with the nuclear envelop i.e., the nuclear lamina. Each chromosome resides in its territory (multicolor areas), but there are areas of overlapping. **(B)** Schematic illustration of Hi-C maps at the genomic scale of chromosomes. When compared to inter-chromosomal connections, intra-chromosomal interactions are found to be more prevalent. **(C)** Chromatin is organized in "**A**" (yellow) and "**B**" (green) compartments, with "**B**" compartments being at the nuclear lamina. **(D)** Schematic illustration of Hi-C maps at the compartmental scale, where distal chromatin contacts generate a distinctive plaid pattern with A and B compartments. **(E)** TADs are formed via loop extrusion, and architectural proteins are found near the TAD boundaries. Within each TAD, cohesin-mediated loops contribute in chromatin folding. **(F)** Schematic illustration of Hi-C maps at the sub-megabase scale, TADs appear as interaction-rich triangles separated by TAD borders. Through loops, enhancers are brought into closer to the promoters that they control.

during mitosis and cell division and to be re-established only after the formation of cis regulatory interactions, which suggests they are not driving but rather maintaining genome structure (Giorgetti et al., 2013; Naumova et al., 2013; Espinola et al., 2021). Disruption of TAD boundaries can nevertheless alter promoter-enhancer interactions by allowing new or preventing normal interactions (Lupiáñez et al., 2015; Flavahan et al., 2016). While

TAD boundaries are generally conserved across cell types, a small fraction exhibit cell-type specificity with changes observed within boundaries during differentiation (Dixon et al., 2012, 2015; Zheng and Xie, 2019). It is worth mentioning here, that the location of boundaries in single-cells varies from cell-to-cell but is always located close to CTCF and cohesin binding sites. Stable TAD boundaries could only be observed in population averaging

studies (Bintu et al., 2018). Changing the enhancer-promoter distance within a TAD has little effect on the gene's expression level (Symmons et al., 2016), unless multiple genes compete for interactions with the enhancer (Dillon et al., 1997). However, inversions, that disrupt the TAD structure alters expression levels (Lupiáñez et al., 2016; Symmons et al., 2016; Robson et al., 2019). Potentially TAD boundaries could be as barriers to prevent the spread of heterochromatin to active regions (and *vice versa*) and/or the spread of proteins tracking on the chromatin (Austenaa et al., 2015; Narendra et al., 2015). One of the main roles of TADs is to provide an insulation for the enhancer-promoter interactions and contain them within the TAD (Dixon et al., 2012; Rao et al., 2014; Zhan et al., 2017; Gong et al., 2018), although there are cases where enhancer-promoter interactions cross over the TAD boundaries, such as in human hematopoietic cells (Javierre et al., 2016) and between Polycomb-bound regions in mouse ESCs (Schoenfelder et al., 2015b; Bonev et al., 2017; Schoenfelder and Fraser, 2019).

The position of TADs in the nucleus relative to each other, or to the nuclear periphery or substructures is under intense investigation. Localization has been proposed to influence gene expression, such as the observation that TADs containing repressed genes at a particular developmental stage are localized at the nuclear lamina (Guelen et al., 2008). Some heterochromatic TADs correspond to lamina associated domains (LADs) or parts of the genome with repressive histone marks (Nora et al., 2012). This agrees with studies suggesting that LADs are poor in genes and that their transcription is suppressed (Lanctôt et al., 2007; Guelen et al., 2008). LAD and heterochromatic TAD regions overlap, albeit incompletely (van Steensel and Belmont, 2017). Euchromatic TADs are transcriptionally active and correspond to regions with active histone marks (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012). Erasing the histone modifications did not affect TAD conformation, possibly because these histone marks are formed in pre-existing TADs (Nora et al., 2012; Dekker and Heard, 2015). LADs and euchromatic TADs are clearly separated by defined borders of CTCF or active promoters (Guelen et al., 2008). Interestingly, in *D. melanogaster*, most of the TAD borders correspond to regions of active promoters rather than CTCF-binding sites (Ramírez et al., 2018). Similar observations were also made in mESCs (Bonev et al., 2017).
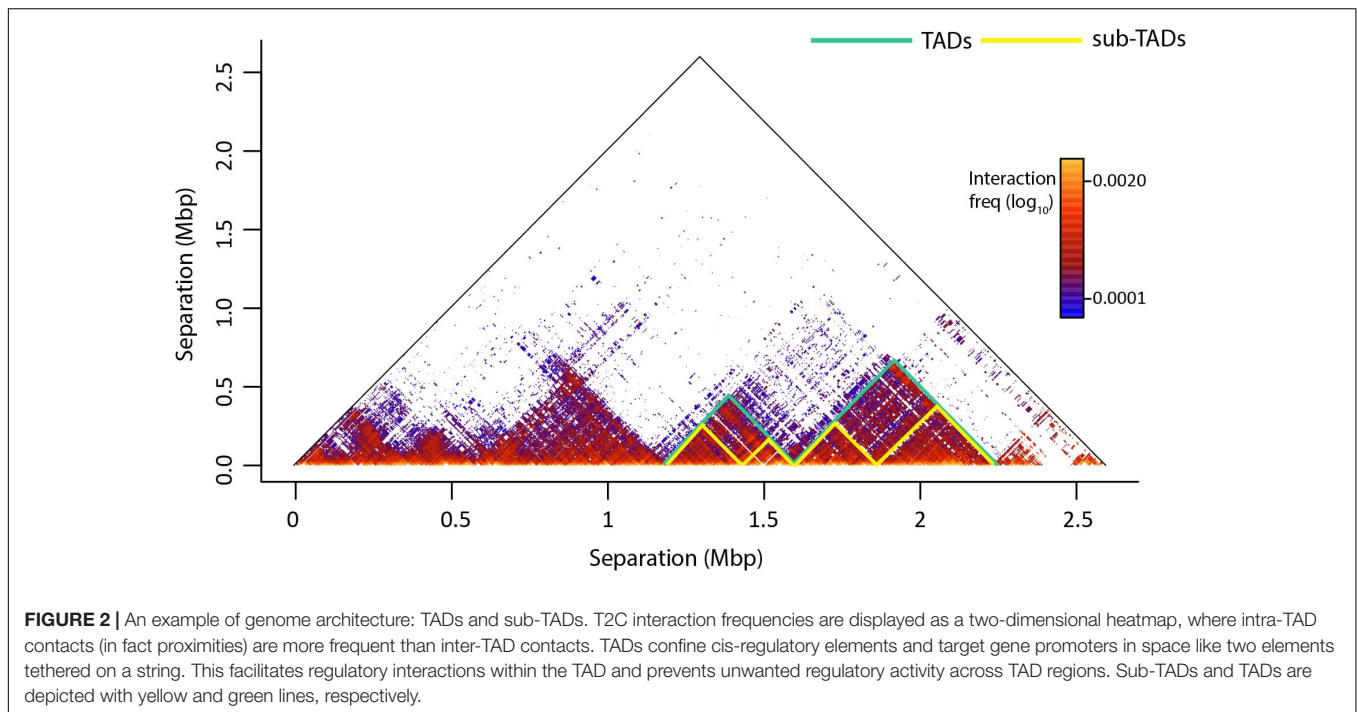
The Important Regulators of Genome Organization

Several key proteins are involved in the establishment of chromatin loops and domains with CTCF and cohesin being among the most studied (Dixon et al., 2012; Rao et al., 2014; Fudenberg et al., 2016; Kim et al., 2019). Proper chromatin interactions require convergent pairs of CTCF bound regions, marking the boundary sites of a TAD (Phillips-Cremins et al., 2013; Zuin et al., 2014; de Wit et al., 2015; Guo et al., 2015; Jia et al., 2020). Inverting or deleting the CTCF sites could affect chromatin conformation, leading to an increase of inter-domain contacts and a decrease of intra-domain contacts (Dixon et al., 2012; de Wit et al., 2015; Hanssen et al., 2017). CTCF is enriched in TAD boundaries (Dixon et al., 2012;

Nora et al., 2012), although its presence is not limited to boundary sites (Zuin et al., 2014). It is also important to note that while CTCF loops define a subset of TADs (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012), not all TADs are surrounded by CTCF sites (Rao et al., 2014). Importantly, CTCF disruption changes TAD structure (de Wit et al., 2015; Guo et al., 2015; Narendra et al., 2015; Sanborn et al., 2015; Nora et al., 2017), while TADs dramatically disappear after depletion of cohesion and compartmentalization is increased (Haarhuis et al., 2017; Rao et al., 2017; Schwarzer et al., 2017; Wutz et al., 2017). Interestingly, these results were corroborated by polymer simulations (Nuebler et al., 2018). Moreover, CTCF interacts with the cohesin complex, which was proposed to organize the genome based on loop extrusion (Fudenberg et al., 2016). It should be noted though that it has not been shown yet that cohesin loops are formed through extrusion *in vivo*. The extrusion mechanism of cohesin is an asymmetric process, which would have certain implications on gene expression. Interestingly recent data indicate that cis regulatory loops are already formed after mitosis before TADs are formed (Espinola et al., 2021).

An example of a topological organization of a locus that could be explained based on the loop extrusion model is that of the α-globin locus (Brown et al., 2018). The self-interacting domain is not present in mES cells, but is formed in differentiating erythroblasts with no apparent change in the binding of CTCF (Brown et al., 2018). Upon perturbations that abolish the expression of α-globin, the domain conformation was unaffected, although interactions within the domain were significantly altered. The convergent pair of CTCF bound regions do not appear as a unique contact, but a broader area of tissue specific contacts was observed around the CTCF borders (Brown et al., 2018). Other mechanisms such as transcription could also lead to loop extrusion. Different cohesin complexes with different subunits (SA1, SA2) seem to act in a different manner mediating different aspects of DNA conformation. SA1-containing complexes promote TAD formation/stabilization while SA2-containing complexes mediate intra-TAD enhancer-promoter contacts (Kojic et al., 2018), suggesting that transcription and transcription factors are important in the formation of those domains. Loop extrusion is also supported by computational modeling (Fudenberg et al., 2016) and also by perturbation assays of important factors of 3D chromatin conformation, such as CTCF and cohesin (Sofueva et al., 2013; Haarhuis et al., 2017; Nora et al., 2017; Rao et al., 2017; Schwarzer et al., 2017; Wutz et al., 2017; Schoenfelder and Fraser, 2019; Thiecke et al., 2020).

DNA is thought to asymmetrically slide through the cohesin ring until it reaches a CTCF site where cohesin is stalled to stabilize the loop (Nuebler et al., 2018). It has been proposed that loop extrusion initiates where cohesin is loaded on DNA through the NIPBL protein. Experiments *in vitro* have shown that human cohesin–NIPBL complexes extrude loops in an ATP-dependent manner (Kim et al., 2019; Golfier et al., 2020).

The removal of NIPBL highlighted two different mechanisms for the genome organization. One is independent of cohesin and organizes the genome into fine-scale compartments (compartmentalization), while the other is dependent on cohesin

FIGURE 2 | An example of genome architecture: TADs and sub-TADs. T2C interaction frequencies are displayed as a two-dimensional heatmap, where intra-TAD contacts (in fact proximities) are more frequent than inter-TAD contacts. TADs confine cis-regulatory elements and target gene promoters in space like two elements tethered on a string. This facilitates regulatory interactions within the TAD and prevents unwanted regulatory activity across TAD regions. Sub-TADs and TADs are depicted with yellow and green lines, respectively.

and contributes to the formation of TADs (Schwarzer et al., 2017; Thiecke et al., 2020). In fact, depletion of CTCF had little effect on A/B compartments, while depletion of cohesin even strengthens it (Nora et al., 2017; Rao et al., 2017; Schwarzer et al., 2017; Wutz et al., 2017; Cremer et al., 2020). This is further supported from experiments where RAD21, a subunit of cohesin complex, was degraded, which disrupted all CTCF loops indicating that CTCF alone cannot stabilize the loops. After restoring RAD21, the majority of CTCF loops appeared within 40 minutes (Fudenberg et al., 2016; Hansen et al., 2017). These findings contradict the hierarchical organization model that suggests that TADs are the compartmental building blocks and suggests that the loop extrusion may change compartmental features (Nuebler et al., 2018). The unloading of cohesin is ensured by other proteins such as WAPL and PDS5 (Wutz et al., 2017). Lack of WAPL contributes to loop collision, with an increase of interactions between distal CTCF sites due to an incremental aggregation of loop domain anchors, and thus, creating a "cohesin traffic jam" (Allahyar et al., 2018). Whether cohesin is "fixed" at CTCF sites remains elusive. It was shown that CTCF and WAPL bind to the same cohesin pocket, with CTCF stabilizing cohesin at TAD boundaries and thus blocking WAPL action (Li et al., 2020). The binding signals at CTCF binding sites are higher than at other position in the genome (Sanborn et al., 2015), but the low general background signal could indicate that cohesin is loaded and extruding continuously and only has a longer dwell time at CTCF sites (Fudenberg et al., 2016, 2017).

CTCF mediated RNA interactions are essential for the proper genome organization (Saldaña-Meyer et al., 2019). Furthermore, many long non-coding RNAs (lncRNAs) have been found to interact with chromatin (Chu et al., 2011; Simon et al., 2011; Engreitz et al., 2013; Li et al., 2017), suggesting that lncRNAs

are involved in structural organization of the genome, like Xist and Firre. During X chromosome inactivation, the lncRNA Xist controls the conformation of the inactive X chromosome (Splinter et al., 2011; Engreitz et al., 2013; Chen et al., 2016), while Firre facilitates the colocalization of genomic regions from different chromosomes (Yang et al., 2015). Moreover, T-cell fate is determined by the lncRNA ThymoD and its role to promote promoter-enhancer interactions (Isoda et al., 2017). Nonetheless, further research is needed in order to conclude, whether lncRNAs play a role in structural organization of the genome.

Higher/Lower Levels of Genome Organization

Topologically associated domains are further divided into smaller organizations, the sub-TADs which have a median size of ∼185 Kbp and are characterized by higher interaction frequencies (**Figure 2**; Rao et al., 2014; Rowley et al., 2017). Sub-TADs should not to be confused with the compartmental domains, which are not formed by CTCF loops but the segregation of A/B compartments (Rowley et al., 2017; Rowley and Corces, 2018). Compartmental domains have been proposed as a model for the organization of chromatin, with architectural proteins and TAD boundaries contributing in the fine-tuning of the transcriptome or regulating a subset of the genes (Stadhouders et al., 2019). On the other hand, a sub-TAD could contain one (or more) gene(s) with its/their regulatory elements, leading to their transcriptional activation or repression (Phillips-Cremins et al., 2013; Rao et al., 2014; Symmons et al., 2014; Bonev et al., 2017). TADs may also contact each other on a higher scale, forming meta-TADs in which inter-TAD interactions are favored (Fraser et al., 2015). sub-TADs and/or meta-TADs exhibit more

tissue specific interaction patterns than the tissue invariant TADs (Dixon et al., 2016; Andrey and Mundlos, 2017).

Other levels of chromatin organization are loop domains and insulated neighborhoods (Rao et al., 2014; Hnisz et al., 2016a; Andrey and Mundlos, 2017). Loop domains are regions with enriched interactions marked by a loop at their border (Rao et al., 2014). A loop domain can represent a whole TAD, but also only a part of it. The current mainstream hierarchical model of chromatin organization promotes, that compartments contain several TADs and subsequently contain several sub-TADs, suggesting that if TADs are the building blocks of the genome, sub-TADs would be the cement holding them together (Bonev and Cavalli, 2016). Insulated neighborhoods are genomic domains, encompassing at least one gene and forming chromatin loops, which are sealed by a CTCF homodimer and co-bound with cohesin (Hnisz et al., 2016a).

Limitations of Methods Unveiling TADs

At present, genome-wide identification of both TADs and sub-TADs relies on the resolution of 3C related technologies and at least 22 different computational methods, contributing to the argument that TADs may not be a "discrete" level of organization of the genome (Fudenberg and Mirny, 2012; Rao et al., 2014; Xu et al., 2020). Nevertheless, genes within the same TAD show similar expression patterns across multiple types of cells and tissues, a trait that is substantially lower at other levels of organizational. This observation favors the role of TADs as a functional level of organization where gene regulation takes place. It is however worrying that different experimental methods result in different estimates of TAD size and numbers (Zufferey et al., 2018), possibly due to the low coverage of the 3C related technologies (Xu et al., 2020) and the different models that each algorithm employs. Adding to this, in single-cell Hi-C experiments, TADs are not reproducibly detected at individual loci, but may be "reassembled" when the individual maps are combined to create a whole population (bulk) experiment (Flyamer et al., 2017). The inherent problem here is that each fragment has only two ends and thus, it could be ligated with four only other fragments. Moreover, contacts are dynamic, created and lost all the time, with TAD borders seeing each other more frequently, strengthens the notion that a TAD is only visible when many cells are analyzed. Thus, the need of improved chromatin conformation capture techniques with increased resolution and coverage as well as algorithms identifying consistently TADs is of prime importance.

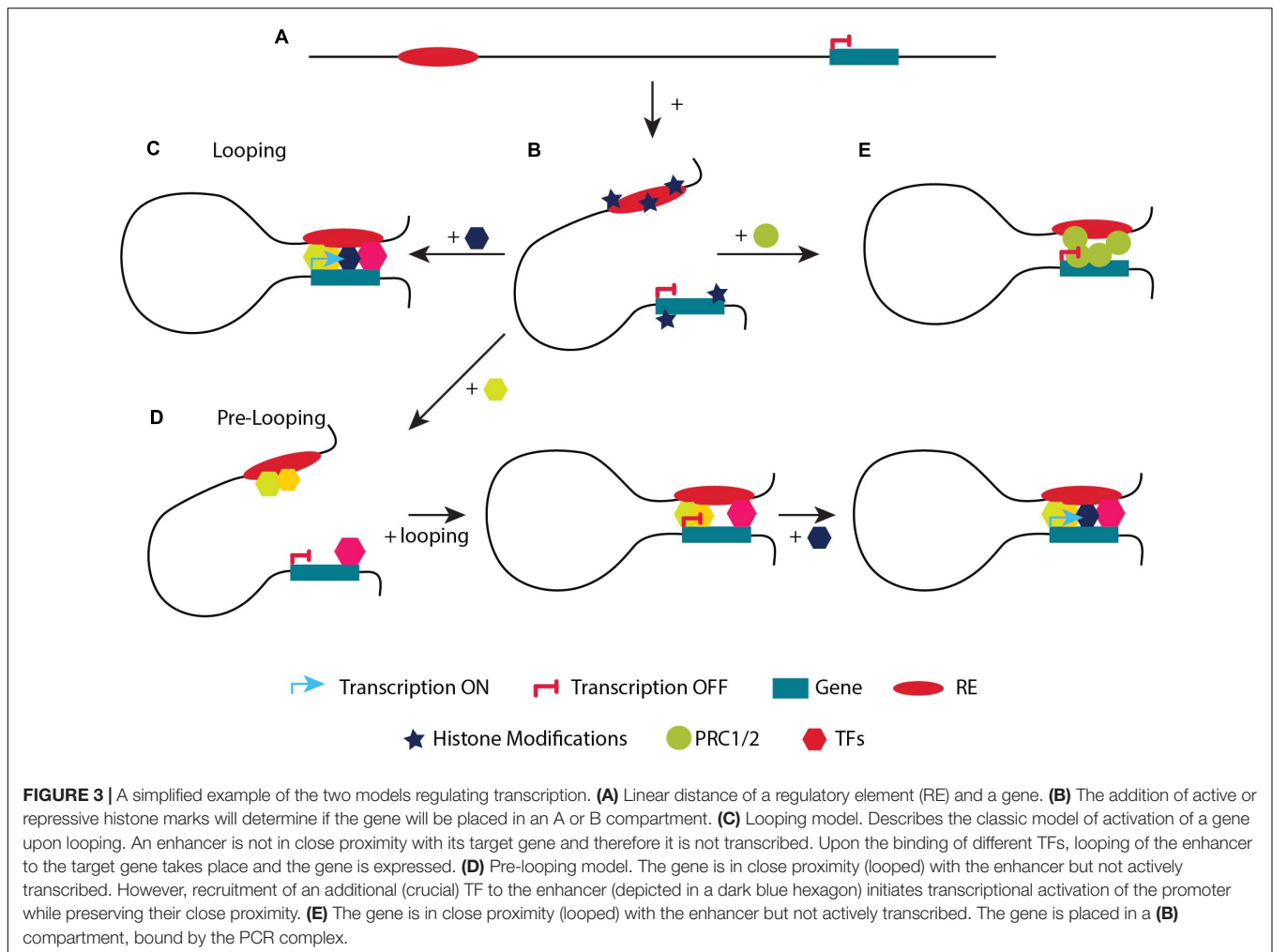## ENHANCER-PROMOTER CONTACTS AS THE DRIVING FORCE OF TRANSCRIPTION

Looping (*de novo* Contacts)

Gene transcription is tightly regulated by regulatory elements (enhancers, insulators, silencers), which can be located at various distances from their cognate gene(s) on the linear DNA strand (**Figure 3A**; Kolovos et al., 2012; Schoenfelder et al., 2015a; Sun et al., 2019). In order to carry out their function, regulatory elements have to be in close proximity to their target gene(s) (Stadhouders et al., 2019). 'Loops' between enhancers and promoters usually result in local interactions, as opposed to CTCF-mediated long-range chromatin loops (TADs), which could facilitate enhancer-promoter interactions either by bringing them closer or by segregating the genome according to its chromatin state (**Figure 3B**; Zheng and Xie, 2019). Recently it was shown that TFs (e.g., YY1 and LDB1), ncRNAs, the Mediator complex, p300 acetyltransferase and the cohesin complex proteins play key roles in the stabilization of chromatin looping or transcription factories (Kagey et al., 2010; Stadhouders et al., 2012; Phillips-Cremins et al., 2013; Fang et al., 2014; Zuin et al., 2014; Schoenfelder et al., 2015a; Boija et al., 2018; Cho et al., 2018; Spielmann et al., 2018; Peñalosa-Ruiz et al., 2019). The function of cohesin varies between various promoter-enhancer interactions. Some promoter–enhancer interactions could also be established only by transcription factors without the involvement of cohesin (Rubin et al., 2017). Four models have been proposed to explain how promoters and enhancers may regulate gene expression with the looping and the transcription factory model being the most prominent (Kolovos et al., 2012; Papantonis and Cook, 2013). Notably, the general notion of the looping model is that an enhancer is in close proximity to its target promoter(s) leading to gene activation, while the gene is silenced when the enhancer and promoter are not in close proximity.

Gene regulation from distal regulatory elements through local looping is now a commonly accepted concept (Lupiáñez et al., 2015; Flavahan et al., 2016; Hnisz et al., 2016a; Bonev et al., 2017; Stadhouders et al., 2018). Before the development of chromosome conformation capture technologies, which are essentially biochemical techniques, there was already strong evidence from biochemical and genetic type experiments that loop formation mediates transcription in both prokaryotic and eukaryotic systems. That was depicted *in vitro* with the lac repressor system (Hochschild and Ptashne, 1986). In eukaryotic systems, *in vitro* assays using a plasmid suggested that an enhancer and a gene could be separated by a protein bridge invoking looping (Müeller-Storm et al., 1989). Strong evidence in eukaryotes, with genes in the normal genome environment, was obtained at the β-*globin* locus after discovery of the Locus Control Region (LCR, (now called super-enhancers), which is located 70 kb upstream of the β-*globin* gene(s). Changing the distance or order of the β-*globin* genes and the LCR could only be explained by looping (Grosveld et al., 1987; Hanscombe et al., 1991; Dillon et al., 1997). A few years later, the effect of natural mutations by defective enhancers located at very long distance, like in the case of polydactyly, was very difficult if not impossible to explain by mechanisms other than looping (Lettice et al., 2003).

The regulation of the β-*globin* like genes by its LCR, was and still is the best-studied example for the looping model (**Figure 3C**; Grosveld et al., 1987). In adults, the LCR and the β-*globin* promoter are located in close proximity contributing to the formation of new chromatin loops by the recruitment of specific TFs such as LDB1, TAL1, GATA1 and KLF1 to the LCR (Noordermeer and de Laat, 2008; Palstra et al., 2008a).

FIGURE 3 | A simplified example of the two models regulating transcription. **(A)** Linear distance of a regulatory element (RE) and a gene. **(B)** The addition of active or repressive histone marks will determine if the gene will be placed in an A or B compartment. **(C)** Looping model. Describes the classic model of activation of a gene upon looping. An enhancer is not in close proximity with its target gene and therefore it is not transcribed. Upon the binding of different TFs, looping of the enhancer to the target gene takes place and the gene is expressed. **(D)** Pre-looping model. The gene is in close proximity (looped) with the enhancer but not actively transcribed. However, recruitment of an additional (crucial) TF to the enhancer (depicted in a dark blue hexagon) initiates transcriptional activation of the promoter while preserving their close proximity. **(E)** The gene is in close proximity (looped) with the enhancer but not actively transcribed. The gene is placed in a **(B)** compartment, bound by the PCR complex.

The different enhancer elements and the gene appear to form a regulatory hub where all the different elements appear to interact with each other (Allahyar et al., 2018). Interestingly, even though the individual enhancers appear to interact, the overall activity of the LCR usually appears to be the result of an additive effect of the individual enhancer elements rather than a synergizing effect, with the individual enhancers exhibiting different properties (Fraser et al., 1993; Bender et al., 2012). Absence of crucial TFs in the LCR results in the disruption of chromatin conformation and in gene mis-expression.

Recent allele specific interaction studies indicate that the LCR interacts with more than one of the (mouse) β-*globin* genes simultaneously (Allahyar et al., 2018), whereas previous studies showed that only the (human) β-*globin* gene can be active at any given moment in time in the situation where two genes are in contact with the LCR at the same time (the γ- and β-*globin* genes in human and the β*maj*- and β*minor*-globin in mouse) (Wijgerde et al., 1995; Trimborn et al., 1999). These observations lead to the conclusion that transcription is a discontinuous process and that the frequency and stability of the promoter-enhancer interactions is a very important parameter in determining the level of transcription. The observation that the mouse LCR would

interact with two β-*globin* genes simultaneously, but that only one would be expressed, sets up the interesting question whether this is perhaps particularly prevalent among genes "competing" for the same enhancers.

Looping interactions are not limited only to enhancers and promoters. Subsequent studies suggest that enhancers make contacts also with gene bodies following the elongating RNAPII (Lee et al., 2015). In parallel, Polycomb proteins (PRC1, PRC2) facilitate the regulatory topology by repressing genes through chromatin interactions and keep them under tight control (Schoenfelder et al., 2015b; Cruz-Molina et al., 2017; Cai et al., 2021). Moreover, some promoters (E-promoters) can act as bona-fide enhancers and are in close proximity with others to activate gene expression (Dao et al., 2017).

An interesting debate is whether gene activation precedes locus conformation or *vice versa* (van Steensel and Furlong, 2019). In a previous study, during neuronal differentiation, promoter–enhancer interactions appeared along with changes in gene expression (Bonev et al., 2017). However, during erythropoiesis, chromatin structure precedes expression and does not require the presence of TFs, but TFs are essential for the advancement to, or maintenance of, a fully functioning active

chromatin hub (Drissen et al., 2004). Moreover, chromatin loops are not altered in the β-*globin* locus upon transcriptional inhibition, suggesting that structure precedes function (Palstra et al., 2008b). Interestingly, the recruitment of LDB1 to the β-*globin* promoter depends on GATA1, in contrast to its recruitment to LCR. In GATA1-null cells that do not express β-*globin*, its expression can be rescued by the tethering of LDB1 via a zinc finger domain to its promoter, mediating its interaction with the LCR, and thus supporting the hypothesis that conformation comes first (Deng et al., 2012). In another study LDB1 was directed to the silenced promoter of the embryonic β-*like globin* (β*h1*) gene in adult mice erythroblasts (Deng et al., 2014). In parallel, during the zygotic genome activation, the formation of TADs coincides with the onset of gene expression (Hug et al., 2017).

Pre-looping (Pre-determined Contacts)

Recent studies propose an additional way on how chromatin conformation controls gene transcription. Genes are often in close proximity to their cognate enhancers without being actively transcribed. Although their cognate enhancer is often bound by various TFs, it lacks the binding of a crucial TF required for gene activation (Kolovos et al., 2016). At the same time, RNAPII is stalled at the promoter (Ghavi-Helm et al., 2014). In that case, when a developmental or a differentiation signal triggers the additional recruitment of crucial TF(s) to the enhancer, looping is maintained and transcription is induced. This model is termed pre-looping (**Figure 3D**; Ghavi-Helm et al., 2014; Kolovos et al., 2016; Rubin et al., 2017). During mouse development, pre-existing chromatin contacts of the Hox genes could help in the recruitment of the necessary transcription factors, in order tissue-specific promoter-enhancer interactions to occur (Lonfat et al., 2014). Moreover, loops mediated by the PRC1 and PRC2 complexes in pluripotent cells are not only repressing the genes inside such loops, but also maintain them in close proximity with their regulatory elements permitting a fast response (activation) to specific differentiation signals (**Figure 3E**; Schoenfelder et al., 2015b; Cruz-Molina et al., 2017). Similarly, the CTCF and cohesin complex bring the *Shh* promoter and ZRS enhancer in close proximity in posterior and anterior limbs (Paliou et al., 2019). Although they are in close proximity, the *Shh* gene is differentially expressed in these tissues (Williamson et al., 2016). An even closer proximity is observed when *Shh* is activated in the posterior limbs (Williamson et al., 2016). As it is clear from the previous examples, specific topological features are not a sufficient criterium to initiate transcription (Ghavi-Helm et al., 2014; Hug et al., 2017).

Most of the interactions of the pre-looping model are not mediated or predicted by CTCF, but by TFs and RNAPII, e.g., in HUVEC cells, *SAMD4A* is not expressed while its promoter is in close proximity with its enhancers. Upon activation by TNFα signaling, the TF NFκB is released from the cytoplasm, enters the nucleus and binds to the enhancer leading to looping maintenance and the activation of *SAMD4A* expression (Kolovos et al., 2016). Other examples of pre-looping were later reported in macrophages, upon adipogenesis, differentiation of the epidermis, during differentiation of mouse embryonic stem

cells (ESCs) to neural progenitors, in the mouse *HoxB* locus and in hypoxia, but also as a mechanism of action for specific transcription factors like PAX5 (Barbieri et al., 2017; Cruz-Molina et al., 2017; Rubin et al., 2017; Siersbæk et al., 2017).

Thus, there is an interesting conundrum. How could transcription be controlled by two different chromatin conformations; looping and pre-looping. According to the pre-looping model, loops formed by CTCF, cohesin, PRC1 or PRC2 could contain poised enhancers and promoters in close proximity only to activate them with subsequent tighter contacts, e.g., after post translational modifications of essential for activation TFs take place (**Figures 3D,E**; Drissen et al., 2004; Robson et al., 2019). According to the looping model, loops appear and disappear dynamically during development, in parallel with transcriptional activation and could flexibly fine-tune transcription (Javierre et al., 2016; Bonev et al., 2017). Another explanation could be that most of the looping paradigms are studied in steady-state systems or when comparing only two stages of differentiation or development (Grosveld et al., 1987; Palstra et al., 2008a; Deng et al., 2012; Kolovos et al., 2014). Maybe some genes have been selected evolutionary to use one of the two ways of chromatin conformation. However, studying more than two stages of differentiation, development or embryogenesis could unveil which of the two mechanisms is used mostly (Stadhouders et al., 2018; Di Stefano et al., 2020). Although the dynamics of nuclear organization have been studied so far during mitosis (Naumova et al., 2013), meiosis (Patel et al., 2019), hormone treatment, differentiation (Bonev et al., 2017) and cell reprogramming (Stadhouders et al., 2018), there is an immediate need for methods that are precisely tailored for the study of time-dependent conformational changes (4D) (Di Stefano et al., 2020).

TRANSCRIPTION FACTORIES AS THE DRIVING FORCE OF TRANSCRIPTION

The established transcription model claims that the polymerase moves along the DNA sequence to produce the transcript. Nowadays, it is believed that transcription takes place in nucleoplasmic hot spots (called "transcription factories" Papantonis and Cook, 2013, see above), mediated by a high local concentration of all the necessary transcription factors. This notion suggests that the polymerase is located primarily, but not fixed in "transcription factories" (Ghamari et al., 2013; Papantonis and Cook, 2013). In the traditional model of transcription, RNAPII leaves the promoter and moves along the DNA template. In "transcription factories", the RNAPII is present these nucleoplasmic hotspots, while genes and their respective promoters diffuse to them, as transcription takes place through the movement of the DNA template via transcription factories (Jackson et al., 1981; Iborra et al., 1996; Papantonis et al., 2010; Cho et al., 2018). Notably a similar type of mechanism/principle has been proposed for "loop extrusion", the mechanism by which loops are formed and where the DNA moves through the cohesion complex (see above). Time course experiments indicated that the enhancer and the promoter of the *Cd47* and

Kit genes are in close proximity during transcription (Lee et al., 2015). "Transcription factories" are most likely a collection of several "active chromatin hubs," that merge in a phase transition type process containing several polymerase complexes, each transcribing a different template (de Laat and Grosveld, 2003; Larson et al., 2017).

Current interests are focused on liquid-liquid phase separation (LLPS) as the driving force to concentrate the necessary elements (e.g., enhancers, transcription factors, RNAPII, etc.) at active chromatin hubs or transcription factories (Sabari et al., 2018; Guo et al., 2019; Nair et al., 2019). The concept of phase transition, LLPS is a mechanism to generate "structures" without membranes (Hildebrand and Dekker, 2020). Molecular seeds are thought to start the process of phase transition leading to a local enrichment of protein-protein complexes. Intrinsically disordered protein domains are thought to play a major role by their ability to have multivalent interactions (multi-modular features) (Li P. et al., 2012). It has been shown that artificial condensates are able to physically pull together specific loci, and thus, LLPS generate mechanical force to the chromatin (Shin et al., 2018). Such compartmentalized hydrogel-like states would have a reduced fluidity and movement of proteins, which would for example fit with the concept that the DNA moves through the polymerase in a transcription factory rather than the polymerase moving along the DNA. Subsequent research has revealed that the Mediator complex, along with other transcription factors, coactivators, and RNAPII, form condensates during transcription (Boija et al., 2018; Cho et al., 2018; Chong et al., 2018; Sabari et al., 2018; Guo et al., 2019). Phase-separated HP1α and RNAPII showed the ability to create phase-separated heterochromatin and euchromatin droplets, respectively (Larson et al., 2017; Lu et al., 2018). Condensation of bound TFs and coactivators is induced by multivalent enhancer sequences via LLPS (Shrinivas et al., 2019). Although this idea has not been thoroughly tested, it has been observed that LLPS causes enhancers that would typically dwell in distant TADs to migrate closer (Nair et al., 2019). The local concentration of RNA can impact condensate formation and dispersion, acting as a transcriptional feedback mechanism (Henninger et al., 2021).

It has also been proposed that the outer edge of phase-separation droplets acts as a barrier that proteins could not pass through (Strom et al., 2017), despite the quick recovery of CDK9-mCherry signal after photo-bleaching, suggesting that CDK9-mCherry is constantly recruited to the stably positioned transcription factories (Ghamari et al., 2013), Chromatin compartmentalization might be the reason that activating transcription factors are not present in B compartments (Laghmach and Potoyan, 2021). Phase separation could explain several confusing observations, like how transcriptional activation occurs without direct physical contact between enhancers and promoters through eRNAs (Cai et al., 2020), or multi-enhancer and multi-promoter contacts (Li G. et al., 2012; Jin et al., 2013), or simultaneous regulation of more than one gene by a single enhancer (Fukaya et al., 2016). In parallel, recent data suggest that forces other than the ones derived from LLPS could also stabilize transcription factories (Ulianov et al., 2021).

The β-*globin* active chromatin hub, containing *Hbb-b1*, its LCR (60 Kbp upstream of *Hbb-b1*) and *Eraf* (encoding an α-globin stabilizing protein, located ~25 Mbp far from *Hbb-b1*) is the best example of different genes in the same transcription factory. Various assays, like 3C-like methods, RNA and DNA FISH coupled to immuno-labeling, confirmed that *Hbb-b1*, its LCR and *Eraf* are found together in sites rich with RNAPII (Bender et al., 2012; Mitchell et al., 2012). As mentioned above, another property of transcription factories is that they encompass groups of genes (located *in cis* or *in trans*), which are co-regulated by specific signaling pathways or activators leading to the idea that co-regulated genes are expressed in "specialized" transcription factories (Schoenfelder et al., 2010). This is corroborated by ChIA-PET of active RNAPII, which uncovered spatial associations between co-regulated and co-transcribed genes in response to various stimulations (Papantonis et al., 2012; Li et al., 2015). Moreover, RNAPII transcribed genes are located in separate factories than RNAPIII genes. TNFα responsive genes and erythropoietic genes are also located in distinct factories (Pombo et al., 1999; Papantonis et al., 2010; Schoenfelder et al., 2010; Baù et al., 2011; Monahan et al., 2019). Therefore, it is tempting to conclude that there are "specialized" transcription factories.

THE INTERPLAY BETWEEN STRUCTURE AND FUNCTION THROUGH DEVELOPMENT, DIFFERENTIATION AND EVOLUTION

Loops within the genome can be separated into two categories according to their role (Kolovos et al., 2014); structural and functional. Structural loops are forming the building blocks of the 3D conformation of the genome. They can take place between DNA segments (none of which is a promoter or an enhancer) through CTCF or cohesin binding, forming large TAD domains with their base defining the domain boundaries (Dixon et al., 2012; Rao et al., 2014; Zuin et al., 2014). Various chromatin conformation capture results suggest that these structural loops are the same between different cell types (Dixon et al., 2015; Schmitt et al., 2016). Therefore, structural loops could contribute indirectly to the regulation of gene expression, via the formation of TADs confining genes and their respective regulatory elements in a dedicated 3D nuclear space. Functional loops, which often appear within structural loops, are the ones bestowing a function/task (activation/repression/poised) to a gene and often correspond to sub-TADs (Grosveld et al., 1987; Splinter et al., 2006; Palstra et al., 2008a; Wendt et al., 2008; Kagey et al., 2010; Schoenfelder et al., 2010; Kolovos et al., 2012; Phillips-Cremins et al., 2013; Sofueva et al., 2013; Fang et al., 2014; Ghavi-Helm et al., 2014; Rao et al., 2014; Zuin et al., 2014; Ji et al., 2016; Kolovos et al., 2016; Phanstiel et al., 2017; Rubin et al., 2017). These interactions could be direct or indirect. The direct interaction is between two DNA segments with one containing a regulatory element (an enhancer or a silencer) and the other the promoter of the target gene

(de Laat and Grosveld, 2003; Palstra et al., 2008a; Stadhouders et al., 2012; Kolovos et al., 2014; Kolovos et al., 2016). The indirect interaction is between an enhancer/silencer and a DNA segment which is not the promoter of the target gene, which subsequently interacts with the promoter creating an active regulatory hub (Stadhouders et al., 2012; Schuijers et al., 2018; Quinodoz et al., 2018). An example is the *Myc* loci where its super-enhancer interacts with its promoter through a CTCF site located 2 Kbp upstream of the *Myc* promoter (Schuijers et al., 2018), similar to the way *Myb* is regulated in mouse erythroid cells (Stadhouders et al., 2012).

In this part, we describe how functional and structural loops are formed, as well as the shape of the 3D chromatin organization at different stages of development and differentiation (**Figure 4**). As already mentioned before, loops are critical for proper gene expression and the integrity of these loops is indispensable for the development of various tissues, differentiation of cells, diseases and cancer. Hence, it is important to understand how or even when they are formed in order to decipher how the local chromatin architecture contributes to different phenotypes.

The chromatin architecture changes significantly during gametogenesis and early embryonic development (Li et al., 2019; Zheng and Xie, 2019). In short, during spermatogenesis A/B compartments and TADs vanish in pachytene spermatocyte and then reappear in round spermatid and mature sperm stages (Wang et al., 2019).The transcriptionally inactive mouse sperm displays chromatin conformation features, with CTCF and cohesin occupying positions similar to those in mESCs, implying the important role of these factors in shaping chromatin conformation even in the absence of transcription (Carone et al., 2014; Du et al., 2017; Jung et al., 2017). Those features, albeit weaker, were also detectable in oocytes (Gassler et al., 2017). During oogenesis, the oocyte shows the typical higher-order structures until the germinal vesicle (GV) stage (Flyamer et al., 2017). The strength of those features declines dramatically from the immature oocytes to mature oocytes (Flyamer et al., 2017), and from this point forward oocytes lack the typical interphase chromatin structures (Du et al., 2017; Ke et al., 2017). Chromatin structure at this point resembles the chromatin structure during mitosis (Naumova et al., 2013).

After fertilization, chromatin conformation undergoes dramatic reprogramming (Zheng and Xie, 2019). Since TADs and A/B compartments are very weak in early-stage mouse embryos, some studies have shown that chromatin adopts a more relaxed state (Du et al., 2017; Ke et al., 2017). However, loops and TADs have also been observed in mouse zygotes (Gassler et al., 2017). Indeed, TADs are maintained during the oocyte-to-zygote transition in mice and gradually become more prominent (Du et al., 2017; Ke et al., 2017). Genes are initially silenced, but after the zygotic genome activation (ZGA), they are activated (Clift and Schuh, 2013). ZGA occurs in the 2-cell embryo in the mouse (Du et al., 2017; Ke et al., 2017). Inhibition of ZGA did not prevent the formation of TADs (Ke et al., 2017), suggesting that TAD formation precedes their main function of transcriptional control (Flyamer et al., 2017; Hug et al., 2017; Ing-Simmons et al., 2021). Thus, TADs act first as building blocks of architecture and then as transcriptional controllers. TADs are

established in Drosophila during ZGA. Compartmentalization of the chromosomes at the zygote stage seems to be driven by a different mechanism than the one of TAD formation (Flyamer et al., 2017; Ke et al., 2017). Specifically, the paternal originated chromosomes maintain all the genome structures, whereas the maternal chromosomes lose the A/B compartments. During the two-to-eight-cell stage, conformation is slowly re-established and become progressively stronger in both, maternal and paternal chromosomes (Du et al., 2017; Flyamer et al., 2017; Gassler et al., 2017; Ke et al., 2017).

Common TADs and A/B compartments that correspond to transcriptionally active regions are present in both pluripotent cells and differentiated cells, but the chromatin of pluripotent cells is less compacted than in other cell types (Melcer and Meshorer, 2010; Gaspar-Maia et al., 2011). In pluripotent cells, pluripotency TFs are found in the same areas of the nucleus, establishing long-range chromatin interactions with each other (Bouwman and de Laat, 2015). The observation that gene loci controlled by pluripotency factors are located in close proximity inside the nucleus, suggests a regulatory mechanism similar to phase separation (De Wit et al., 2013). For example, it was shown that many KLF4-bound regions are in close proximity to each other in pluripotent cells and released upon differentiation or KLF4 depletion (Wei et al., 2013).

Early in differentiation, pluripotent genes are initially repressed and subsequently activated (Phillips-Cremins et al., 2013). Early differentiation genes exhibit a permissive architecture and are in close proximity to their associated poised enhancers. Upon differentiation, their enhancers become active and activate their target gene(s) (Cruz-Molina et al., 2017). This suggests that conformation structures mediated by Polycomb proteins create a permissive regulatory environment, where poised regulatory elements are ready to be expressed (Cruz-Molina et al., 2017). Similar observations have been also made in other differentiation pathways, such as adipogenesis (Siersbæk et al., 2017).

An intriguing question is how regulatory elements are generated during evolution, because it is clear that a gene can use different regulatory elements in different cell types or during differentiation to more mature cell types. Interestingly, neocortical enhancers start out as basic proto-enhancers and evolve in complexity and size over time (Emera et al., 2016). Moreover, the rapid evolution of enhancers in liver across 20 mammalian species (*18 placental species from Primates, Rodents, Ungulates, Carnivores and 2 marsupial species*) is a general feature of mammalian genome as observed by profiling genomic enrichment of H3K27ac and H3K4me3 of liver enhancer regions (Villar et al., 2015). Interestingly, the majority of the recently evolved enhancers are derived from ancestral DNA exaptation and are significantly over-represented in the vicinity of positively selected genes in a species-specific manner (Villar et al., 2015). Thus, it would be tempting to speculate that species, which were less "evolved", have developed "simpler" regulatory elements to control their gene expression. Since these species were more primitive, gene expression profiles were less complicated and more specific for each of the much smaller number of different cell types. During evolution and the appearance of
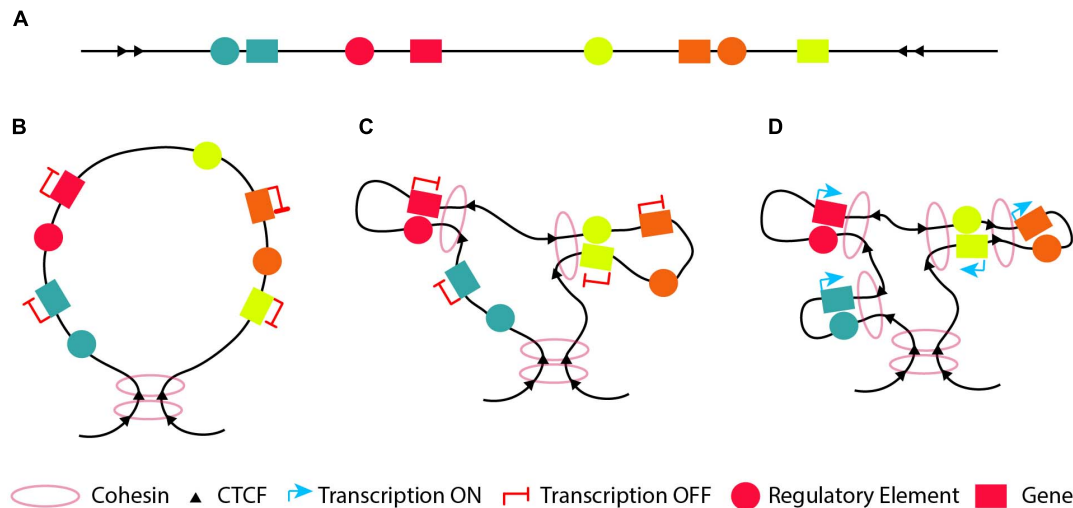
FIGURE 4 | The "loop within loops" model. **(A)** An example of a DNA segment which contains four genes (blue, red, green and orange) depicted with boxes and their cognate regulatory elements (circles with the respective colors). **(B)** A structural loop of 1-2Mb forms a TAD with its base to define the domain boundary. **(C)** At an early developmental/differentiation stage most genes are silenced. Thus, inside the structural loop, the genes will either not form any loops with their cognate regulatory element (looping model; blue and orange genes) or form functional silencing-loops within structural-loops (loops-within-loops) with their cognate regulatory element lacking a crucial TF (pre-looping model; red and green genes). **(D)** At later developmental stages, new functional loops are formed within the pre-existing functional loops (orange gene) and/or the structural loop (blue gene) forming "loops within loops" in order to activate the orange and blue genes, respectively. At the same time, the previously pre-looped genes (red and green) are activated as a result of a recruitment of the necessary TF to their cognate enhancer or due to conversion of their cognate poised enhancer to an active one.

more complex organisms that require an increased diversity of cell composition, the control of gene expression became more complex and new regulatory elements appeared (Ong and Corces, 2011).

Thus, during the early stage(s) of development, differentiation or evolution, a DNA segment with various genes and regulatory elements (**Figure 4A**) will mostly form structural loops to shape the chromatin (**Figure 4B**), since chromatin conformation during ZGA is independent of activation of gene expression (Hug et al., 2017; Ing-Simmons et al., 2021). At an early developmental/differentiation stage or during the oocyte-to-zygote transition, genes are often silenced. Based on the pre-looping model, some genes will already be in close proximity with their enhancer, which lacks one or more necessary TFs and is in a poised state (**Figure 4C**, red and yellow genes and their respective regulatory elements) to promote their activation or silencing, forming functional "loops-within-loops". In parallel based on the looping model, the genes will be far apart from their cognate enhancer in the 3D space (**Figure 4C**, green and orange genes and their respective regulatory elements). At a later developmental/differentiation stages, genes which do not have a poised functional loop, will have to form new functional loops within the pre-existing structural or silencing-functional loops ("loops-within-loops") in order to become transcriptionally active (**Figure 4D** green and orange genes and their respective regulatory elements). At the same time, the previously silenced genes in a poised loop (**Figure 4D**, red and yellow genes) are activated as a result of a recruitment of the necessary TF to their cognate enhancer or due to conversion of their cognate poised enhancer to an active one.

In this context, at initial stages of development, differentiation or evolution, we speculate that the genome must have an initially regulatory element located at a distance from its target gene (**Figure 5A**, green regulatory element), which interacts with its target gene via a specific loop (**Figure 5B**). At a subsequent developmental, differentiation or evolutionary stage, we hypothesize that new (cell/tissue type specific) regulatory elements are developed between the gene and its original "early" regulatory element, which can interact with their target gene (**Figure 5C**, orange and red regulatory elements). Thus, we could observe an initial big loop, which can be functional or either structural, containing other loops at later stages. The latter will form new "loops-within-loops" to accommodate new expression patterns. This type of regulation is observed when comparing the activity of different regulatory elements in multiple stages of differentiation/development/evolution (de Laat and Grosveld, 2003; Palstra et al., 2008a; Mylona et al., 2013; Pimkin et al., 2014; Villar et al., 2015; Goode et al., 2016). However, we cannot exclude the possibility that in rare cases during development, differentiation or evolution, a regulatory element outside the original "early" loop would develop, which could also interact with its target gene at subsequent stage (**Figure 5D**, yellow regulatory element). Finally, all these aforementioned interactions could satisfy either the pre-looping or the looping model (**Figure 3**). In an evolutionary sense, developing novel enhancers is an almost inevitable feature of multicellular organisms with different cell types and functions. Other mechanisms are very difficult to envision for the enormous diversity in gene expression patterns, which is ultimately due to the fact that DNA is a linear molecule.
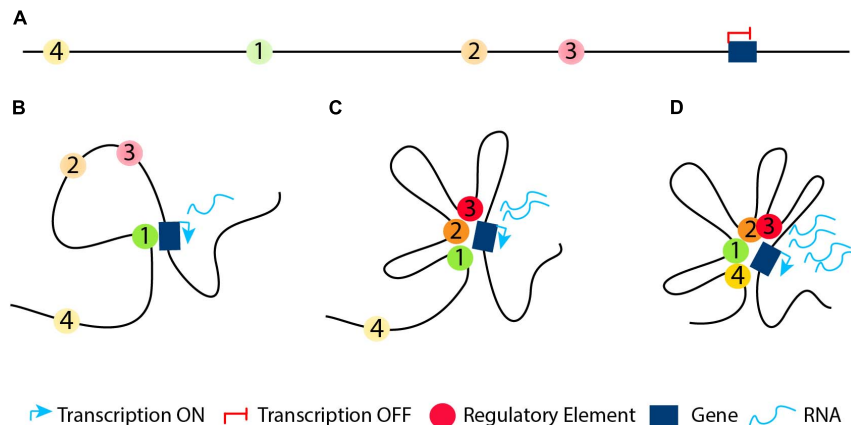
FIGURE 5 | Regulatory elements in development, differentiation or evolution. **(A)** At an initial stage in development, differentiation or evolution, the genome has a silenced gene and a regulatory element located in a distance from it (regulatory element 1, depicted with a green circle). **(B)** The original "early" regulatory element interacts with its target gene in order to activate it. **(C)** At subsequent developmental, differentiation or evolutional stages, the genome could develop new regulatory elements (regulatory elements 2 and 3, depicted with orange and red circle, respectively) between the gene and its original "early" regulatory element, which interact with their target gene. **(D)** At some cases, at later developmental, differentiation or evolutional stages, we could observe a new regulatory element (regulatory element 4, depicted with a yellow circle) outside the original "early" loop, which could also interact with its target gene via a formation of a new loop.

CHROMATIN CONFORMATION FROM EARLY DEVELOPMENT TO DIFFERENTIATION

The internal structure of TADs becomes more organized during development and differentiation, as TADs enable more enhancer-promoter contacts (Bonev and Cavalli, 2016). This is important during development, where specific genes need to be activated or repressed to promote specific cell programs and lineage commitment. For example, during limb development, the *HoxD* cluster is located at the border of two flanking regulatory elements, which are contained into two separated TADs (Lonfat and Duboule, 2015). In the beginning, the 3′ TAD is active and regulates the proximal patterning. Subsequently, this TAD is switched off while the 5′ TAD becomes active and controls distal structure. Activation of *Hox13* switches off the 3′ TAD through a global repressive mechanism and interacts with enhancers at the 5′ TAD that sustains its activity (Beccari et al., 2016). Thus, the *HoxD* cluster contains a dynamic TAD boundary, regulating the switching between the flanking TADs and enabling a proper limb development (Rodríguez-Carballo et al., 2017).

Early studies before the discovery of TADs, showed that the lack of CTCF or its disruption on one of the binding sites in the mouse β-*globin* locus resulted in an altered interactome (Splinter et al., 2006). New insights in the significance of TADs during development came from a study of the *HOXA* locus, which is important for development of many tissues such as limb. The *HOXA* locus is organized in two different TADs, with CTCF and cohesin binding sites at their boundary. The disruption of CTCF or Cohesin recruitment at the boundary sites of these two TADs allows the spreading of euchromatin into heterochromatin domains and the subsequent ectopic activation of *HOX* genes during cell differentiation due to new chromatin contacts (Zuin et al., 2014; Lonfat and Duboule, 2015). Another example is the

Tfap2c and *Bmp7* locus, which is split into two functional and structural domains, with each gene being present in separate TADs with their cognate enhancers. Inversions at the TAD boundary, changes the position of *Bmp7*'s cognate enhancer into the TAD containing *Tfap2c*, thus leading to upregulation of the latter gene and downregulation of Bmp7 (Tsujimura et al., 2015). This illustrates the extent to which proper topology influences the regulation of expression of developmentally essential genes. A fine example of regulatory specificity of enhancers controlled by chromatin architecture is that of *Pitx1*, a regulator of hindlimb development (Kragesteen et al., 2018). In hindlimbs, *Pitx1* is in close proximity with its enhancer (active), allowing for normal leg morphogenesis. In forelimbs, *Pitx1* is physically separated from the enhancer (inactive), allowing for normal arm development. The disturbance of that specificity (e.g., due to structural variants) can cause gene mis-expression and disease *in vivo* (Kragesteen et al., 2018). Transcription after activation of the glucocorticoid receptor occurs without significant changes of the pre-looped chromatin interactions, enabling its rapid reaction (Hakim et al., 2011). Changes in chromatin topology and conformation have already been associated and described in muscle progenitor specification and myogenic differentiation (Zhang et al., 2020), sensory experience during post-natal brain development (Tan et al., 2021), dendritic cell development and differentiation (Chauvistré and Seré, 2020), and neural development (Kishi and Gotoh, 2018).

An interesting question is whether conformation accompanies cell lineage decision and what the role of TFs is. During reprogramming, TFs reorganize genome structural features before changes in gene expression occur (Stadhouders et al., 2018). Somatic cell reprogramming is a useful model for investigating how genome topology affects cell fate decisions. A recent study, investigating chromatin interactions in ESC, iPSC and NPCs, revealed that reprogramming does not completely

restore a number of pluripotency-related interactions (Beagan et al., 2016). CTCF was abundant in these regions in ESC, while poor in differentiated NPCs. CTCF binding was not restored in iPSC causing an inadequate pluripotent genome topology recovery. The embryonic and trophoblast lineages have significant differences between them, in their epigenetic landscapes and their 3D conformation (Schoenfelder et al., 2018). ESCs have an enrichment for repressive interactions between gene promoters and also involving poised/silenced enhancers (marked with H3K27me3), whereas trophoblasts have an enrichment for active enhancer-gene interactions (Schoenfelder et al., 2018). Similarly, during neuronal differentiation of ESCs, Polycomb repressive complex 1 and 2 (PRC1 and PRC2) are known to have important functions in mediating repressing interactions. PRC1 mediated interactions are disrupted and gene-enhancer interactions become prominent (Bonev et al., 2017). Interestingly, poised enhancers in ESCs are already in close proximity with their target genes in a PRC2 dependent manner (Ngan et al., 2020). Deletion of PRC2 core components leads to activation of their target genes and embryonic lethality (Boyer et al., 2006; Bracken et al., 2006). When these enhancers are activated during differentiation of ESCs to neural progenitors, the interaction with their cognate genes is preserved, leading to their activation (Cruz-Molina et al., 2017). This is similar to the aforementioned pre-looping phenomenon. All in all, these results demonstrate that chromatin architecture changes may not cause instant transcriptional changes. As an alternative, structure seems to set the stage for future transcriptional changes by sculpturing the chromatin environment.

X chromosome inactivation (XCI) is another well-studied example to show how the 3D chromatin organization impacts development, as well as the differences between the two homologs (Lee and Bartolomei, 2013). One of the two X-chromosomes in female cells is randomly inactivated to equalize the expression levels of the X-linked genes between female and male cells, early during embryonic development or upon differentiation of female ESCs (Gribnau and Grootegoed, 2012). Several regulatory elements and genes directing the XCI process are located in a small region, the X inactivation center (Xic) (Barakat et al., 2014). This region harbors the best-studied mammalian lncRNA, *Xist* and its negative regulator *Tsix* (Lee et al., 1999). While *Xist* silences one X chromosome *in cis*, *Tsix* represses *Xist* also *in cis* and thus these two lncRNAs form a regulatory switch locus (Nora et al., 2012). The *Xist* locus has been proposed to be organized in two big TADs and XCI is initiated by the upregulation of *Xist* in one of the two X-chromosomes (Chaumeil et al., 2006; Engreitz et al., 2013). In another study, the two X chromosomes were shown to have distinct and different chromatin organization. The active X presented distinct compartmentalization of active and inactive regions, while the inactive X compartments were more uniform (Tan et al., 2018). TADs were present in the active X chromosome, but not in the inactivated X chromosome (Splinter et al., 2011; Nora et al., 2012; Giorgetti et al., 2016). In comparison, two mega-structures appeared on the inactivated X chromosome, separated by a microsatellite repeat containing several CTCF-binding sites (Horakova et al., 2012; Rao et al., 2014; Wang et al., 2016). Interestingly, a study using

mathematical prediction and experimental validation suggested that three internal elements (CTCF/binding sites within the *Linx*, *Chic1* and *Xite/Tsix* loci) might work in partnership with boundary elements for the formation and the stabilization of the two TADs (Bartman and Blobel, 2015). The deletion of these internal elements is sufficient to disrupt the TADs and subsequently triggers ectopic expression of genes at the neighboring TAD hence disturbing XCI process (Nora et al., 2012; Dixon et al., 2016).

Overall, TAD formation and maintenance as well as specificity of the enhancer-promoter interaction play key roles during development and differentiation to ensure the finely tuned regulation of gene expression and lineage decision.

CHROMATIN CONFORMATION IN DISEASE AND CANCER

Human diseases are often caused by structural variations (SVs) in the genome, through disruption of genes or changes in gene dosage (Spielmann et al., 2018; Ibrahim and Mundlos, 2020). While their effect in coding regions can be easily predicted, their occurrence in non-coding regions requires further investigation to address its influence on gene expression, for example in the case of limb formation involving the TAD-spanning *WNT6/IHH/EPHA4/PAX3* locus (Lupiáñez et al., 2016). SVs have the potential to interfere with genome architecture causing disease phenotypes (Lupiáñez et al., 2015; Spielmann et al., 2018). Depending on the type but also the extent of the SV, the effect on gene regulation may vary a lot (Ibrahim and Mundlos, 2020).

Disruption of genome architecture may lead to altered gene expression in a variety of ways and, as a result, disease phenotypes. This disruption is separated into inter-TAD and intra-TAD alterations.

Inter-TAD alterations can disrupt and rewire the 3D chromatin architecture resulting in changes of TAD boundaries, mis-regulation of important genes with deleterious effects and relocation of regulatory elements such as enhancers and/or silencers. Inter-TAD alterations are caused by many reasons. Genome architecture disruption involves the disruption of TADs borders, leading to contacts of enhancers and genes, otherwise insulated from each other, and thereby, the ectopic activation of those genes. This phenomenon is called "enhancer adoption" or "enhancer hijacking" (**Table 1**; Lettice et al., 2011; Northcott et al., 2014; Lupiáñez et al., 2016; Kaiser and Semple, 2017). Deletions result in TAD fusion (**Table 1**; Katainen et al., 2015; Lupiáñez et al., 2015; Flavahan et al., 2016), inversions in a swap of DNA regions (TAD shuffling) and duplications or translocations of regulatory or structural elements in new domains (neo-TADs) (**Table 2**; Gröschel et al., 2014; Northcott et al., 2014; Lupiáñez et al., 2015; Franke et al., 2016; Weischenfeldt et al., 2017). Furthermore, inter-TAD alterations could be caused by inversions, translocations of regulatory elements which may result in gain-of-function events by coupling enhancers with newly associated promoters, or loss-of-function events by separating enhancers from their associated promoters or a combination of the two (**Table 2**;

TABLE 1 | Summary of inter-TAD alterations (Enhancer adoption and TAD fusion), the disease/abnormality they caused and their description.

Enhancer adoption

LMNB1 locus	Adult-onset demyelinating leukodystrophy (ADLD)	A deletion eliminates a TAD boundary, leading to new interactions between the *LMNB1* promoter and three non-cognate enhancers and its subsequent activation, resulting in the progressive central nervous system demyelination (Giorgio et al., 2014)
FOXG1 locus	Rett syndrome	A telomeric deletion, including the TAD border, results in the ectopic activation of *FOXG1* by active enhancers in the brain (Allou et al., 2012)
GFI1B locus	Medulloblastoma	Somatic structural variants place *GFI1* or *GFI1B* near active enhancer sites, resulting in activation (Northcott et al., 2014)
SNCAIP locus	Group 4 medulloblastomas	A duplication of *SNCAIP* gene results in the ectopic activation of the putative oncogene *PRDM6* (Arabzade et al., 2020)

TAD fusion

EPHA4 locus	Brachydactyly	Deletions in the *EPHA4* locus that include a TAD border result in a fusion of the neighboring TADs, which attaches a cluster of limb-associated *EPHA4* enhancers to the *PAX3* gene and its concomitant mis-expression (Lupiáñez et al., 2015)
Six TAD boundaries encompassing T-ALL related genes	T-cell acute-lymphoblastic leukemia (T-ALL) or medulloblastoma	TAD disruption leads to ectopic proto-oncogene activation and abnormal cell proliferation (Northcott et al., 2014; Hnisz et al., 2016b; Weischenfeldt et al., 2017). CRISPR-engineered deletions of the TAD boundaries near the known oncogenes *TAL1* and *LMO2* result in new interactomes of those oncogenes with distal enhancers, leading to their aberrant activation (Hnisz et al., 2016b)
NOTCH1 locus	Ovarian cancer	Downregulation of *NOTCH1* gene due to its altered interactome as a result of mutations in the CTCF sites that disrupt the TAD boundary (Ji et al., 2016)
Various CTCF binding sites	Colorectal cancer	Frequently mutated *CTCF binding sites* lead to TAD boundary disruption and altered interactomes between genes and their regulatory elements (Katainen et al., 2015)
IRS4 locus	Lung squamous carcinoma, sarcomas and cervical squamous carcinoma	Deletions occurring at TAD boundaries coinciding with CTCF recruitment downstream of the *IRS4* locus led to *IRS4* overexpression (Weischenfeldt et al., 2017)
NEK6 locus	B cell lymphoma cell lines	Deletion of all CTCF-binding sites in the *NEK6* super-enhancer borders decreased the expression of *NEK6* while increased the expression of the neighboring *LHX2* gene (Huang et al., 2017)

Lupiáñez et al., 2016; Spielmann et al., 2018). To mention here, that while the above studies stress out the insulating role of TAD boundaries, it is important to keep in mind that TAD boundaries may not be the only component needed to maintain them (Anania and Lupiáñez, 2020). TADs did not fuse completely after serial deletions at the *Sox9* locus. This occurred only after the deletion of other CTCF sites within the locus (Despang et al., 2019). Deletions of CTCF-binding sites at the *Shh* locus result in structural changes, but TAD insulation is maintained (Paliou et al., 2019). Overall, these results reveal the ability of TAD borders to successfully organize the genome into distinct regulatory domains, as well as their ability to work and communicate with the internal structure elements.

Interestingly, in multiple myeloma 30% of the breakpoints are located at, or close to, TAD boundaries. The number of TADs is increased by 25% and they are smaller when compared to normal B cells, indicating that genomic rearrangements and translocations are driving forces in chromatin topology and creating new TADs (Wu et al., 2017). The smaller size of TADs in cancer cells when compared to their healthy controls can also be observed in prostate cancer and therefore seem to be most likely a general phenomenon in cancer cells. In the case of prostate cancer, this smaller size is the consequence of the splitting of one TAD in two, the majority of the TAD boundaries (∼98%) being the same between the prostate cancer cells and the normal ones (Taberlay et al., 2016). In prostate cancer, a deletion on 17p13.1 encompassing the *TP53* tumor suppressor locus leads to the division of a single TAD into two distinct smaller TADs, resulting into new chromatin interactomes of

the enhancers, promoters and insulators within the TADs and changing gene expression (Taberlay et al., 2016). Similarly, in mammary epithelial and breast cancer several TADs were divided into multiple sub-TADs but kept the same boundaries, as a result of various genomic alterations (Barutcu et al., 2015). In prostate cancer cells (and probably in most cancers) the size of the TADs (2–4 MB) is smaller compared to normal prostate cells (∼8 MB). These new small-TADs reside within the normal TAD architecture rather than forming new TADs, with the majority of the TAD boundaries (∼98%) to be the same between the prostate cancer cells and the normal ones (Taberlay et al., 2016).

Because of their ability to co-localize in the nucleus and/or their abundance within TAD boundaries, transposable elements (TEs) have been related to genome architecture (Dixon et al., 2012; Cournac et al., 2016). It has been shown that during the evolution of mammalian lineages, activation of retro-transposable elements triggered an increase of CTCF-binding events (Schmidt et al., 2012). As shown by changes in chromatin states, many of the new CTCF sites acted as chromatin insulators, affecting genome architecture and transcription. According to this observation, human T-lymphotropic virus type 1 (*HTLV-1*) translocation introduced an ectopic CTCF-binding site, which could form new loops and induce transcriptional changes in the new locus (Melamed et al., 2018). *HTLV-1* results in chronic inflammation in 10% of infect hosts.

Changes in the interactome and local chromatin architecture have also been associated to single nucleotide polymorphisms (SNPs) causing intra-TAD alterations. Intra-TAD alterations lead to abnormal transcriptional control of the genes inside the

TABLE 2 | Summary of inter-TAD alterations (TAD shuffling, Inter-TAD loss- or gain-of-function alterations and Neo-TADs), the disease/abnormality they caused and their description.

TAD shuffling		
Wnt6/Epha4 locus	F-syndrome	An inversion at the *Wnt6/Epha4* locus that misplaces the *Epha4* enhancers near *Wnt6* gene, causing its mis-expression in the developing limb bud (Lupiáñez et al., 2015; Kraft et al., 2019)
Ihh/Epha4 locus	Polydactyly	Duplications of the previous enhancers and rearranging them in front of the *Ihh* gene induce overexpression of *Ihh* (Kraft et al., 2019)
TFAP2A locus	Branchio-oculofacial syndrome	Inversion of the *TFAP2A* TAD resulted in lower *TFAP2A* expression due to the fact that the promoter was separated from its associated enhancers (Laugsch et al., 2019)
Shh locus	Digit syndactyly	An inversion at the *Shh* locus places the *Shh* gene in a TAD together with a limb enhancer, that induces its activation (Lettice et al., 2011)
MEF2C locus	5q14.3 microdeletion syndrome	Patients with balanced *MEF2C* translocations have been shown to be affected by the separation of promoters from their associated enhancers. The influence of these translocations was confirmed in patient-derived LCLs, which showed lower *MEF2C* expression (Redin et al., 2017)
GATA2 locus	Acute myeloid leukemia sub-types	A chromosomal inversion and translocation in chromosome 3 at two different breakpoints place the *GATA2* enhancer in the same TAD as the *EVI1* oncogene. The enhancer is then in close proximity with the *EVI1* promoter triggering its activation, which is responsible for the development of the disease (Gröschel et al., 2014)
IGF2 locus	Colorectal cancer	Recurrent tandem duplications encompassing a TAD boundary result into new interactions between *IGF2* and a cell specific super-enhancer located in the adjacent TAD, leading to its > 250-fold overexpression (Weischenfeldt et al., 2017). The duplications in the abovementioned TAD boundary are tandem rather inverted or dispersed, suggesting that the orientation of the enhancer and *IGF2* is probably important for the activation of *IGF2* (Beroukhim et al., 2016)
Inter-TAD loss- or gain-of-function alterations		
IDH locus	Gliomas	Mutations in the *IDH* gene results in accumulation of 2-hydroxyglutarate, which subsequently represses TET proteins. This causes hyper-methylation of CpG sites and increased methylation of CTCF sites affecting CTCF binding and the respective TAD boundaries. New interactions are consequently established between the oncogene *PDGFRA* with constitutive enhancers, which are normally located outside its normal TAD (Flavahan et al., 2016)
FMR1 locus	Fragile X syndrome (FXS)	The CGG triplet repeat (short tandem repeat or STRs) within the *FMR1* gene expands in an erratic way and the *FMR1* locus boundary is disrupted due to inability of CTCF to bound, caused by the abnormal DNA methylation levels. *FMR1* is silenced as the boundary is disrupted, because of the separation from its associated regulatory elements, which are now located in another TAD (Anania and Lupiáñez, 2020).
Neo-TADs		
Kcnj2 and Sox9 loci	Limb malformation	A neo-TAD where *Kcnj2* interacts with the Sox9 regulatory region resulted in overexpression of *Kcnj2* (Franke et al., 2016)
IGF2 locus	Cancer	Due to duplications of neighboring TADs, the new TAD incorporates the *IGF2* gene and a lineage-specific super enhancer, resulting in oncogenic locus mis-regulation (Weischenfeldt et al., 2017)

TAD, without altering its overall conformation. Many GWAS SNPs have now been connected to putative causative genes in hematopoietic cell types (Javierre et al., 2016; Mumbach et al., 2017). How sequence variations in putative regulatory elements lead to gene expression alterations that drive complex illnesses is largely unknown. On one hand, SNPs could restrict TFs or architectural proteins from interacting with their regulatory elements, leading to lower expression of their associated genes (**Table 3**; Schoenfelder and Fraser, 2019). SNPs can affect the recruitment of the LDB1 complex to the MYB enhancer, impairing its interaction with the MYB promoter, decreasing its expression and resulting in an increase of HbF expression (Stadhouders et al., 2014). On the other hand, SNPs could result in overexpression of target genes and/or their mis-expression in different cell types (Schoenfelder and Fraser, 2019). Gain or loss of function mutations in regulatory elements such as enhancers (or silencers) can affect the transcription of their cognate gene(s) (Spielmann and Mundlos, 2016; Schoenfelder and Fraser, 2019), provided that there is no other regulatory element compensating that gain or loss (Heinz et al., 2013). Another study attempted to identify the causative gene at GWAS

neurological disease loci by linking the SNPs with gene promoters and enhancers (Lu et al., 2020). They concluded that a SNP may only have subtle effects on looped target gene in healthy donors, but plays a more prominent role when the locus gains a disease-specific enhancer in patients. Their results indicated that high-quality Hi-C loops have a unique value in the study of disease genetics (Lu et al., 2020). Other GWAS studies have identified mutations in regulatory elements that could contribute to the Inflammatory bowel disease etiology by altering gene expression (Meddens et al., 2016). Duplications can change the copy number of regulatory elements, resulting in loss- or gain-of-function mutations, similar to the principle of gene dosage alterations occurring in the inter-TAD duplications and translocations (**Tables 3, 4**; Spielmann et al., 2018).

While SNPs could alter the content of specific enhancers, resulting to abnormal expression patterns, mutations in genes encoding TFs or architectural proteins could also have similar results (Schoenfelder and Fraser, 2019). Cohesinopathies and laminopathies, however, are the two groups of structural protein-associated human diseases that receive the most attention. Cohesinopathies are caused by mutations in genes associated

TABLE 3 | Summary of intra-TAD alterations (Single nucleotide polymorphisms and gain-of-function alterations), the disease/abnormality they caused and their description.

Single nucleotide polymorphisms (SNPs)

HSBL1-MYB locus	Hemoglobinopathies	SNPs affect the recruitment of the LDB1 complex to the *MYB* enhancer, impairing its interaction with the *MYB* promoter. Consequently, decrease of *MYB* expression results in an increase of *HbF* expression (Stadhouders et al., 2014)
CLEC16A locus	Autoimmune disease	SNPs in intron 19 of the *CLEC16A* gene have been shown to promote the interaction of the intron with the adjacent *DEXI* gene, resulting to its expression (Davison et al., 2012)
FTO locus	Obesity	An intron of the *FTO* gene containing obesity-associated SNPs interacts with the distal *IRX3* gene, and thus controlling its expression (Smemo et al., 2014; Schoenfelder and Fraser, 2019)
SNCA locus	Parkinson disease	A common Parkinson disease SNP in a non-coding distal enhancer factor prevents two repressive transcription factors, EMX2 and NKX6-1, from binding to a regulatory element, and thus, resulting in *SNCA* transcriptional upregulation (Soldner et al., 2016)
Various loci	Chronic Kidney Disease (CKD)	SNPs in both coding and non-coding regions have been discovered in studies of CKD, and dysregulation of gene expression of the 23 genes identified to be associated with such SNPs is possibly a contributing factor in CKD pathophysiology (Brandt et al., 2018)

Intra-TAD gain-of-function alterations

SHH locus	Polydactyly	Point mutations in the Sonic hedgehog (*SHH*) regulatory region *ZRS* result in the ectopic expression of *SHH* at the anterior margin in mouse. Although not formerly demonstrated in this study, these mutations allow the formation of chromatin looping between the *ZRS* region and the *SHH* promoter (Lettice et al., 2003).
MYC locus	Lung adenocarcinoma	Amplification of *MYC*-regulating enhancers results in a slightly higher *MYC* expression than in samples without amplification of *MYC* enhancers. The enhancer-amplified samples had a comparable *MYC* expression levels when compared to samples with *MYC* coding area amplification (Zhang et al., 2016; Agrawal et al., 2019)
IHH locus	Craniosynostosis and synpolydactyly	Duplications of regulatory elements within the *IHH* locus led to misexpression or overexpression of *IHH* and by this affect the complex regulatory signaling network during digit and skull development respectively (Klopocki et al., 2011)
CTSB locus	Keratolytic winter erythema	Overexpression of *CTSB* as a result of enhancer duplications (Ngcungcu et al., 2017)
Various loci	Prostate cancer	SNPs, associated with prostate cancer, co-localize/affect regions of active histone modification and transcription factor binding sites. 15 of the 17 identified genes in these loci exhibit a substantial change in expression, suggesting that the genes physically interacting with risk loci are associated to prostate cancer (Du et al., 2016)
Various loci	Atherosclerotic disease	294 additional candidate expressed genes for coronary artery disease (CAD) and large artery stroke (LAS) have been identified as potential factors in the pathophysiology of human atherosclerotic disease (Haitjema et al., 2017)
Various loci	Inflammatory bowel disease (IBD)	Mutations in DNA regulatory elements (DREs) can contribute to IBD etiology by altering gene expression (Meddens et al., 2016)
Pitx1 locus	Liebenberg syndrome	Deletion mutations upstream of the hindlimb expressed *Pitx1* gene result in intra-TAD conformation changes, merging a forelimb and hindlimb *Pitx1* gene enhancer (Kragesteen et al., 2018)
PITX1 locus	As previous	Translocation of two enhancers from chromosome 18 upstream of the *PITX1* on chromosome 5 (TAD shuffling), resulted in an increased *PITX1* expression (Spielmann et al., 2012)

with Cohesin complex and/or its regulators (Banerji et al., 2017; Norton and Phillips-Cremins, 2017; Davis et al., 2018; Olley et al., 2018; Krumm and Duan, 2019). CTCF and cohesin associated SNPs have been related to a number of human disorders and developmental defects. The significance and role of genome organizing factors like CTCF and the cohesin complex has been highlighted for a number of diseases. For example, CTCF depletion leads to pathological effects that are quite comparable to heart failure (Rosa-Garrido et al., 2017). Altered interactions and accessibility was shown at a substantial number of enhancer areas and the genes in the surrounding chromosomal areas were implicated in cardiac pathological pathways (Rosa-Garrido et al., 2017). Another example are the laminopathies caused by mutations in nuclear lamins (LMNA) and the lamin B receptor (LBR) genes. Given that LADs organize a large portion of the genome, the nuclear lamina and its components appear to play an important role in genome architecture. Laminopathies are distinct from other disorders in that a variety of disorders may be developed from just different mutations located in the same gene (Worman and Bonne, 2007).

Cancer is a particularly important area of disease where changes in the interactome are important. Alterations in TAD boundaries, which are observed in cancer, can lead to oncogene activation by affecting gene regulation in the flanking TADs via the establishment of new unusual promoter-enhancer interactions (**Figure 6**). Oncogene activation by TAD disruption and consequent enhancer adoption has been described in leukemia (Gröschel et al., 2014; Hnisz et al., 2016b), neuroblastoma (Peifer et al., 2015), colorectal cancer (Weischenfeldt et al., 2017), medulloblastoma (Northcott et al., 2014), glioma (Flavahan et al., 2016), sarcoma and squamous cancers (Weischenfeldt et al., 2017). Notably, the most prominent alterations in binding sequences at TAD boundaries, are located at CTCF binding motifs (Ji et al., 2016), although it should be noted that many CTCF binding sites are not boundaries. Approximately 11% of 922 deletion cases affect TAD boundaries at the vicinity of a disease-associated gene, resulting in "enhancer

TABLE 4 | Summary of intra-TAD alterations (loss-of-function alterations), the disease/abnormality they caused and their description.

Intra-TAD loss-of-function alterations

Shh locus	Preaxial polydactyly (PPD)	In the ZRS, two ETV4/ETV5 binding sites have been discovered. In transgenics, a single ETV binding site is sufficient to suppress ectopic expression; the absence of both sites leads in repressor activity loss and, as a result, in ectopic *Shh* expression in the limb bud (Lettice et al., 2012)
PAX6 locus	Aniridia	Point mutations (disruption of binding sites) in enhancers of *PAX6*, *PTF1A* and *TBX5* impair the expression of these genes. While it has not been demonstrated properly, these studies suggest that these point mutations impair the chromatin looping between these enhancers and their associated promoters (Smemo et al., 2012; Bhatia et al., 2013; Weedon et al., 2014)
PTF1A locus	Pancreatic agenesis	
TBX5 locus	Congenital heart disease	
SOX9 locus	Campomelic dysplasia	Sex reversal occurs when the relevant testis enhancer of *SOX9* is deleted, while deletions and point mutations further upstream induce the Pierre-Robin syndrome, which is characterized by cranial skeleton growth defects but normal sexual development (Benko et al., 2011)
DYNC1I1 locus	Split-hand/split-foot malformation (SHFM)	Exons 15 and 17 of *DYNC1I1* act as tissue specific limb enhancers of *DLX5/6*. Enhancer deletions in the *DYNC1I1* gene result in down regulation of the *DLX5/6* genes about 1Mb away (Allen et al., 2014; Tayebi et al., 2014)
ATOH7 locus	Non-syndromic congenital retinal non-attachment (NCRNA)	A deletion that covers a distal cis regulatory element upstream from *ATOH7* is responsible for NCRNA (Ghiasvand et al., 2011)
SHH locus	Holoprosencephaly (HPE)	The loss of function (disruption of binding sites) of *Shh* brain enhancer-2 (SBE2) in the hypothalamus of transgenic mouse embryos was caused by a rare nucleotide variant upstream of *SHH gene* found in an individual with HPE (Jeong et al., 2008)
MYC locus	Cleft lip with or without cleft palate (CL/P)	Deletion of a 640-kb non-coding region at 8q24, which contains distal cis-acting enhancers that regulate *Myc* expression in the developing face, causes modest facial morphological changes in mice and, on rare occasions, CL/P (Uslu et al., 2014)
BCL11A locus	β-hemoglobinopathies	A common variant in an erythroid enhancer of *BCL11A* is associated with reduced TF binding, modestly diminished BCL11A expression, and elevated HbF (Bauer et al., 2013)

adoption" (Swaminathan et al., 2012). A comprehensive analysis among various cancer cell lines, indicated that the formation of neo-TADs, encompassing cancer driver genes, is the result of SV alterations in cancer cells (Dixon et al., 2018). However, whether neo-TAD formation is a recurrent phenomenon in a given cancer cell type needs to be investigated further.

THE IMPORTANCE OF A REFINED IDENTIFICATION OF CHROMATIN CONFORMATION AND POTENTIAL THERAPEUTIC APPROACHES

An important question is what underlying mechanism protects TAD boundaries from deletions and disruptions? Using machine learning approaches, TAD boundaries were recently categorized based on strength (Gong et al., 2018). Strong TAD boundaries are less frequently lost in cancer, as they act as building blocks of the genome and encompass super-enhancers (Gong et al., 2018). In cancer, strong boundaries are notably safe from SVs and co-duplicated with super-enhancer elements (Gong et al., 2018). These observations and the observations that enhancers lead to aberrant activation of oncogenes due to genetic or epigenetic alterations highlight the importance of the chromatin architecture integrity (Lupiáñez et al., 2015; Flavahan et al., 2016; Hnisz et al., 2016b; Weischenfeldt et al., 2017).

An interesting question is whether mis-regulation of TFs causes the altered 3D chromatin organization or whether the opposite takes place? Intriguingly, studies advocate both options. A gene fusion in prostate cancer, causes the overexpression of oncogenic *ERG* resulting in changes in chromatin organization

and territories encompassing genes associated with aggressive prostate cancer (Rickman et al., 2012). This hypothesis may also be true for other TFs whose aberrant expression is involved in many other cancers (Rickman et al., 2012). In contrast, chromosomal inversion and translocation in chromosome 3 at two different breakpoints, tethering the enhancer of *GATA2* in the same TAD as *EVI1*, activate expression of *EVI1* and downregulate *GATA2*, resulting in the development of acute myeloid leukemia (Gröschel et al., 2014).

Thus, is chromatin architecture characteristic of each disease and can we predict the effect of SVs in chromatin organization? A support vector machine classifier (3D-SP) can separate leukemia sub-types based on the information contained in the chromatin architecture and specifically the interactome of the *HOXA* gene cluster in various leukemia cell lines (Rousseau et al., 2014), while a recently developed approach can be used to predict *in silico* the altered 3D conformation resulting from structural variants (Bianco et al., 2018). Hence, the improvement of new chromatin conformation techniques can help to better understand the biological effect of newly discovered structural variants and TAD alterations in the human genome, that are linked to uncharacterized genetic disorders or diseases and to evaluate their role on chromatin architecture and transcriptional control. Interestingly, chromatin conformation capture techniques which employ selection based on oligonucleotides, like T2C (Kolovos et al., 2018) and capture-promoter Hi-C (cpHi-C) (Schoenfelder et al., 2015a), can identify the interactome of those specific fragments. Especially in the cases of where SNPs are heterozygous in these fragments, oligonucleotides designed for the two alleles can discriminate the interactome of the wild type allele compared to the allele containing the SNP.
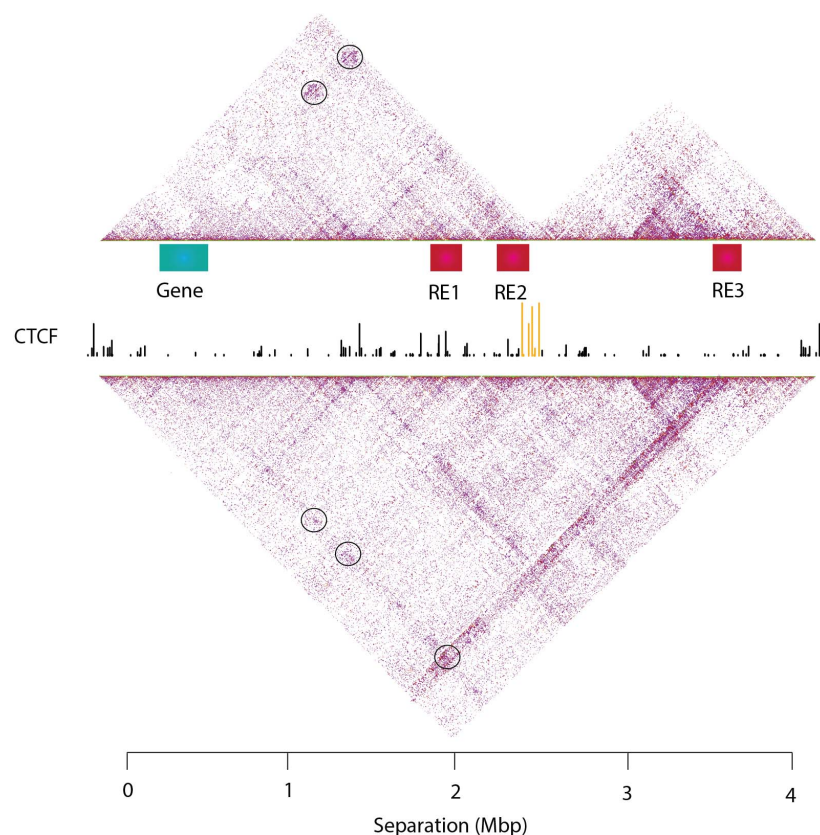
FIGURE 6 | An example of TAD disruption in cancer and rewiring of promoter-enhancer proximity. The upper panel depicts two distinct TADs, the left containing a gene (depicted with a green box) and two regulatory elements (RE1 and RE2 depicted with red boxes), which can be either an enhancer, or a poised enhancer or a silencer. The right TAD contains one regulatory element (RE3) that would be compatible with the gene. In the upper panel, the gene is located in a confined place with RE1 and RE2 (depicted with round black circles) resulting in its normal transcriptional activation (if RE1 and RE2 are enhancers) or its repression (if RE1 and RE2 are poised enhancers or silencers). Mutation or deletion of the CTCF sites (depicted with yellow) located at the boundary between the TADs, results in the reorganization of the TAD topology and fusion of the two TADs into one. Thus, in the bottom panel, the gene is now in close proximity (and interacts frequently) with RE3 (depicted with round black circle), leading to its expression also by the RE3, if RE3 is an enhancer or its downregulation if RE3 is a silencer. The different combinations of REs could have different results in the expression levels of the gene. Since, RE1 and RE2 contacts are diminished, it could lead to less expression by those two enhancers while the expression levels of the gene remain the same. On the other hand, the combination of RE1, RE2, and RE3 could lead to a super-enhancer and higher levels of expression of the gene.

Targeting chromatin interactions could potentially provide therapeutic approaches (Babu and Fullwood, 2015). Perturbing promoter-enhancer interactions would permit the fine tuning of expression of target genes, in a reversible and specific manner. However, it faces many difficulties that would need to be overcome. CTCF, cohesin and other TFs mediate many different chromatin interactions. Frequently, these TFs are also involved in signaling pathways. Thus, a systemic perturbation of TFs would cause many off-target effects. Moreover, proteins, which mediate chromatin interactions, are often found in the nucleus and are therefore difficult to perturb by antibodies or small molecule inhibitors. Various epigenetic regulators are involved in cancer, but whether they are involved in chromatin organization is poorly understood. Many drugs have been developed for epigenetic regulators but again it has not been examined yet whether they affect chromatin interactions and compartmentalization, although it is likely that many will affect genomic interactions directly by enabling or preventing the

binding of TF type protein (e.g., CTCF is DNA methylation sensitive) or indirectly via changes in the transcriptome. Interestingly, a recent study has identified 50 factors that are potentially important for genome organization (Shachar et al., 2015). However, this study applied an siRNA approach, which is known to cause off target effects. To overcome the non-specificity of targeting such proteins, a new tool (CLOuD9) for the precise manipulation of 3D chromatin structure and chromatin looping has been developed by employing the CRISPR/Cas9 approach and establishing stable chromatin loops (Morgan et al., 2017). This approach may be useful in cancer diagnostics, where chromosomal rearrangements interrupt genomic organization and alter gene expression. Thus, screening studies preferably with the use of drugs or CRISPR/Cas9 approach targeting alterations of chromatin conformation structure, could unveil new factors, which mediate chromatin interactions and unveil them as potential new therapeutic targets. More promising would perhaps be the development of genome editing tools to alter the

binding sites of TFs or CTCF using Crispr/Cas9 and homing technology to target the appropriate cells (Cruz et al., 2021).

It is clear from the studies above and many others that chromatin conformation plays a key role in cancer. Thus, understanding the modulation of chromatin interactions will unveil the underlying mechanisms of diseases, development and cancer and identify new promising therapeutic targets.

CONCLUSION AND FUTURE PERSPECTIVES

The integrity of the 3D chromatin architecture and the genome interactome is important to ensure proper transcriptional control. Alterations of this topology are often correlated with diseases such as cancer. Since the genome of cancer cells or cells derived from other pathologies are often instable, TAD disruption is observed often, that result in altered gene expression profiles leading to tumorigenesis or other pathology. Hence, mapping the precise location of TAD topology, their boundaries and other structures is an integral part of deciphering the genetic basis of gene expression in cancer and other diseases, and possibly provide new therapeutic targets. Moreover, the recent development of CRISPR-Cas9 technique (Ran et al., 2013) could lead to correcting altered TAD boundaries in patient cells, offering an exciting potential therapeutic strategy. Recently developed high resolution chromatin conformation techniques [e.g., Hi-C (Rao et al., 2014), T2C (Kolovos et al., 2018)] that offer sub-Kbp resolution, could unveil the precise location of TAD boundaries and their detailed features, holding the key to better understand diseases. Finally, we propose a model integrating recent developments in chromatin architecture with the formation of either structural or functional loops, controlling

proper transcriptional control. Understanding how these loops are formed and how they evolve is essential to identify new mechanisms triggering pathologies such as cancer and to develop new efficient therapeutic strategies.

However, despite the spectacular recent advances in the field of chromatin architecture and gene regulation, many questions still remain to be answered. Some of those are: is gene activation preceding locus conformation or *vice versa*? What is the underlying mechanism creating TADs and protecting TAD boundaries from deletions and disruptions, e.g., is it continuous loop extrusion? Is 3D conformation accompanying cell lineage decisions? How were regulatory elements generated during evolution? Are there as yet unknown TFs, which contribute to 3D genome structure? How can we efficiently identify these? Deciphering all these questions could further lead to our understanding of the dynamics and forces of chromatin organization to enable all the necessary functions of cells.

AUTHOR CONTRIBUTIONS

All authors structured, wrote and proofread the manuscript.

FUNDING

REFERENCES

Agrawal, P., Heimbruch, K. E., and Rao, S. (2019). Genome-wide maps of transcription regulatory elements and transcription enhancers in development and disease. *Compr. Physiol.* 9, 439–455. doi: 10.1002/cphy.c180028

Allahyar, A., Vermeulen, C., Bouwman, B. A. M., Krijger, P. H. L., Verstegen, M. J. A. M., Geeven, G., et al. (2018). Enhancer hubs and loop collisions identified from single-allele topologies. *Nat. Genet.* 50, 1151–1160. doi: 10.1038/s41588-018-0161-5

Allen, H. L., Caswell, R., Xie, W., Xu, X., Wragg, C., Turnpenny, P. D., et al. (2014). Next generation sequencing of chromosomal rearrangements in patients with split-hand/split-foot malformation provides evidence for DYNC1I1 exonic enhancers of DLX5/6 expression in humans. *J. Med. Genet.* 51, 264–267. doi: 10.1136/jmedgenet-2013-102142

Allou, L., Lambert, L., Amsallem, D., Bieth, E., Edery, P., Destrée, A., et al. (2012). 14q12 and severe Rett-like phenotypes: new clinical insights and physical mapping of FOXG1-regulatory elements. *Eur. J. Hum. Genet.* 20, 1216–1223. doi: 10.1038/ejhg.2012.127

Anania, C., and Lupiáñez, D. G. (2020). Order and disorder: abnormal 3D chromatin organization in human disease. *Brief. Funct. Genomics* 19, 128–138. doi: 10.1093/bfgp/elz028

Andrey, G., and Mundlos, S. (2017). *The Three-Dimensional Genome: regulating Gene Expression During Pluripotency and Development*. United Kingdom: Company of Biologists Ltd.

Arabzade, A., Stuckert, A. J., Bertrand, K. C., and Mack, S. C. (2020). *Invited Review: the Role and Contribution of Transcriptional Enhancers in Brain Cancer*. United States: Blackwell Publishing Ltd.

Austenaa, L. M., Barozzi, I., Simonatto, M., Masella, S., Della Chiara, G., Ghisletti, S., et al. (2015). Transcription of Mammalian cis-Regulatory Elements Is Restrained by Actively Enforced Early Termination. *Mol. Cell.* 60, 460–474. doi: 10.1016/j.molcel.2015.09.018

Azagra, A., Marina-Zárate, E., Ramiro, A. R., Javierre, B. M., and Parra, M. (2020). *From Loops to Looks: transcription Factors and Chromatin Organization Shaping Terminal B Cell Differentiation*. Netherlands: Elsevier Ltd.

Babu, D., and Fullwood, M. J. (2015). 3D genome organization in health and disease: emerging opportunities in cancer translational medicine. *Nucleus* 6, 382–393. doi: 10.1080/19491034.2015.1106676

Banerji, R., Skibbens, R. V., and Iovine, M. K. (2017). How many roads lead to cohesinopathies? *Dev. Dyn.* 246, 881–888. doi: 10.1002/dvdy.24510

Barakat, T. S., Loos, F., van Staveren, S., Myronova, E., Ghazvini, M., Grootegoed, J. A., et al. (2014). The trans-activator RNF12 and cis-acting elements effectuate X chromosome inactivation independent of X-pairing. *Mol. Cell.* 53, 965–978. doi: 10.1016/j.molcel.2014.02.006

Barbieri, M., Xie, S. Q., Torlai Triglia, E., Chiariello, A. M., Bianco, S., de Santiago, I., et al. (2017). Active and poised promoter states drive folding of the extended HoxB locus in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* 24, 515–524. doi: 10.1038/nsmb.3402

Bartman, C. R., and Blobel, G. A. (2015). Perturbing Chromatin Structure to Understand Mechanisms of Gene Expression. *Cold Spring Harb. Symp. Quant. Biol.* 80, 207–212. doi: 10.1101/sqb.2015.80.027359

Barutcu, A. R., Lajoie, B. R., McCord, R. P., Tye, C. E., Hong, D., Messier, T. L., et al. (2015). Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. *Genome Biol.* 16:214. doi: 10.1186/s13059-015-0768-0

Baù, D., Sanyal, A., Lajoie, B. R., Capriotti, E., Byron, M., Lawrence, J. B., et al. (2011). The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules. *Nat. Struct. Mol. Biol.* 18, 107–114. doi: 10.1038/nsmb.1936

Bauer, D. E., Kamran, S. C., Lessard, S., Xu, J., Fujiwara, Y., Lin, C., et al. (2013). An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. *Science* 342, 253–257. doi: 10.1126/science.1242088

Beagan, J. A., Gilgenast, T. G., Kim, J., Plona, Z., Norton, H. K., Hu, G., et al. (2016). Local Genome Topology Can Exhibit an Incompletely Rewired 3D-Folding State during Somatic Cell Reprogramming. *Cell Stem Cell* 18, 611–624. doi: 10.1016/j.stem.2016.04.004

Beccari, L., Yakushiji-Kaminatsui, N., Woltering, J. M., Necsulea, A., Lonfat, N., Rodríguez-Carballo, E., et al. (2016). A role for HOX13 proteins in the regulatory switch between TADs at the HoxD locus. *Genes Dev.* 30, 1172–1186. doi: 10.1101/gad.281055.116

Bender, M. A., Ragoczy, T., Lee, J., Byron, R., Telling, A., Dean, A., et al. (2012). The hypersensitive sites of the murine β-globin locus control region act independently to affect nuclear localization and transcriptional elongation. *Blood* 119, 3820–3827. doi: 10.1182/blood-2011-09-380485

Benko, S., Gordon, C. T., Mallet, D., Sreenivasan, R., Thauvin-Robinet, C., Brendehaug, A., et al. (2011). Disruption of a long distance regulatory region upstream of SOX9 in isolated disorders of sex development. *J. Med. Genet.* 48, 825–830. doi: 10.1136/jmedgenet-2011-100255

Beroukhim, R., Zhang, X., and Meyerson, M. (2016). Copy number alterations unmasked as enhancer hijackers. *Nat. Genet.* 49, 5–6. doi: 10.1038/ng.3754

Bhatia, S., Bengani, H., Fish, M., Brown, A., Divizia, M. T., de Marco, R., et al. (2013). Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia. *Am. J. Hum. Genet.* 93, 1126–1134. doi: 10.1016/j.ajhg.2013.10.028

Bianco, S., Lupiáñez, D. G., Chiariello, A. M., Annunziatella, C., Kraft, K., Schöpflin, R., et al. (2018). Polymer physics predicts the effects of structural variants on chromatin architecture. *Nat. Genet.* 50, 662–667. doi: 10.1038/s41588-018-0098-8

Bintu, B., Mateo, L. J., Su, J. H., Sinnott-Armstrong, N. A., Parker, M., Kinrot, S., et al. (2018). Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* 362:eaau1783. doi: 10.1126/science.aau1783

Boija, A., Klein, I. A., Sabari, B. R., Dall'Agnese, A., Coffey, E. L., Zamudio, A. V., et al. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* 175, 1842–1855.e16. doi: 10.1016/j.cell.2018.10.042

Bompadre, O., and Andrey, G. (2019). Chromatin topology in development and disease. *Curr. Opin. Genet. Dev.* 55, 32–38. doi: 10.1016/j.gde.2019.04.007

Bonev, B., and Cavalli, G. (2016). Organization and function of the 3D genome. *Nat. Rev. Genet.* 17, 661–678. doi: 10.1038/nrg.2016.112

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G. L., Lubling, Y., et al. (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* 171, 557–572.e24. doi: 10.1016/j.cell.2017.09.043

Bouwman, B. A., and de Laat, W. (2015). Architectural hallmarks of the pluripotent genome. *FEBS Lett.* 589, 2905–2913. doi: 10.1016/j.febslet.2015.04.055

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441, 349–353. doi: 10.1038/nature04733

Bracken, A. P., Dietrich, N., Pasini, D., Hansen, K. H., and Helin, K. (2006). Genome-wide mapping of polycomb target genes unravels their roles in cell fate transitions. *Genes Dev.* 20, 1123–1136. doi: 10.1101/gad.381706

Branco, M. R., and Pombo, A. (2006). Intermingling of Chromosome Territories in Interphase Suggests Role in Translocations and Transcription-Dependent Associations. *PLoS Biol.* 4:e138. doi: 10.1371/journal.pbio.0040138

Brandt, M. M., Meddens, C. A., Louzao-Martinez, L., van den Dungen, N. A. M., Lansu, N. R., Nieuwenhuis, E. E. S., et al. (2018). Chromatin Conformation Links Distal Target Genes to CKD Loci. *J. Am. Soc. Nephrol.* 29, 462–476. doi: 10.1681/ASN.2016080875

Brown, J. M., Roberts, N. A., Graham, B., Waithe, D., Lagerholm, C., Telenius, J. M., et al. (2018). A tissue-specific self-interacting chromatin domain forms independently of enhancer-promoter interactions. *Nat. Commun.* 9, 1–15. doi: 10.1038/s41467-018-06248-4

Cai, Y., Zhang, Y., Loh, Y. P., Tng, J. Q., Lim, M. C., Cao, Z., et al. (2021). H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions. *Nat. Commun.* 12:719. doi: 10.1038/s41467-021-20940-y

Cai, Z., Cao, C., Ji, L., Ye, R., Wang, D., Xia, C., et al. (2020). RIC-seq for global in situ profiling of RNA-RNA spatial interactions. *Nature* 582, 432–437. doi: 10.1038/s41586-020-2249-1

Carone, B. R., Hung, J. H., Hainer, S. J., Chou, M. T., Carone, D. M., Weng, Z., et al. (2014). High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. *Dev. Cell* 30, 11–22. doi: 10.1016/j.devcel.2014.05.024

Chaumeil, J., Le Baccon, P., Wutz, A., and Heard, E. (2006). A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev.* 20, 2223–2237. doi: 10.1101/gad.380906

Chauvistré, H., and Seré, K. (2020). Epigenetic aspects of DC development and differentiation. *Mol. Immunol.* 128, 116–124. doi: 10.1016/j.molimm.2020.10.011

Chen, C. K., Blanco, M., Jackson, C., Aznauryan, E., Ollikainen, N., Surka, C., et al. (2016). Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science* 354, 468–472. doi: 10.1126/science.aae0047

Cho, W. K., Spille, J. H., Hecht, M., Lee, C., Li, C., Grube, V., et al. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* 361, 412–415. doi: 10.1126/science.aar4199

Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G. M., Cattoglio, C., et al. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* 361:eaar2555. doi: 10.1126/science.aar2555

Chu, C., Qu, K., Zhong, Franklin, L., Artandi, S. E., Chang, et al. (2011). Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. *Mol. Cell* 44, 667–678. doi: 10.1016/j.molcel.2011.08.027

Clift, D., and Schuh, M. (2013). Restarting life: fertilization and the transition from meiosis to mitosis. *Nat. Rev. Mol. Cell. Biol.* 14, 549–562. doi: 10.1038/nrm3643

Corces, M. R., Buenrostro, J. D., Wu, B., Greenside, P. G., Chan, S. M., Koenig, J. L., et al. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* 48, 1193–1203. doi: 10.1038/ng.3646

Cournac, A., Koszul, R., and Mozziconacci, J. (2016). The 3D folding of metazoan genomes correlates with the association of similar repetitive elements. *Nucleic Acids Res.* 44, 245–255. doi: 10.1093/nar/gkv1292

Cremer, M., Brandstetter, K., Maiser, A., Rao, S. S. P., Schmid, V. J., Guirao-Ortiz, M., et al. (2020). Cohesin depleted cells rebuild functional nuclear compartments after endomitosis. *Nat. Commun.* 11, 1–16. doi: 10.1038/s41467-020-19876-6

Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* 2, 292–301. doi: 10.1038/35066075

Cruz, L. J., van Dijk, T., Vepris, O., Li, T. M. W. Y., Schomann, T., Baldazzi, F., et al. (2021). PLGA-Nanoparticles for Intracellular Delivery of the CRISPR-Complex to Elevate Fetal Globin Expression in Erythroid Cells. *Biomaterials* 268:120580. doi: 10.1016/j.biomaterials.2020.120580

Cruz-Molina, S., Respuela, P., Tebartz, C., Kolovos, P., Nikolic, M., Fueyo, R., et al. (2017). PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. *Cell Stem Cell* 20, 689–705.e9. doi: 10.1016/j.stem.2017.02.004

Dao, L. T. M., Galindo-Albarrán, A. O., Castro-Mondragon, J. A., Andrieu-Soler, C., Medina-Rivera, A., Souaid, C., et al. (2017). Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nat. Genet.* 49, 1073–1081. doi: 10.1038/ng.3884

Davis, L., Onn, I., and Elliott, E. (2018). *The Emerging Roles for the Chromatin Structure Regulators CTCF and Cohesin in Neurodevelopment and Behavior.* Switzerland: Birkhauser Verlag AG.

Davison, L. J., Wallace, C., Cooper, J. D., Cope, N. F., Wilson, N. K., Smyth, D. J., et al. (2012). Long-range DNA looping and gene expression analyses identify DEXI as an autoimmune disease candidate gene. *Hum. Mol. Genet.* 21, 322–333. doi: 10.1093/hmg/ddr468

de Laat, W., and Grosveld, F. (2003). Spatial organization of gene expression: the active chromatin hub. *Chromosome Res.* 11, 447–459. doi: 10.1023/a: 1024922626726

De Wit, E., Bouwman, B. A. M., Zhu, Y., Klous, P., Splinter, E., Verstegen, M. J. A. M., et al. (2013). The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature* 501, 227–231. doi: 10.1038/nature12420

de Wit, E., Vos, E. S., Holwerda, S. J., Valdes-Quezada, C., Verstegen, M. J., Teunissen, H., et al. (2015). CTCF Binding Polarity Determines Chromatin Looping. *Mol. Cell.* 60, 676–684. doi: 10.1016/j.molcel.2015.09.023

Dekker, J., and Heard, E. (2015). Structural and functional diversity of Topologically Associating Domains. *FEBS Lett.* 589, 2877–2884. doi: 10.1016/j.febslet.2015.08.044

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P. D., et al. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* 149, 1233–1244. doi: 10.1016/j.cell.2012.03.051

Deng, W., Rupon, J. W., Krivega, I., Breda, L., Motta, I., Jahn, K. S., et al. (2014). Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell* 158, 849–860. doi: 10.1016/j.cell.2014.05.050

Denker, A., and de Laat, W. (2016). The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev.* 30, 1357–1382. doi: 10.1101/gad.281964.116

Despang, A., Schöpflin, R., Franke, M., Ali, S., Jerković, I., Paliou, C., et al. (2019). Functional dissection of the Sox9–Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat. Genet.* 51, 1263–1271. doi: 10.1038/s41588-019-0466-z

Di Stefano, M., Stadhouders, R., Farabella, I., Castillo, D., Serra, F., Graf, T., et al. (2020). Transcriptional activation during cell reprogramming correlates with the formation of 3D open chromatin hubs. *Nat. Commun.* 11, 1–12. doi: 10.1038/s41467-020-16396-1

Dillon, N., Trimborn, T., Strouboulis, J., Fraser, P., and Grosveld, F. (1997). The effect of distance on long-range chromatin interactions. *Mol. Cell.* 1, 131–139. doi: 10.1016/s1097-2765(00)80014-3

Dixon, J. R., Gorkin, D. U., and Ren, B. (2016). Chromatin Domains: the Unit of Chromosome Organization. *Mol. Cell.* 62, 668–680. doi: 10.1016/j.molcel.2016.05.018

Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature* 518, 331–336. doi: 10.1038/nature14222

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380. doi: 10.1038/nature11082

Dixon, J. R., Xu, J., Dileep, V., Zhan, Y., Song, F., Le, V. T., et al. (2018). Integrative detection and analysis of structural variation in cancer genomes. *Nat. Genet.* 50, 1388–1398. doi: 10.1038/s41588-018-0195-8

Drissen, R., Palstra, R. J., Gillemans, N., Splinter, E., Grosveld, F., Philipsen, S., et al. (2004). The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev.* 18, 2485–2490. doi: 10.1101/gad.317004

Du, M., Tillmans, L., Gao, J., Gao, P., Yuan, T., Dittmar, R. L., et al. (2016). Chromatin interactions and candidate genes at ten prostate cancer risk loci. *Sci. Rep.* 6, 23202–23202. doi: 10.1038/srep23202

Du, Z., Zheng, H., Huang, B., Ma, R., Wu, J., Zhang, X. X., et al. (2017). Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* 547, 232–235. doi: 10.1038/nature23263

Emera, D., Yin, J., Reilly, S. K., Gockley, J., and Noonan, J. P. (2016). Origin and evolution of developmental enhancers in the mammalian neocortex. *Proc. Natl. Acad. Sci. U. S. A.* 113, E2617–E2626. doi: 10.1073/pnas.1603718113

Engreitz, J. M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., et al. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* 341:1237973. doi: 10.1126/science.1237973

Eres, I. E., and Gilad, Y. (2021). *A TAD Skeptic: is 3D Genome Topology Conserved?*. Netherlands: Elsevier Ltd.

Espinola, S. M., Götz, M., Bellec, M., Messina, O., Fiche, J. B., Houbron, C., et al. (2021). Cis-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early Drosophila development. *Nat. Genet.* 53, 477–486. doi: 10.1038/s41588-021-00816-z

Falk, M., Feodorova, Y., Naumova, N., Imakaev, M., Lajoie, B. R., Leonhardt, H., et al. (2019). Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* 570, 395–399. doi: 10.1038/s41586-019-1275-3

Fang, F., Xu, Y., Chew, K. K., Chen, X., Ng, H. H., and Matsudaira, P. (2014). Coactivators p300 and CBP maintain the identity of mouse embryonic stem cells by mediating long-range chromatin structure. *Stem Cells* 32, 1805–1816. doi: 10.1002/stem.1705

Fanucchi, S., Shibayama, Y., Burd, S., Weinberg, M. S., and Mhlanga, M. M. (2013). Chromosomal contact permits transcription between coregulated genes. *Cell* 155, 606–620. doi: 10.1016/j.cell.2013.09.051

Finn, E. H., Pegoraro, G., Brandão, H. B., Valton, A. L., Oomen, M. E., Dekker, J., et al. (2019). Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell* 176, 1502–1515.e10. doi: 10.1016/j.cell.2019.01.020

Flavahan, W. A., Drier, Y., Liau, B. B., Gillespie, S. M., Venteicher, A. S., Stemmer-Rachamimov, A. O., et al. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* 529, 110–114. doi: 10.1038/nature16490

Flyamer, I. M., Gassler, J., Imakaev, M., Brandão, H. B., Ulianov, S. V., Abdennur, N., et al. (2017). Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* 544, 110–114. doi: 10.1038/nature21711

Franke, M., Ibrahim, D. M., Andrey, G., Schwarzer, W., Heinrich, V., Schöpflin, R., et al. (2016). Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* 538, 265–269. doi: 10.1038/nature19800

Fraser, J., Ferrai, C., Chiariello, A. M., Schueler, M., Rito, T., Laudanno, G., et al. (2015). Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Mol. Syst. Biol.* 11:852. doi: 10.15252/msb.20156492

Fraser, P., Pruzina, S., Antoniou, M., and Grosveld, F. (1993). Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes. *Genes Dev.* 7, 106–113. doi: 10.1101/gad.7.1.106

Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A., and Mirny, L. A. (2017). Emerging Evidence of Chromosome Folding by Loop Extrusion. *Cold Spring Harb. Symp. Quant. Biol.* 82, 45–55. doi: 10.1101/sqb.2017.82.034710

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L. A. (2016). Formation of Chromosomal Domains by Loop Extrusion. *Cell. Rep.* 15, 2038–2049. doi: 10.1016/j.celrep.2016.04.085

Fudenberg, G., and Mirny, L. A. (2012). Higher-order chromatin structure: bridging physics and biology. *Curr. Opin. Genet. Dev.* 22, 115–124. doi: 10.1016/j.gde.2012.01.006

Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer Control of Transcriptional Bursting. *Cell* 166, 358–368. doi: 10.1016/j.cell.2016.05.025

Gaspar-Maia, A., Alajem, A., Meshorer, E., and Ramalho-Santos, M. (2011). *Open Chromatin in Pluripotency and Reprogramming*. Germany: Nature Publishing Group.

Gassler, J., Brandão, H. B., Imakaev, M., Flyamer, I. M., Ladstätter, S., Bickmore, W. A., et al. (2017). A mechanism of cohesin–dependent loop extrusion organizes zygotic genome architecture. *EMBO J.* 36, 3600–3618. doi: 10.15252/embj.201798083

Ghamari, A., van de Corput, M. P., Thongjuea, S., van Cappellen, W. A., van Ijcken, W., van Haren, J., et al. (2013). In vivo live imaging of RNA polymerase II transcription factories in primary cells. *Genes Dev.* 27, 767–777. doi: 10.1101/gad.216200.113

Ghavi-Helm, Y., Klein, F. A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., et al. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* 512, 96–100. doi: 10.1038/nature13417

Ghiasvand, N. M., Rudolph, D. D., Mashayekhi, M., Brzezinski, J. A., Goldman, D., and Glaser, T. (2011). Deletion of a remote enhancer near ATOH7 disrupts retinal neurogenesis, causing NCRNA disease. *Nat. Neurosci.* 14, 578–588. doi: 10.1038/nn.2798

Giorgetti, L., Lajoie, B. R., Carter, A. C., Attia, M., Zhan, Y., Xu, J., et al. (2016). Structural organization of the inactive X chromosome in the mouse. *Nature* 535, 575–579. doi: 10.1038/nature18589

Giorgetti, L., Servant, N., and Heard, E. (2013). Changes in the organization of the genome during the mammalian cell cycle. *Genome Biol.* 14:142. doi: 10.1186/gb4147

Giorgio, E., Robyr, D., Spielmann, M., Ferrero, E., Di Gregorio, E., Imperiale, D., et al. (2014). A large genomic deletion leads to enhancer adoption by the lamin B1 gene: a second path to autosomal dominant adult-onset demyelinating leukodystrophy (ADLD). *Hum. Mol. Genet.* 24, 3143–3154. doi: 10.1093/hmg/ddv065

Golfier, S., Quail, T., Kimura, H., and Brugués, J. (2020). Cohesin and condensin extrude DNA loops in a cell-cycle dependent manner. *Elife* 9, 1–34. doi: 10.7554/eLife.53885

Gong, Y., Lazaris, C., Sakellaropoulos, T., Lozano, A., Kambadur, P., Ntziachristos, P., et al. (2018). Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. *Nat. Commun.* 9:542. doi: 10.1038/s41467-018-03017-1

Goode, D. K., Obier, N., Vijayabaskar, M. S., Lie-A-Ling, M., Lilly, A. J., Hannah, R., et al. (2016). Dynamic Gene Regulatory Networks Drive Hematopoietic Specification and Differentiation. *Dev. Cell.* 36, 572–587. doi: 10.1016/j.devcel.2016.01.024

Gribnau, J., and Grootegoed, J. A. (2012). Origin and evolution of X chromosome inactivation. *Curr. Opin. Cell. Biol.* 24, 397–404. doi: 10.1016/j.ceb.2012.02.004

Gröschel, S., Sanders, M. A., Hoogenboezem, R., de Wit, E., Bouwman, B. A. M., Erpelinck, C., et al. (2014). A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell* 157, 369–381. doi: 10.1016/j.cell.2014.02.019

Grosveld, F., van Assendelft, G. B., Greaves, D. R., and Kollias, G. (1987). Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* 51, 975–985. doi: 10.1016/0092-8674(87)90584-8

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M. B., Talhout, W., et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 453, 948–951. doi: 10.1038/nature06947

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D. U., et al. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* 162, 900–910. doi: 10.1016/j.cell.2015.07.038

Guo, Y. E., Manteiga, J. C., Henninger, J. E., Sabari, B. R., Dall'Agnese, A., Hannett, N. M., et al. (2019). Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* 572, 543–548. doi: 10.1038/s41586-019-1464-0

Haarhuis, J. H. I., van der Weide, R. H., Blomen, V. A., Yáñez-Cuna, J. O., Amendola, M., van Ruiten, M. S., et al. (2017). The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* 169, 693–707.e14. doi: 10.1016/j.cell.2017.04.013

Haitjema, S., Meddens, C. A., Van Der Laan, S. W., Kofink, D., Harakalova, M., Tragante, V., et al. (2017). Additional Candidate Genes for Human Atherosclerotic Disease Identified Through Annotation Based on Chromatin Organization. *Circ. Cardiovasc. Genet.* 10:e001664. doi: 10.1161/CIRCGENETICS.116.001664

Hakim, O., Sung, M. H., Voss, T. C., Splinter, E., John, S., Sabo, P. J., et al. (2011). Diverse gene reprogramming events occur in the same spatial clusters of distal regulatory elements. *Genome Res.* 21, 697–706. doi: 10.1101/gr.111153.110

Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N., et al. (1991). Importance of globin gene order for correct developmental expression. *Genes Dev.* 5, 1387–1394. doi: 10.1101/gad.5.8.1387

Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R., and Darzacq, X. (2017). CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife* 6:e25776. doi: 10.7554/eLife.25776.001

Hanssen, L. L. P., Kassouf, M. T., Oudelaar, A. M., Biggs, D., Preece, C., Downes, D. J., et al. (2017). Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nat. Cell Biol.* 19, 952–961. doi: 10.1038/ncb3573

Harmston, N., Ing-Simmons, E., Tan, G., Perry, M., Merkenschlager, M., and Lenhard, B. (2017). Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. *Nat. Commun.* 8:441. doi: 10.1038/s41467-017-00524-5

Heinz, S., Romanoski, C. E., Benner, C., Allison, K. A., Kaikkonen, M. U., Orozco, L. D., et al. (2013). Effect of natural genetic variation on enhancer selection and function. *Nature* 503, 487–492. doi: 10.1038/nature12615

Henninger, J. E., Oksuz, O., Shrinivas, K., Sagi, I., LeRoy, G., Zheng, M. M., et al. (2021). RNA-Mediated Feedback Control of Transcriptional Condensates. *Cell* 184, 207–225.e24. doi: 10.1016/j.cell.2020.11.030

Hildebrand, E. M., and Dekker, J. (2020). *Mechanisms and Functions of Chromosome Compartmentalization*. Netherlands: Elsevier Ltd.

Hnisz, D., Day, D. S., and Young, R. A. (2016a). *Insulated Neighborhoods: structural and Functional Units of Mammalian Gene Control*. United States: Cell Press.

Hnisz, D., Weintraub, A. S., Day, D. S., Valton, A. L., Bak, R. O., Li, C. H., et al. (2016b). Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* 351, 1454–1458. doi: 10.1126/science.aad9024

Hochschild, A., and Ptashne, M. (1986). Homologous interactions of lambda repressor and lambda Cro with the lambda operator. *Cell* 44, 925–933. doi: 10.1016/0092-8674(86)90015-2

Horakova, A. H., Moseley, S. C., McLaughlin, C. R., Tremblay, D. C., and Chadwick, B. P. (2012). The macrosatellite DXZ4 mediates CTCF-dependent long-range intrachromosomal interactions on the human inactive X chromosome. *Hum. Mol. Genet.* 21, 4367–4377. doi: 10.1093/hmg/dds270

Huang, Y., Koues, O. I., Zhao, J. Y., Liu, R., Pyfrom, S. C., Payton, J. E., et al. (2017). cis-Regulatory Circuits Regulating NEK6 Kinase Overexpression in Transformed B Cells Are Super-Enhancer Independent. *Cell Rep.* 18, 2918–2931. doi: 10.1016/j.celrep.2017.02.067

Hug, C. B., Grimaldi, A. G., Kruse, K., and Vaquerizas, J. M. (2017). Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell* 169, 216–228.e19. doi: 10.1016/j.cell.2017.03.024

Iborra, F. J., Pombo, A., Jackson, D. A., and Cook, P. R. (1996). Active RNA polymerases are localized within discrete transcription "factories" in human nuclei. *J. Cell. Sci.* 361, 412–415.

Ibrahim, D. M., and Mundlos, S. (2020). *Three-Dimensional Chromatin in Disease: what Holds us Together and What Drives Us Apart?*. Netherlands: Elsevier Ltd.

Ing-Simmons, E., Vaid, R., Bing, X. Y., Levine, M., Mannervik, M., and Vaquerizas, J. M. (2021). Independence of chromatin conformation and gene regulation during Drosophila dorsoventral patterning. *Nat. Genet.* 53, 487–499. doi: 10.1038/s41588-021-00799-x

Isoda, T., Moore, A. J., He, Z., Chandra, V., Aida, M., Denholtz, M., et al. (2017). Non-coding Transcription Instructs Chromatin Folding and Compartmentalization to Dictate Enhancer-Promoter Communication and T Cell Fate. *Cell* 171, 103–119.e18. doi: 10.1016/j.cell.2017.09.001

Jackson, D. A., McCready, S. J., and Cook, P. R. (1981). RNA is synthesized at the nuclear cage. *Nature* 292, 552–555. doi: 10.1038/292552a0

Javierre, B. M., Burren, O. S., Wilder, S. P., Kreuzhuber, R., Hill, S. M., Sewitz, S., et al. (2016). Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* 167, 1369–1384.e19. doi: 10.1016/j.cell.2016.09.037

Jeong, Y., Leskow, F. C., El-Jaick, K., Roessler, E., Muenke, M., Yocum, A., et al. (2008). Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. *Nat. Genet.* 40, 1348–1353. doi: 10.1038/ng.230

Ji, X., Dadon, D. B., Powell, B. E., Fan, Z. P., Borges-Rivera, D., Shachar, S., et al. (2016). 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. *Cell Stem Cell* 18, 262–275. doi: 10.1016/j.stem.2015.11.007

Jia, Z., Li, J., Ge, X., Wu, Y., Guo, Y., and Wu, Q. (2020). Tandem CTCF sites function as insulators to balance spatial chromatin contacts and topological enhancer-promoter selection. *Genome Biol.* 21, 75–75. doi: 10.1186/s13059-020-01984-7

Jin, F., Li, Y., Dixon, J. R., Selvaraj, S., Ye, Z., Lee, A. Y., et al. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* 503, 290–294. doi: 10.1038/nature12644

Jung, Y. H., Sauria, M. E. G., Lyu, X., Cheema, M. S., Ausio, J., and Taylor, J. (2017). Chromatin States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes. *Cell Rep.* 18, 1366–1382. doi: 10.1016/j.celrep.2017.01.034

Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., van Berkum, N. L., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 430–435. doi: 10.1038/nature09380

Kaiser, V. B., and Semple, C. A. (2017). *When TADs Go Bad: chromatin Structure and Nuclear Organisation in Human Disease*. United Kingdom: Faculty of 1000 Ltd.

Katainen, R., Dave, K., Pitkänen, E., Palin, K., Kivioja, T., Välimäki, N., et al. (2015). CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat. Genet.* 47, 818–821. doi: 10.1038/ng.3335

Ke, Y., Xu, Y., Chen, X., Feng, S., Liu, Z., Sun, Y., et al. (2017). 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. *Cell* 170, 367–381.e20. doi: 10.1016/j.cell. 2017.06.029

Kim, Y., Shi, Z., Zhang, H., Finkelstein, I. J., and Yu, H. (2019). Human cohesin compacts DNA by loop extrusion. *Science* 366, 1345–1349. doi: 10.1126/science. aaz4475

Kishi, Y., and Gotoh, Y. (2018). *Regulation of Chromatin Structure During Neural Development*. Switzerland: Frontiers Media S.A.

Klopocki, E., Lohan, S., Brancati, F., Koll, R., Brehm, A., Seemann, P., et al. (2011). Copy-number variations involving the IHH locus are associated with syndactyly and craniosynostosis. *Am. J. Hum. Genet.* 88, 70–75. doi: 10.1016/j.ajhg.2010.11. 006

Kohwi, M., Lupton, J. R., Lai, S. L., Miller, M. R., and Doe, C. Q. (2013). Developmentally regulated subnuclear genome reorganization restricts neural progenitor competence in Drosophila. *Cell* 152, 97–108. doi: 10.1016/j.cell.2012. 11.049

Kojic, A., Cuadrado, A., De Koninck, M., Giménez-Llorente, D., Rodríguez-Corsino, M., Gómez-López, G., et al. (2018). Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization. *Nat. Struct. Mol. Biol.* 25, 496–504. doi: 10.1038/s41594-018-0070-4

Kolovos, P., Brouwer, R. W. W., Kockx, C. E. M., Lesnussa, M., Kepper, N., Zuin, J., et al. (2018). Investigation of the spatial structure and interactions of the genome at sub-kilobase-pair resolution using T2C. *Nat. Protoc.* 13, 459–477. doi: 10.1038/nprot.2017.132

Kolovos, P., Georgomanolis, T., Koeferle, A., Larkin, J. D., Brant, L., Nikolić, M., et al. (2016). Binding of nuclear factor κB to noncanonical consensus sites reveals its multimodal role during the early inflammatory response. *Genome Res.* 26, 1478–1489. doi: 10.1101/gr.210005.116

Kolovos, P., Knoch, T. A., Grosveld, F. G., Cook, P. R., and Papantonis, A. (2012). Enhancers and silencers: an integrated and simple model for their function. *Epigenetics Chromatin.* 5:1. doi: 10.1186/1756-8935-5-1

Kolovos, P., van de Werken, H. J., Kepper, N., Zuin, J., Brouwer, R. W., Kockx, C. E., et al. (2014). Targeted Chromatin Capture (T2C): a novel high resolution high throughput method to detect genomic interactions and regulatory elements. *Epigenetics Chromatin.* 7:10. doi: 10.1186/1756-8935-7-10

Kosak, S. T., Skok, J. A., Medina, K. L., Riblet, R., Le Beau, M. M., Fisher, A. G., et al. (2002). Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* 296, 158–62. doi: 10.1126/science.1068768

Kraft, K., Magg, A., Heinrich, V., Riemenschneider, C., Schöpflin, R., Markowski, J., et al. (2019). *Serial Genomic Inversions Induce Tissue-specific Architectural Stripes, Gene Misexpression and Congenital Malformations*. Germany: Nature Publishing Group.

Kragesteen, B. K., Spielmann, M., Paliou, C., Heinrich, V., Schöpflin, R., Esposito, A., et al. (2018). Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. *Nat. Genet.* 50, 1463–1473. doi: 10.1038/ s41588-018-0221-x

Krefting, J., Andrade-Navarro, M. A., and Ibn-Salem, J. (2018). Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biol.* 16:87. doi: 10.1186/s12915-018-0 556-x

Krumm, A., and Duan, Z. (2019). *Understanding the 3D Genome: emerging Impacts on Human Disease*. Netherlands: Elsevier Ltd.

Laghmach, R., and Potoyan, D. A. (2021). Liquid-liquid phase separation driven compartmentalization of reactive nucleoplasm. *Phys. Biol.* 18:015001. doi: 10. 1088/1478-3975/abc5ad

Lanctôt, C., Cheutin, T., Cremer, M., Cavalli, G., and Cremer, T. (2007). Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat. Rev. Genet.* 8, 104–115. doi: 10.1038/nrg2041

Larson, A. G., Elnatan, D., Keenen, M. M., Trnka, M. J., Johnston, J. B., Burlingame, A. L., et al. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* 547, 236–240. doi: 10.1038/nature22822

Laugsch, M., Bartusel, M., Rehimi, R., Alirzayeva, H., Karaolidou, A., Crispatzu, G., et al. (2019). Modeling the Pathological Long-Range Regulatory Effects of Human Structural Variation with Patient-Specific hiPSCs. *Cell Stem Cell* 24, 736–752.e12. doi: 10.1016/j.stem.2019.03.004

Lee, J. T., and Bartolomei, M. S. (2013). X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* 152, 1308–1323. doi: 10.1016/j. cell.2013.02.016

Lee, J. T., Davidow, L. S., and Warshawsky, D. (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat. Genet.* 21, 400–404. doi: 10.1038/7734

Lee, K., Hsiung, C. C., Huang, P., Raj, A., and Blobel, G. A. (2015). Dynamic enhancer-gene body contacts during transcription elongation. *Genes Dev.* 29, 1992–1997. doi: 10.1101/gad.255265.114

Lettice, L. A., Daniels, S., Sweeney, E., Venkataraman, S., Devenney, P. S., Gautier, P., et al. (2011). Enhancer-adoption as a mechanism of human developmental disease. *Hum. Mutat.* 32, 1492–1499. doi: 10.1002/humu.21615

Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., et al. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* 12, 1725–1735. doi: 10.1093/hmg/ddg180

Lettice, L. A., Williamson, I., Wiltshire, J. H., Peluso, S., Devenney, P. S., Hill, A. E., et al. (2012). Opposing functions of the ETS factor family define Shh spatial expression in limb buds and underlie polydactyly. *Dev. Cell.* 22, 459–467. doi: 10.1016/j.devcel.2011.12.010

Li, F., An, Z., and Zhang, Z. (2019). The Dynamic 3D Genome in Gametogenesis and Early Embryonic Development. *Cells* 8:788. doi: 10.3390/cells8080788

Li, G., Ruan, X., Auerbach, R. K., Sandhu, K. S., Zheng, M., Wang, P., et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 148, 84–98. doi: 10.1016/j.cell.2011.12.014

Li, L., Lyu, X., Hou, C., Takenaka, N., Nguyen, H. Q., Ong, C. T., et al. (2015). Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing. *Mol. Cell.* 58, 216–231. doi: 10.1016/j.molcel. 2015.02.023

Li, P., Banjade, S., Cheng, H. C., Kim, S., Chen, B., Guo, L., et al. (2012). Phase transitions in the assembly of multivalent signalling proteins. *Nature* 483, 336–340. doi: 10.1038/nature10879

Li, X., Zhou, B., Chen, L., Gou, L. T., Li, H., and Fu, X. D. (2017). GRID-seq reveals the global RNA-chromatin interactome. *Nat. Biotechnol.* 35, 940–950. doi: 10.1038/nbt.3968

Li, Y., Haarhuis, J. H. I, Sedeño, C. Á, Oldenkamp, R., van Ruiten, M. S., Willems, L., et al. (2020). The structural basis for cohesin–CTCF-anchored loops. *Nature* 578, 472–476. doi: 10.1038/s41586-019-1910-z

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293. doi: 10.1126/science.1181369

Lonfat, N., and Duboule, D. (2015). Structure, function and evolution of topologically associating domains (TADs) at HOX loci. *FEBS Lett.* 589, 2869–2876. doi: 10.1016/j.febslet.2015.04.024

Lonfat, N., Montavon, T., Darbellay, F., Gitto, S., and Duboule, D. (2014). Convergent evolution of complex regulatory landscapes and pleiotropy at Hox loci. *Science* 346, 1004–1006. doi: 10.1126/science.1257493

Lu, H., Yu, D., Hansen, A. S., Ganguly, S., Liu, R., Heckert, A., et al. (2018). Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature* 558, 318–323. doi: 10.1038/s41586-018-0174-3

Lu, L., Liu, X., Huang, W. K., Giusti-Rodríguez, P., Cui, J., Zhang, S., et al. (2020). Robust Hi-C Maps of Enhancer-Promoter Interactions Reveal the Function of Non-coding Genome in Neural Development and Diseases. *Mol. Cell.* 79, 521.e–534.e. doi: 10.1016/j.molcel.2020.06.007

Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., and Klopocki, E. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 161, 1012–1025. doi: 10.1016/j.cell. 2015.04.004

Lupiáñez, D. G., Spielmann, M., and Mundlos, S. (2016). Breaking TADs: how Alterations of Chromatin Domains Result in Disease. *Trends Genet.* 32, 225–237. doi: 10.1016/j.tig.2016.01.003

Meddens, C. A., Harakalova, M., van den Dungen, N. A. M., Foroughi Asl, H., Hijma, H. J., Cuppen, E. P. J. G., et al. (2016). Systematic analysis of chromatin interactions at disease associated loci links novel candidate genes to inflammatory bowel disease. *Genome Biol.* 17, 247–247. doi: 10.1186/s13059-016-1100-3

Melamed, A., Yaguchi, H., Miura, M., Witkover, A., Fitzgerald, T. W., Birney, E., et al. (2018). The human leukemia virus HTLV-1 alters the structure and transcription of host chromatin in cis. *Elife* 7:e36245. doi: 10.7554/eLife.36245

Melcer, S., and Meshorer, E. (2010). Chromatin plasticity in pluripotent cells. *Essays Biochem.* 48, 245–262. doi: 10.1042/BSE0480245

Mitchell, J. A., Clay, I., Umlauf, D., Chen, C. Y., Moir, C. A., Eskiw, C. H., et al. (2012). Nuclear RNA sequencing of the mouse erythroid cell transcriptome. *PLoS One* 7:e49274. doi: 10.1371/journal.pone.0049274

Monahan, K., Horta, A., and Lomvardas, S. (2019). LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. *Nature* 565, 448–453. doi: 10.1038/s41586-018-0845-0

Morgan, S. L., Mariano, N. C., Bermudez, A., Arruda, N. L., Wu, F., Luo, Y., et al. (2017). Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nat. Commun.* 8:15993. doi: 10.1038/ncomms15993

Müeller-Storm, H. P., Sogo, J. M., and Schaffner, W. (1989). An enhancer stimulates transcription in trans when attached to the promoter via a protein bridge. *Cell* 58, 767–777. doi: 10.1016/0092-8674(89)90110-4

Mumbach, M. R., Satpathy, A. T., Boyle, E. A., Dai, C., Gowen, B. G., Cho, S. W., et al. (2017). Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.* 49, 1602–1612. doi: 10.1038/ng.3963

Mylona, A., Andrieu-Soler, C., Thongjuea, S., Martella, A., Soler, E., Jorna, R., et al. (2013). Genome-wide analysis shows that Ldb1 controls essential hematopoietic genes/pathways in mouse early development and reveals novel players in hematopoiesis. *Blood* 121, 2902–2913. doi: 10.1182/blood-2012-11-467654

Nair, S. J., Yang, L., Meluzzi, D., Oh, S., Yang, F., Friedman, M. J., et al. (2019). Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly. *Nat. Struct. Mol. Biol.* 26, 193–203. doi: 10.1038/s41594-019-0190-5

Narendra, V., Rocha, P. P., An, D., Raviram, R., Skok, J. A., Mazzoni, E. O., et al. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* 347, 1017–1021. doi: 10.1126/science.1262088

Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B. R., Mirny, L. A., et al. (2013). Organization of the mitotic chromosome. *Science* 342, 948–53. doi: 10.1126/science.1236083

Ngan, C. Y., Wong, C. H., Tjong, H., Wang, W., Goldfeder, R. L., Choi, C., et al. (2020). Chromatin interaction analyses elucidate the roles of PRC2-bound silencers in mouse development. *Nat. Genet.* 52, 264–272. doi: 10.1038/s41588-020-0581-x

Ngcungcu, T., Oti, M., Sitek, J. C., Haukanes, B. I., Linghu, B., Bruccoleri, R., et al. (2017). Duplicated Enhancer Region Increases Expression of CTSB and Segregates with Keratolytic Winter Erythema in South African and Norwegian Families. *Am. J. Hum. Genet.* 100, 737–750. doi: 10.1016/j.ajhg.2017.03.012

Noordermeer, D., and de Laat, W. (2008). Joining the loops: beta-globin gene regulation. *IUBMB Life* 60, 824–833. doi: 10.1002/iub.129

Nora, E. P., Goloborodko, A., Valton, A. L., Gibcus, J. H., Uebersohn, A., Abdennur, N., et al. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* 169, 930–944.e22. doi: 10.1016/j.cell.2017.05.004

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385. doi: 10.1038/nature11049

Northcott, P. A., Lee, C., Zichner, T., Stütz, A. M., Erkek, S., Kawauchi, D., et al. (2014). Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* 511, 428–434. doi: 10.1038/nature13379

Norton, H. K., and Phillips-Cremins, J. E. (2017). *Crossed Wires: 3D Genome Misfolding in Human Disease*. United States: Rockefeller University Press.

Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N., and Mirny, L. A. (2018). Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl. Acad. Sci. U. S. A.* 115, E6697–E6706. doi: 10.1073/pnas.1717730115

Olley, G., Ansari, M., Bengani, H., Grimes, G. R., Rhodes, J., Von Kriegsheim, A., et al. (2018). BRD4 interacts with NIPBL and BRD4 is mutated in a Cornelia de Lange-like syndrome. *Nat. Genet.* 50, 329–332. doi: 10.1038/s41588-018-0042-y

Ong, C. T., and Corces, V. G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.* 12, 283–293. doi: 10.1038/nrg2957

Paliou, C., Guckelberger, P., Schöpflin, R., Heinrich, V., Esposito, A., Chiariello, A. M., et al. (2019). Preformed chromatin topology assists transcriptional robustness of Shh during limb development. *Proc. Natl. Acad. Sci. U. S. A.* 116, 12390–12399. doi: 10.1073/pnas.1900672116

Palstra, R. J., de Laat, W., and Grosveld, F. (2008a). Beta-globin regulation and long-range interactions. *Adv. Genet.* 61, 107–142. doi: 10.1016/S0065-2660(07)00004-1

Palstra, R. J., Simonis, M., Klous, P., Brasset, E., Eijkelkamp, B., and de Laat, W. (2008b). Maintenance of long-range DNA interactions after inhibition of ongoing RNA polymerase II transcription. *PLoS One* 3:e1661. doi: 10.1371/journal.pone.0001661

Papantonis, A., and Cook, P. R. (2013). Transcription factories: genome organization and gene regulation. *Chem. Rev.* 113, 8683–8705. doi: 10.1021/cr300513p

Papantonis, A., Kohro, T., Baboo, S., Larkin, J. D., Deng, B., Short, P., et al. (2012). TNFα signals through specialized factories where responsive coding and miRNA genes are transcribed. *EMBO J.* 31, 4404–4414. doi: 10.1038/emboj.2012.288

Papantonis, A., Larkin, J. D., Wada, Y., Ohta, Y., Ihara, S., Kodama, T., et al. (2010). Active RNA polymerases: mobile or immobile molecular machines? *PLoS Biol.* 8:e1000419. doi: 10.1371/journal.pbio.1000419

Patel, L., Kang, R., Rosenberg, S. C., Qiu, Y., Raviram, R., Chee, S., et al. (2019). Dynamic reorganization of the genome shapes the recombination landscape in meiotic prophase. *Nat. Struct. Mol. Biol.* 26, 164–174. doi: 10.1038/s41594-019-0187-0

Peifer, M., Hertwig, F., Roels, F., Dreidax, D., Gartlgruber, M., Menon, R., et al. (2015). Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature* 526, 700–704. doi: 10.1038/nature14980

Peñalosa-Ruiz, G., Bright, A. R., Mulder, K. W., and Veenstra, G. J. C. (2019). *The Interplay of Chromatin and Transcription Factors During Cell Fate Transitions in Development and Reprogramming*. Netherlands: Elsevier B.V.

Phanstiel, D. H., Van Bortle, K., Spacek, D., Hess, G. T., Shamim, M. S., Machol, I., et al. (2017). Static and Dynamic DNA Loops form AP-1-Bound Activation Hubs during Macrophage Development. *Mol. Cell.* 67, 1037–1048.e6. doi: 10.1016/j.molcel.2017.08.006

Phillips-Cremins, J. E., Sauria, M. E., Sanyal, A., Gerasimova, T. I., Lajoie, B. R., Bell, J. S., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* 153, 1281–1295. doi: 10.1016/j.cell.2013.04.053

Pimkin, M., Kossenkov, A. V., Mishra, T., Morrissey, C. S., Wu, W., Keller, C. A., et al. (2014). Divergent functions of hematopoietic transcription factors in lineage priming and differentiation during erythro-megakaryopoiesis. *Genome Res.* 24, 1932–1944. doi: 10.1101/gr.164478.113

Pombo, A., Jackson, D. A., Hollinshead, M., Wang, Z., Roeder, R. G., and Cook, P. R. (1999). Regional specialization in human nuclei: visualization of discrete sites of transcription by RNA polymerase III. *EMBO J.* 18, 2241–2253. doi: 10.1093/emboj/18.8.2241

Quinodoz, S. A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J. M., Detmar, E., et al. (2018). Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* 174, 744–757.e24. doi: 10.1016/j.cell.2018.05.024

Ramírez, F., Bhardwaj, V., Arrigoni, L., Lam, K. C., Grüning, B. A., Villaveces, J., et al. (2018). High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* 9, 1–15. doi: 10.1038/s41467-017-02525-w

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–2308. doi: 10.1038/nprot.2013.143

Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680. doi: 10.1016/j.cell.2014.11.021

Rao, S. S. P., Huang, S. C., Glenn St Hilaire, B., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K. R., et al. (2017). Cohesin Loss Eliminates All Loop Domains. *Cell* 171, 305–320.e24. doi: 10.1016/j.cell.2017.09.026

Redin, C., Brand, H., Collins, R. L., Kammin, T., Mitchell, E., Hodge, J. C., et al. (2017). The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nat. Genet.* 49, 36–45. doi: 10.1038/ng.3720

Rickman, D. S., Soong, T. D., Moss, B., Mosquera, J. M., Dlabal, J., Terry, S., et al. (2012). Oncogene-mediated alterations in chromatin conformation. *Proc. Natl. Acad. Sci. U. S. A.* 109, 9083–9088. doi: 10.1073/pnas.1112570109

Robson, M. I., Ringel, A. R., and Mundlos, S. (2019). *Regulatory Landscaping: how Enhancer-Promoter Communication Is Sculpted in 3D*. United States: Cell Press.

Rodríguez-Carballo, E., Lopez-Delisle, L., Zhan, Y., Fabre, P. J., Beccari, L., El-Idrissi, I., et al. (2017). The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. *Genes Dev.* 31, 2264–2281. doi: 10.1101/gad.307769.117

Rosa-Garrido, M., Chapski, D. J., Schmitt, A. D., Kimball, T. H., Karbassi, E., Monte, E., et al. (2017). High-resolution mapping of chromatin conformation in cardiac myocytes reveals structural remodeling of the epigenome in heart failure. *Circulation* 136, 1613–1625. doi: 10.1161/CIRCULATIONAHA.117.029430

Rousseau, M., Ferraiuolo, M. A., Crutchley, J. L., Wang, X. Q., Miura, H., Blanchette, M., et al. (2014). Classifying leukemia types with chromatin conformation data. *Genome Biol.* 15:R60. doi: 10.1186/gb-2014-15-4-r60

Rowley, M. J., and Corces, V. G. (2018). Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* 19, 789–800. doi: 10.1038/s41576-018-0060-8

Rowley, M. J., Nichols, M. H., Lyu, X., Ando-Kuri, M., Rivera, I. S. M., Hermetz, K., et al. (2017). Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol. Cell* 67, 837–852.e7. doi: 10.1016/j.molcel.2017.07.022

Rubin, A. J., Barajas, B. C., Furlan-Magaril, M., Lopez-Pajares, V., Mumbach, M. R., Howard, I., et al. (2017). Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nat. Genet.* 49, 1522–1528. doi: 10.1038/ng.3935

Rudan, M. V., Barrington, C., Henderson, S., Ernst, C., Odom, D. T., Tanay, A., et al. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell. Rep.* 10, 1297–1309. doi: 10.1016/j.celrep.2015.02.004

Sabari, B. R., Dall'Agnese, A., Boija, A., Klein, I. A., Coffey, E. L., Shrinivas, K., et al. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. *Science* 361:eaar3958. doi: 10.1126/science.aar3958

Saldaña-Meyer, R., Rodriguez-Hernaez, J., Escobar, T., Nishana, M., Jácome-López, K., Nora, E. P., et al. (2019). RNA Interactions Are Essential for CTCF-Mediated Genome Organization. *Mol. Cell.* 76, 412–422.e5. doi: 10.1016/j.molcel.2019.08.015

Sanborn, A. L., Rao, S. S., Huang, S. C., Durand, N. C., Huntley, M. H., Jewett, A. I., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U. S. A.* 112, E6456–E6465. doi: 10.1073/pnas.1518552112

Sauerwald, N., and Kingsford, C. (2018). Quantifying the similarity of topological domains across normal and cancer human cell types. *Bioinformatics* 34, i475–i483. doi: 10.1093/bioinformatics/bty265

Schmidt, D., Schwalie, P. C., Wilson, M. D., Ballester, B., Gonalves, Â, Kutter, C., et al. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* 148, 335–348. doi: 10.1016/j.cell.2011.11.058

Schmitt, A. D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C. L., et al. (2016). A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell. Rep.* 11, 2042–2059. doi: 10.1016/j.celrep.2016.10.061

Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* 20, 437–455. doi: 10.1038/s41576-019-0128-0

Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B. M., et al. (2015a). The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* 25, 582–597. doi: 10.1101/gr.185272.114

Schoenfelder, S., Sugar, R., Dimond, A., Javierre, B. M., Armstrong, H., Mifsud, B., et al. (2015b). Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat. Genet.* 47, 1179–1186. doi: 10.1038/ng.3393

Schoenfelder, S., Mifsud, B., Senner, C. E., Todd, C. D., Chrysanthou, S., Darbo, E., et al. (2018). Divergent wiring of repressive and active chromatin interactions between mouse embryonic and trophoblast lineages. *Nat. Commun.* 9:4189. doi: 10.1038/s41467-018-06666-4

Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N. F., Horton, A., Andrews, S., et al. (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat. Genet.* 42, 53–61. doi: 10.1038/ng.496

Schuijers, J., Manteiga, J. C., Weintraub, A. S., Day, D. S., Zamudio, A. V., Hnisz, D., et al. (2018). Transcriptional Dysregulation of MYC Reveals Common Enhancer-Docking Mechanism. *Cell. Rep.* 23, 349–360. doi: 10.1016/j.celrep.2018.03.056

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., et al. (2017). Two independent modes of chromatin organization revealed by cohesin removal. *Nature* 551, 51–56. doi: 10.1038/nature24281

Sexton, T., and Cavalli, G. (2015). The role of chromosome domains in shaping the functional genome. *Cell* 160, 1049–1059. doi: 10.1016/j.cell.2015.02.040

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., et al. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* 148, 458–472. doi: 10.1016/j.cell.2012.01.010

Shachar, S., and Misteli, T. (2017). *Causes and Consequences of Nuclear Gene Positioning*. United Kingdom: Company of Biologists Ltd.

Shachar, S., Voss, T. C., Pegoraro, G., Sciascia, N., and Misteli, T. (2015). Identification of Gene Positioning Factors Using High-Throughput Imaging Mapping. *Cell* 162, 911–923. doi: 10.1016/j.cell.2015.07.035

Shin, Y., Chang, Y. C., Lee, D. S. W., Berry, J., Sanders, D. W., Ronceray, P., et al. (2018). Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. *Cell* 175, 1481–1491.e13. doi: 10.1016/j.cell.2018.10.057

Shrinivas, K., Sabari, B. R., Coffey, E. L., Klein, I. A., Boija, A., Zamudio, A. V., et al. (2019). Enhancer Features that Drive Formation of Transcriptional Condensates. *Mol. Cell.* 75, 549–561.e7. doi: 10.1016/j.molcel.2019.07.009

Siersbæk, R., Madsen, J. G. S., Javierre, B. M., Nielsen, R., Bagge, E. K., Cairns, J., et al. (2017). Dynamic Rewiring of Promoter-Anchored Chromatin Loops during Adipocyte Differentiation. *Mol. Cell.* 66, 420–435.e5. doi: 10.1016/j.molcel.2017.04.010

Simon, M. D., Wang, C. I., Kharchenko, P. V., West, J. A., Chapman, B. A., Alekseyenko, A. A., et al. (2011). The genomic binding sites of a noncoding RNA. *Proc. Natl. Acad. Sci. U. S. A.* 108, 20497–20502. doi: 10.1073/pnas.1113536108

Smemo, S., Campos, L. C., Moskowitz, I. P., Krieger, J. E., Pereira, A. C., and Nobrega, M. A. (2012). Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Hum. Mol. Genet.* 21, 3255–3263. doi: 10.1093/hmg/dds165

Smemo, S., Tena, J. J., Kim, K. H., Gamazon, E. R., Sakabe, N. J., Gómez-Marín, C., et al. (2014). Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* 507, 371–375. doi: 10.1038/nature13138

Sofueva, S., Yaffe, E., Chan, W. C., Georgopoulou, D., Vietri Rudan, M., Mira-Bontenbal, H., et al. (2013). Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* 32, 3119–3129. doi: 10.1038/emboj.2013.237

Soldner, F., Stelzer, Y., Shivalila, C. S., Abraham, B. J., Latourelle, J. C., Barrasa, M. I., et al. (2016). Parkinson-associated risk variant in distal enhancer of α-synuclein modulates target gene expression. *Nature* 533, 95–99. doi: 10.1038/nature17939

Spielmann, M., Brancati, F., Krawitz, P. M., Robinson, P. N., Ibrahim, D. M., Franke, M., et al. (2012). Homeotic arm-to-leg transformation associated with genomic rearrangements at the PITX1 locus. *Am. J. Hum. Genet.* 91, 629–635. doi: 10.1016/j.ajhg.2012.08.014

Spielmann, M., Lupiáñez, D. G., and Mundlos, S. (2018). Structural variation in the 3D genome. *Nat. Rev. Genet.* 19, 453–467. doi: 10.1038/s41576-018-0007-0

Spielmann, M., and Mundlos, S. (2016). *Looking Beyond the Genes: the role of Non-coding Variants in Human Disease*. Oxford: Oxford University Press.

Splinter, E., de Wit, E., Nora, E. P., Klous, P., van de Werken, H. J., Zhu, Y., et al. (2011). The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes Dev.* 25, 1371–1383. doi: 10.1101/gad.633311

Splinter, E., Heath, H., Kooren, J., Palstra, R. J., Klous, P., Grosveld, F., et al. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* 20, 2349–2354. doi: 10.1101/gad.399506

Stadhouders, R., Aktuna, S., Thongjuea, S., Aghajanirefah, A., Pourfarzad, F., van Ijcken, W., et al. (2014). HBS1L-MYB intergenic variants modulate fetal hemoglobin via long-range MYB enhancers. *J. Clin. Invest.* 124, 1699–1710. doi: 10.1172/JCI71520

Stadhouders, R., Filion, G. J., and Graf, T. (2019). *Transcription Factors and 3D Genome Conformation in Cell-fate Decisions*. Germany: Nature Publishing Group.

Stadhouders, R., Thongjuea, S., Andrieu-Soler, C., Palstra, R. J., Bryne, J. C., van den Heuvel, A., et al. (2012). Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development. *EMBO J.* 31, 986–999. doi: 10.1038/emboj.2011.450

Stadhouders, R., Vidal, E., Serra, F., Di Stefano, B., Le Dily, F., Quilez, J., et al. (2018). Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. *Nat. Genet.* 50, 238–249. doi: 10.1038/s41588-017-0030-7

Strom, A. R., Emelyanov, A. V., Mir, M., Fyodorov, D. V., Darzacq, X., and Karpen, G. H. (2017). Phase separation drives heterochromatin domain formation. *Nature* 547, 241–245. doi: 10.1038/nature22989

Sun, F., Chronis, C., Kronenberg, M., Chen, X. F., Su, T., Lay, F. D., et al. (2019). Promoter-Enhancer Communication Occurs Primarily within Insulated Neighborhoods. *Mol. Cell* 73, 250–263.e5. doi: 10.1016/j.molcel.2018.10.039

Swaminathan, G. J., Bragin, E., Chatzimichali, E. A., Corpas, M., Bevan, A. P., Wright, C. F., et al. (2012). DECIPHER: web-based, community resource for clinical interpretation of rare variants in developmental disorders. *Hum. Mol. Genet.* 21, R37–R44. doi: 10.1093/hmg/dds362

Symmons, O., Pan, L., Remeseiro, S., Aktas, T., Klein, F., Huber, W., et al. (2016). The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances. *Dev. Cell* 39, 529–543. doi: 10.1016/j.devcel.2016.10.015

Symmons, O., Uslu, V. V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., et al. (2014). Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* 24, 390–400. doi: 10.1101/gr.163519.113

Szabo, Q., Bantignies, F., and Cavalli, G. (2019). Principles of genome folding into topologically associating domains. *Sci. Adv.* 5:eaaw1668. doi: 10.1126/sciadv.aaw1668

Taberlay, P. C., Achinger-Kawecka, J., Lun, A. T., Buske, F. A., Sabir, K., Gould, C. M., et al. (2016). Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. *Genome Res.* 26, 719–731. doi: 10.1101/gr.201517.115

Tan, L., Ma, W., Wu, H., Zheng, Y., Xing, D., Chen, R., et al. (2021). Changes in genome architecture and transcriptional dynamics progress independently of sensory experience during post-natal brain development. *Cell* 184, 741–758.e17. doi: 10.1016/j.cell.2020.12.032

Tan, L., Xing, D., Chang, C. H., Li, H., and Xie, X. S. (2018). Three-dimensional genome structures of single diploid human cells. *Science* 361, 924–928. doi: 10.1126/science.aat5641

Tayebi, N., Jamsheer, A., Flöttmann, R., Sowinska-Seidler, A., Doelken, S. C., Oehl-Jaschkowitz, B., et al. (2014). Deletions of exons with regulatory activity at the DYNC1I1 locus are associated with split-hand/split-foot malformation: array CGH screening of 134 unrelated families. *Orphanet J. Rare Dis.* 9, 108–108. doi: 10.1186/s13023-014-0108-6

Therizols, P., Illingworth, R. S., Courilleau, C., Boyle, S., Wood, A. J., and Bickmore, W. A. (2014). Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science* 346, 1238–1242. doi: 10.1126/science.1259587

Thiecke, M. J., Wutz, G., Muhar, M., Tang, W., Bevan, S., Malysheva, V., et al. (2020). Cohesin-Dependent and -Independent Mechanisms Mediate Chromosomal Contacts between Promoters and Enhancers. *Cell Rep.* 32, 107929–107929. doi: 10.1016/j.celrep.2020.107929

Trimborn, T., Gribnau, J., Grosveld, F., and Fraser, P. (1999). Mechanisms of developmental control of transcription in the murine alpha- and beta-globin loci. *Genes Dev.* 13, 112–124. doi: 10.1101/gad.13.1.112

Tsujimura, T., Klein, F. A., Langenfeld, K., Glaser, J., Huber, W., and Spitz, F. (2015). A discrete transition zone organizes the topological and regulatory autonomy of the adjacent tfap2c and bmp7 genes. *PLoS Genet.* 11:e1004897. doi: 10.1371/journal.pgen.1004897

Ulianov, S. V., Velichko, A. K., Magnitov, M. D., Luzhin, A. V., Golov, A. K., Ovsyannikova, N., et al. (2021). Suppression of liquid-liquid phase separation by 1,6-hexanediol partially compromises the 3D genome organization in living cells. *Nucleic Acids Res.* 9:gkab249. doi: 10.1093/nar/gkab249

Uslu, V. V., Petretich, M., Ruf, S., Langenfeld, K., Fonseca, N. A., Marioni, J. C., et al. (2014). Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. *Nat. Genet.* 46, 753–758. doi: 10.1038/ng.2971

van Steensel, B., and Belmont, A. S. (2017). *Lamina-Associated Domains: links with Chromosome Architecture, Heterochromatin, and Gene Repression*. United States: Cell Press.

van Steensel, B., and Furlong, E. E. M. (2019). *The role of Transcription in Shaping the Spatial Organization of the Genome*. Germany: Nature Publishing Group.

Vilarrasa-Blasi, R., Soler-Vila, P., Verdaguer-Dot, N., Russiñol, N., Di Stefano, M., Chapaprieta, V., et al. (2021). Dynamics of genome architecture and chromatin function during human B cell differentiation and neoplastic transformation. *Nat. Commun.* 12, 1–18. doi: 10.1038/s41467-020-20849-y

Villar, D., Berthelot, C., Aldridge, S., Rayner, T. F., Lukk, M., Pignatelli, M., et al. (2015). Enhancer evolution across 20 mammalian species. *Cell* 160, 554–566. doi: 10.1016/j.cell.2015.01.006

Wang, S., Su, J. H., Beliveau, B. J., Bintu, B., Moffitt, J. R., Wu, C. T., et al. (2016). Spatial organization of chromatin domains and compartments in single chromosomes. *Science* 353, 598–602. doi: 10.1126/science.aaf8084

Wang, Y., Wang, H., Zhang, Y., Du, Z., Si, W., Fan, S., et al. (2019). Reprogramming of Meiotic Chromatin Architecture during Spermatogenesis. *Mol. Cell* 73, 547–561.e6. doi: 10.1016/j.molcel.2018.11.019

Weedon, M. N., Cebola, I., Patch, A. M., Flanagan, S. E., De Franco, E., Caswell, R., et al. (2014). Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat. Genet.* 46, 61–64. doi: 10.1038/ng.2826

Wei, Z., Gao, F., Kim, S., Yang, H., Lyu, J., An, W., et al. (2013). Klf4 organizes long-range chromosomal interactions with the OCT4 locus inreprogramming andpluripotency. *Cell Stem Cell* 13, 36–47. doi: 10.1016/j.stem.2013.05.010

Weischenfeldt, J., Dubash, T., Drainas, A. P., Mardin, B. R., Chen, Y., Stütz, A. M., et al. (2017). Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nat. Genet.* 49, 65–74. doi: 10.1038/ng.3722

Wendt, K. S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451, 796–801. doi: 10.1038/nature06634

Wijgerde, M., Grosveld, F., and Fraser, P. (1995). Transcription complex stability and chromatin dynamics in vivo. *Nature* 377, 209–213. doi: 10.1038/377209a0

Williamson, I., Lettice, L. A., Hill, R. E., and Bickmore, W. A. (2016). Shh and ZRS enhancer colocalisation is specific to the zone of polarising activity. *Development* 143, 2994–3001. doi: 10.1242/dev.139188

Woodcock, C. L., and Ghosh, R. P. (2010). Chromatin higher-order structure and dynamics. *Cold Spring Harb. Perspect. Biol.* 2:a000596. doi: 10.1101/cshperspect.a000596

Worman, H. J., and Bonne, G. (2007). *"Laminopathies": a Wide Spectrum of Human Diseases*. United States: Academic Press Inc.

Wu, P., Li, T., Li, R., Jia, L., Zhu, P., Liu, Y., et al. (2017). 3D genome of multiple myeloma reveals spatial genome disorganization associated with copy number variations. *Nat. Commun.* 8:1937. doi: 10.1038/s41467-017-01793-w

Wutz, G., Várnai, C., Nagasaka, K., Cisneros, D. A., Stocsits, R. R., Tang, W., et al. (2017). Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* 36, 3573–3599. doi: 10.15252/embj.201798004

Xie, W., Schultz, M. D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., et al. (2013). Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* 153, 1134–1148. doi: 10.1016/j.cell.2013.04.022

Xu, H., Zhang, S., Yi, X., Plewczynski, D., and Li, M. J. (2020). *Exploring 3D Chromatin Contacts in Gene Regulation: the Evolution of Approaches for the Identification of Functional Enhancer-Promoter Interaction*. Netherlands: Elsevier B.V.

Yang, F., Deng, X., Ma, W., Berletch, J. B., Rabaia, N., Wei, G., et al. (2015). The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. *Genome Biol.* 16:52. doi: 10.1186/s13059-015-0618-0

Zhan, Y., Mariani, L., Barozzi, I., Schulz, E. G., Blüthgen, N., Stadler, M., et al. (2017). Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes. *Genome Res.* 27, 479–490. doi: 10.1101/gr.212803.116

Zhang, N., Mendieta-Esteban, J., Magli, A., Lilja, K. C., Perlingeiro, R. C. R., Marti-Renom, M. A., et al. (2020). Muscle progenitor specification and

myogenic differentiation are associated with changes in chromatin topology. *Nat. Commun.* 11:6222. doi: 10.1038/s41467-020-19999-w

Zhang, X., Choi, P. S., Francis, J. M., Imielinski, M., Watanabe, H., Cherniack, A. D., et al. (2016). Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. *Nat. Genet.* 48, 176–182. doi: 10.1038/ng.3470

Zheng, H., and Xie, W. (2019). *The role of 3D Genome Organization in Development and Cell Differentiation*. Germany: Nature Publishing Group.

Zufferey, M., Tavernari, D., Oricchio, E., and Ciriello, G. (2018). Comparison of computational methods for the identification of topologically associating domains. *Genome Biol.* 19:217. doi: 10.1186/s13059-018-1596-9

Zuin, J., Dixon, J. R., van der Reijden, M. I., Ye, Z., Kolovos, P., Brouwer, R. W., et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. U. S. A.* 111, 996–1001. doi: 10.1073/pnas.1317788111

The Role of rDNA Clusters in Global Epigenetic Gene Regulation

Nickolai A. Tchurikov* and Yuri V. Kravatsky

Engelhardt Institute of Molecular Biology Russian Academy of Sciences, Moscow, Russia

The regulation of gene expression has been studied for decades, but the underlying mechanisms are still not fully understood. As well as local and distant regulation, there are specific mechanisms of regulation during development and physiological modulation of gene activity in differentiated cells. Current research strongly supports a role for the 3D chromosomal structure in the regulation of gene expression. However, it is not known whether the genome structure reflects the formation of active or repressed chromosomal domains or if these structures play a primary role in the regulation of gene expression. During early development, heterochromatinization of ribosomal DNA (rDNA) is coupled with silencing or activation of the expression of different sets of genes. Although the mechanisms behind this type of regulation are not known, rDNA clusters shape frequent inter-chromosomal contacts with a large group of genes controlling development. This review aims to shed light on the involvement of clusters of ribosomal genes in the global regulation of gene expression. We also discuss the possible role of RNA-mediated and phase-separation mechanisms in the global regulation of gene expression by nucleoli.

Keywords: rDNA, inter-chromosomal contacts, epigenetics, nucleoli, cancer, H3K27ac mark, super-enhancers, phase separation

INTRODUCTION

Nucleoli are the largest organelles in nuclei. They are not separated from chromosomes by any kind of membrane and potentially could shape contacts with chromosomal regions in interphase cells either without any particular order, or in some order to attain structural or functional features. If ordered, these contacts should be re-established in the course of cell division and epigenetic mechanisms may be involved. In interphase chromosomes, chromatin forms loops of different sizes that are required for both the compaction of chromosomes and for establishing a regulatory network. The close contacts between nucleoli and the chromosomal DNA were demonstrated by the co-isolation of chromosomal regions that are rather strongly attached to nucleoli during the isolation of nucleoli preparations (Németh et al., 2010; van Koningsbruggen et al., 2010). However, the size of the attached chromosomal DNA fragments (up to 1 Mb) did not allow a precise estimation of the contact sites of nucleoli in chromosomes or to determine their roles.

The Hi-C approach (Lieberman-Aiden et al., 2009) allows more precise mapping of the genome-wide chromatin contacts including those of the rDNA units. Using Hi-C or its derivative, the 4C (circular chromatin conformation capture) approach, it was possible to determine the rDNA contacts in human and *Drosophila* genomes. Of particular interest, the novel data suggested a role for nucleoli in differentiation. Localized heterochromatization of rDNA genes initiates the appearance of condensed chromatin structures in different genomic regions coupled with transcriptional activation of differentiation genes and the loss of pluripotency of embryonic stem cells (Feinberg, 2014; Savić et al., 2014). Modulating the rDNA expression fosters changes in the cell fate, growth, and proliferation of female *Drosophila*

ovarian germline stem cells and their daughters (Zhang et al., 2014). The mechanisms of the regulation of rDNA units and the factors involved are described in more detail in the recent review by Kresoja-Rakic and Santoro (2019).

There are two possible ways that rDNA units could modulate differentiation. The first is rDNA-mediated regulation by a remote mechanism that works at the level of unknown protein or RNA factors from active or silent rDNA units that initiate activation or silencing of different target genes. The second possible mechanism is the formation of dynamic direct contacts between rDNA units and different chromosomal regions that contain development-regulating genes. At present, both mechanisms should be considered. In this review, we discuss the recent data supporting the view that nucleoli are involved in the formation of 3D inter-chromosomal structures and that they shape contacts with different chromosomal genes, as well as the data on the role of phase-separation mechanisms in this type of regulation. We do not attempt to exhaustively review the literature and only refer to the main papers describing the most important ideas and findings in this area.
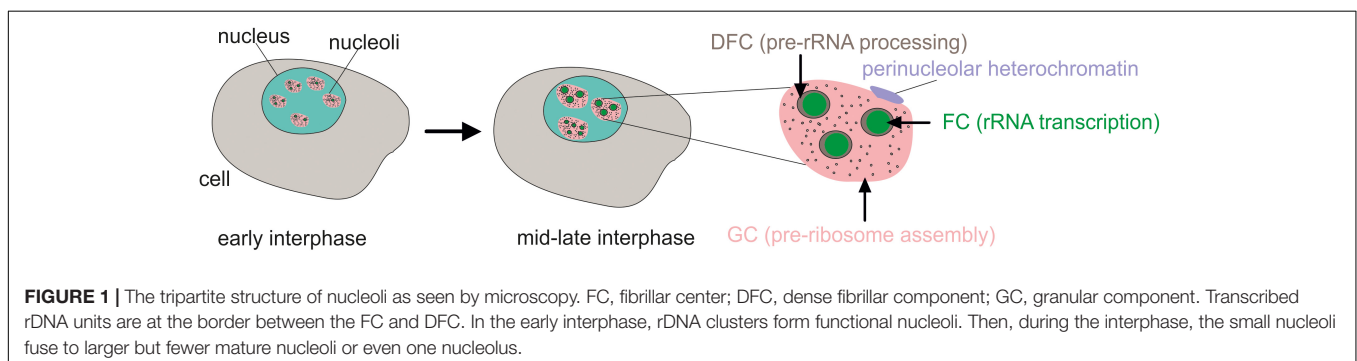
Genetic and Molecular Evidence of the Regulatory Role of Nucleoli

Nucleoli are the largest membrane-less organelles in the nucleus. By light and electron microscopy, the tripartite structure of nucleoli can be observed including the fibrillar center (FC), dense fibrillar component (DFC), and the granular component (GC) (**Figure 1**). The clusters of rDNA genes reside around the FC while at the border of the FC and DFC, the chromatin loops that contain rDNA units are transcribed (Tiku and Antebi, 2018). The processing of 47S pre-rRNA and ribosomal protein assembly occurs in the DFC, and then the assembly of pre-ribosomal subunits is performed in the GC (Granneman and Baserga, 2005). Pre-ribosomal particles are formed in the GC using 5S rRNA, which is synthesized by RNA polymerase III from independent genes outside of the nucleolus, and the ribosomal proteins, which are transported from the cytoplasm to the nucleolus (Baßler and Hurt, 2019). There are several dozen FC–DFC modules in each nucleolus in human cells (Lafontaine et al., 2021). The number of FC–DFC modules is relatively constant for a particular cell type but differs widely between cell types, making it a powerful biomarker for cell classification (Lafontaine et al., 2021).

Some rDNA clusters are silent and form constitutive heterochromatin and are not associated with nucleoli (Akhmanova et al., 2000). Active rDNA units, which represent about half of the rDNA copies, are bound with upstream binding factor (UBF) and form nucleoli. Some inactive rDNA copies in the nucleolus are attached to the periphery of the nucleolus and shape co-called perinucleolar heterochromatin (Lindström et al., 2018). During the cell cycle, nucleoli disassemble at the prophase stage and begin to reassemble during the telophase (Pederson, 2011).

One way to determine the global role of rDNA expression on cell function is using a genetic approach to change the level of transcription by damaging some components of the Pol I machinery, pre-rRNA processing, or ribosome assembly. The *Drosophila* Pol I regulatory complex includes Under-developed (Udd) and TAF1B factors. Damaging *udd* or *TAF1B* leads to a reduced number of germ stem cell clones that produce differentiating cysts over time (Zhang et al., 2014). Similarly, active rDNA expression delays the differentiation of ovarian germline stem cells, whereas reduced rRNA production induces morphological changes that accompany early germline differentiation. These findings demonstrate that modulating rRNA synthesis promotes changes in the cell fate, growth, and proliferation of female *Drosophila* germline stem cells. The underlying mechanisms are not known; however, it is speculated that changes in ribosome biogenesis lead to changes in the expression of specific proteins that direct cell fate decisions, growth, and proliferation within an *in vivo* stem cell lineage more rapidly or to a greater extent (Zhang et al., 2014).

Important evidence in favor of the regulatory role of nucleoli in differentiation was obtained during studies of mouse embryonic stem cells (ESCs). Topological-associated domains (TADs) in ESCs are similar in different cell types and the chromatin is generally less condensed (Dixon et al., 2012; Nora et al., 2012). During differentiation of ESCs in mammals and *Drosophila*, large-scale silencing occurs and highly condensed heterochromatin appears in different chromosomes, including the regions of heterochromatic centric and pericentric repeats (Bhattacharya et al., 2009). In the course of differentiation, condensed heterochromatin regions are formed inside particular subsets of rDNA clusters. The nucleolar repressor TIP5, in association with long non-coding pRNA (promoter RNA), transcribed from the intergenic spacer (IGS) of rDNA, and

FIGURE 1 | The tripartite structure of nucleoli as seen by microscopy. FC, fibrillar center; DFC, dense fibrillar component; GC, granular component. Transcribed rDNA units are at the border between the FC and DFC. In the early interphase, rDNA clusters form functional nucleoli. Then, during the interphase, the small nucleoli fuse to larger but fewer mature nucleoli or even one nucleolus.

some other factors are required for heterochromatin formation in some rDNA units (Santoro et al., 2010; Guetg et al., 2012). The analysis of the levels of pre-rRNA, rDNA methylation, and histone repressive marks in rDNA and satellites revealed that the formation of silenced rDNA units takes place during the transition from ESCs to neural progenitor cells and coincides with the switch to a more condensed heterochromatic form of centric and pericentric repeats (Savić et al., 2014). Interestingly, the silencing of particular rDNA units promotes the transcriptional activation or downregulation of hundreds of differentiation genes. These data suggest that nucleoli are involved in the regulation of chromatin states and the expression of genes associated with differentiation. The underlying mechanisms by which nucleoli control the expression of developmental genes in this model are unknown. How specific lncRNAs selectively locate the corresponding interaction sites in the genome is not understood and the nature of lncRNA-chromatin interactions, as well as their possible functional roles, is not yet clear (Rinn and Chang, 2012). It is possible that rDNA-derived lncRNAs are involved in targeting and regulating a specific set of developmental genes.

Role of Nucleoli in Genome Stability, Aging, and Cancer

There is much evidence suggesting an important role for rDNA clusters in the regulation of cellular processes that are unrelated to ribosome biogenesis. For example, rDNA plays an important role in the DNA-damage response and in maintaining genome stability. The expression of rDNA is inhibited by DNA double-strand breaks (DSBs) induced by exogenous agents, e.g., ionizing radiation (Moore et al., 2011). DSBs occur under normal physiological conditions throughout the human genome, but the most fragile sites in the human genome coincide with actively transcribed rDNA genes, which possess hot spots of DSBs (Tchurikov et al., 2015). High transcriptional activity that leads to the formation of R-loops and to conflicts between transcription and replication within rDNA gene clusters are responsible for the DNA breakage of rDNA genes (Takeuchi et al., 2003; Lindström et al., 2018). There are nine hot spots of DSBs in the IGS of the rDNA repeats, denoted Pleiades (Tchurikov et al., 2016). The sites of these hot spots coincide with γ-H2AX marks, which suggests that the *in vivo* origin of DSBs is associated with transcription. However, Pleiades are only characteristic of active rDNA clusters that possess the UBF mark. It follows that a high level of DNA breakage inside the nucleoli should be accompanied by a high level of DNA repair (Korsholm et al., 2019).

The presence of hot spots of DSBs in rDNA explains the fact that there are 166 DNA-damage response (DDR) proteins found in the nucleolus (Hutten et al., 2011; Ogawa and Baserga, 2017). Among the proteins that are phosphorylated by kinases in response to DNA damage by ataxia-telangiectasia-mutated (ATM) and ataxia-telangiectasia and Rad3-related (ATR) kinases are 98 nucleolar proteins involved in ribosome biogenesis, ribosome function, and epigenetic regulation of rDNA genes. These facts led to the conclusion that the nucleolus is an

important hub of the DDR (Matsuoka et al., 2007; Larsen and Stucki, 2016). The data suggesting a general role of nucleoli in chromosomal DNA repair were confirmed by the finding that many DNA repair proteins can freely relocalize from nucleoli to the nucleoplasm and contribute to DNA repair at different chromosomal loci (Antoniali et al., 2014). Nucleolar proteins constantly move between the nucleolus and the nucleoplasm (Hernandez-Verdun, 2006; Sirri et al., 2008). This constant movement is associated with other novel nucleoli functions beyond the formation of ribosomes, including ribonucleoprotein biogenesis and the regulation of mitosis and the cell cycle, as well as the response to several types of stress (Boisvert et al., 2007; Boulon et al., 2010; Lindström and Latonen, 2013; Latonen, 2019). Thus, nucleoli are dynamic functional hubs that coordinate genome integrity, DNA repair mechanisms, stress response, and other cellular functions.

Upon proteotoxic insults, such as proteasome inhibition or heat shock treatment, nucleolar aggresomes are formed within the nucleolus in nucleolar cavities and intranucleolar bodies (Latonen, 2019). Similar structures are formed in certain neurodegenerative disorders in which proteins and RNA accumulate and aggregate. Interestingly, several non-coding RNAs that are transcribed from the IGS can recruit proteins to the aggresomes (Audas et al., 2012).

Nucleoli are shaped by the most conserved DNA sequences and thus, could potentially serve as markers of cellular longevity and aging mechanisms (Tiku et al., 2017; Wang and Lemos, 2019). The nucleolus is considered to be a convergent point of regulation of major longevity pathways, which strikingly reduce nucleolar size and diminish the expression of the nucleolar protein FIB-1, ribosomal RNA, and ribosomal proteins across species; furthermore, the development of small nucleoli correlates with longevity in higher organisms (Tiku et al., 2017). The underlying mechanisms of this correlation are unknown. However, nucleolar size positively correlates with rRNA synthesis and the TOR signaling pathway regulates nucleolar size (Tiku and Antebi, 2018). The reduced TOR signaling leads to diminished nucleolar size and function, as well as increased longevity, in different organisms. On the other hand, active TOR signaling promotes growth and proliferation and is often hyperactivated in tumors, leading to increased nucleolar size (Derenzini et al., 1998). Genome instability can accelerate cellular senescence, which restricts the lifespan of a cell, and the stability of rDNA affects the lifespan (Kobayashi, 2014). It has also been proposed that rDNA clusters play a key role in maintaining the stability of the whole genome and the control of the cellular lifespan.

It has been suggested that changes in cytosine-5 methylation within CpG dinucleotides sites across the genome can be used to predict human chronological age, as well as aspects of biological age (Horvath, 2013). However, the mechanisms linking DNA methylation changes with age are also unclear. The methylation status inside rDNA units may explain both observed correlations (longevity with small nucleoli size and aging with DNA methylation). Human rDNA genes possess a high density of CpGs and potentially could be regulated by DNA methylation mechanisms during aging. The putative association of rDNA methylation with age was tested during aging in humans, mice,

and dogs (Wang and Lemos, 2019). A significant age-associated hypermethylation of the rDNA relative to other regions of the genome was detected. However, the underlying mechanisms of this association are yet to be elucidated. Whatever the mechanism may be, it is not universal. In *Drosophila* and yeast, there is only low rDNA methylation and so other mechanisms of aging must exist beyond rDNA methylation.

rDNA genes are hot spots of DNA damage, and they often make intra- and inter-chromosomal contacts with different genomic regions that also possess hot spots of DSBs. These rDNA features lead to a high potential for translocations with different chromosomal regions, as well as with other rDNA clusters (Tchurikov et al., 2015). The latter could explain the origin of Robertsonian translocations that involve one or two rDNA-containing human acrocentric chromosomes (13, 14, 15, 21, and 22). The rDNA-mediated genome rearrangements could change the regulation of critical target genes and give rise to a cancer cell. As rDNA clusters consist of tandemly repeated genes, damage inside rDNA could be repaired by recombination with another rDNA copy and, as a result, the cluster could lose copies. It was observed that in 54% of solid tumors, there are rDNA cluster rearrangements before the start of clonal tumor expansion (Stults et al., 2009). The link between nucleoli and cancer was established more than 100 years ago by the observation of large and abnormal nucleoli in cancer cells (Pianese, 1896), which is thought to be due to hyper-activated transcription of rDNA (Hein et al., 2013). Cancer cells boost rDNA expression mainly via the genes involved in Pol-I-mediated transcription and through stimulation of their activity via different signaling pathways (for a detailed review see Gaviraghi et al., 2019). Although the link between nucleoli and cancer is well proved, the mechanisms of rDNA-mediated cancer genesis are not yet clear, possibly because, although we are aware that rDNA has many roles beyond ribosome biogenesis, the full list of cellular functions is unknown.

Nucleoli Shape Frequent Contacts With Genes Controlling Differentiation

Evidence of the role of nucleoli in differentiation, aging, and cancer raises questions on the nature of the underlying mechanisms by which rDNA clusters regulate different cellular processes. From a general point of view, there are two possible ways for such regulation to occur. One is the regulation by factors that act at a distance, e.g., non-coding RNAs or regulatory proteins that are dependent on the expression of rDNA genes. The second possible way is through direct contacts of rDNA clusters with particular sets of genes by the formation of a net of nucleoli-mediated 3D chromosomal structures. The first microscopic evidence in favor of the formation of reproducible contacts of nucleoli with specific bands in *Drosophila* polytene chromosomes was found many years ago (Ananiev et al., 1981). More examples from *Deptera* were described later (Zhimulev, 1998). Then, molecular indications for the interactions between nucleoli and different chromosomal regions were gained from experiments on the co-purification of large stretches of chromosomal DNA (up to 1 Mb) with nucleoli preparations (Németh et al., 2010; van Koningsbruggen et al., 2010). However,

this approach cannot accurately localize the contact sites of rDNA clusters within particular chromosomal regions or genes and so cannot reveal the regulatory targets. Therefore, high-resolution analyses are required because the regulatory influence of rDNA contacts could only spread to the nearest gene(s).
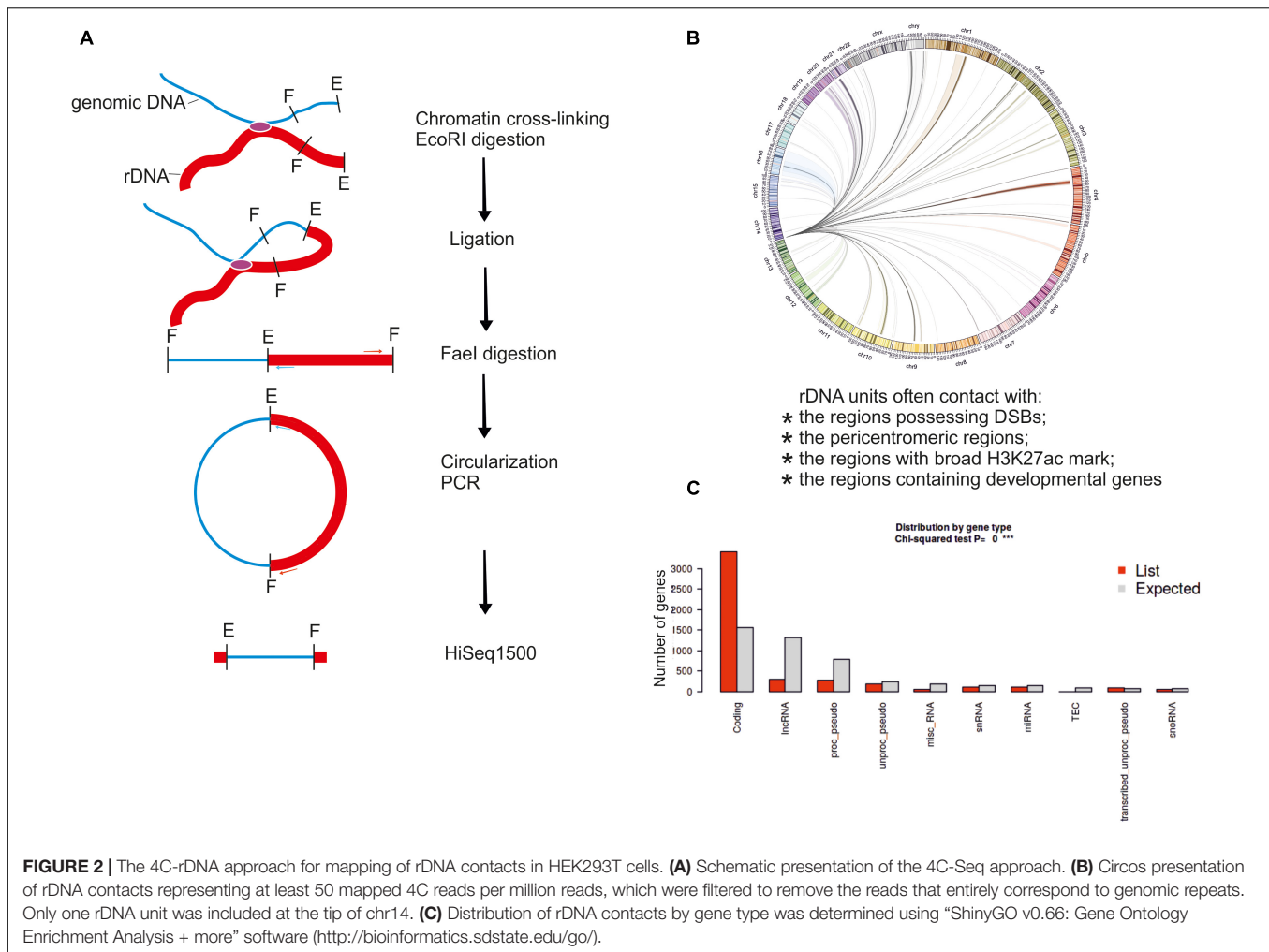
The high-resolution Hi-C and 4C approaches were used to more precisely localize the patterns of rDNA contacts in human cells (Yu and Lemos, 2018; Diesch et al., 2019). About 15 billion Hi-C reads from several experiments were used to map the rDNA-genome interactions with 1-Mb resolution. It was found that rDNA contacts are enriched in segments of closed, repressed, and late replicating chromatin, as well as CTCF binding sites (Yu and Lemos, 2018). Only a small portion of Hi-C reads represents the rDNA contacts. In contrast, the 4C-rDNA approach (**Figure 2A**) is more productive and allows amplification of only the DNA regions at the contact sites of rDNA. This approach was used to map rDNA contacts at better resolution (5 kb or less) using a MYC-driven lymphoma model or HEK293T cells (Diesch et al., 2019; Tchurikov et al., 2019).

The increased resolution confirmed the role of direct contacts of nucleoli with particular genes. First, the rDNA contacts are dynamic and their pattern changes during differentiation or in response to physiological stimuli.

Secondly, these changes in the contacts correlate with the changes in the expression of rDNA-contacting genes. Interestingly, in the lymphoma cells, gene expression changes at the rDNA-contacting loci include genes controlling B-cell differentiation, cell growth, and metabolism (Diesch et al., 2019). In HEK293T cells of neuronal origin, the nucleoli regulate the contacts with hundreds of genes controlling nervous system and neuron development (Tchurikov et al., 2019). In these cells, the contacts are detected in all chromosomes and often correspond to protein-coding genes (**Figures 2B,C**). In the MYC-driven lymphoma model, during the cellular transition from premalignancy to malignancy, there is a correlation between interactions of associated genes with the rDNA and transcriptional repression. These results suggest that the interactions with nucleoli contribute to Pol II gene regulation during the development of malignancy (Diesch et al., 2019).

In mice, the nucleolus may act as a hub for the location and regulation of repressive genomic domains, whereas nuclear speckles are hubs of the location and regulation of active genomic domains (Quinodoz et al., 2018; Kresoja-Rakic and Santoro, 2019). These findings were supported by the observation of repressive histone modifications at rDNA-containing sites. Nevertheless, the detailed analysis of profiles ± 1.5 kb around rDNA-contacting sites in HEK293T cells revealed both active and repressive states around the rDNA contacts (examples shown in **Figure 3**), while in *Drosophila,* the contacting sites are enriched with repressive chromatin marks (Tchurikov et al., 2019, 2020).

The depletion of Pol II and enrichment with repressive H3K9me3 marks and the binding sites of the transcriptional repressor ZNF274 suggest the presence of silent chromatin at rDNA-contacting sites (**Figure 3**). At the same time, rDNA also makes contact with active chromatin regions, where TAF15 and active chromatin H3K4me3 marks are present.

FIGURE 2 | The 4C-rDNA approach for mapping of rDNA contacts in HEK293T cells. **(A)** Schematic presentation of the 4C-Seq approach. **(B)** Circos presentation of rDNA contacts representing at least 50 mapped 4C reads per million reads, which were filtered to remove the reads that entirely correspond to genomic repeats. Only one rDNA unit was included at the tip of chr14. **(C)** Distribution of rDNA contacts by gene type was determined using "ShinyGO v0.66: Gene Ontology Enrichment Analysis + more" software (http://bioinformatics.sdstate.edu/go/).

Interestingly, small RNAs often occur at rDNA-contacting sites, which suggests that RNA-mediated mechanisms may be involved in the contacts.
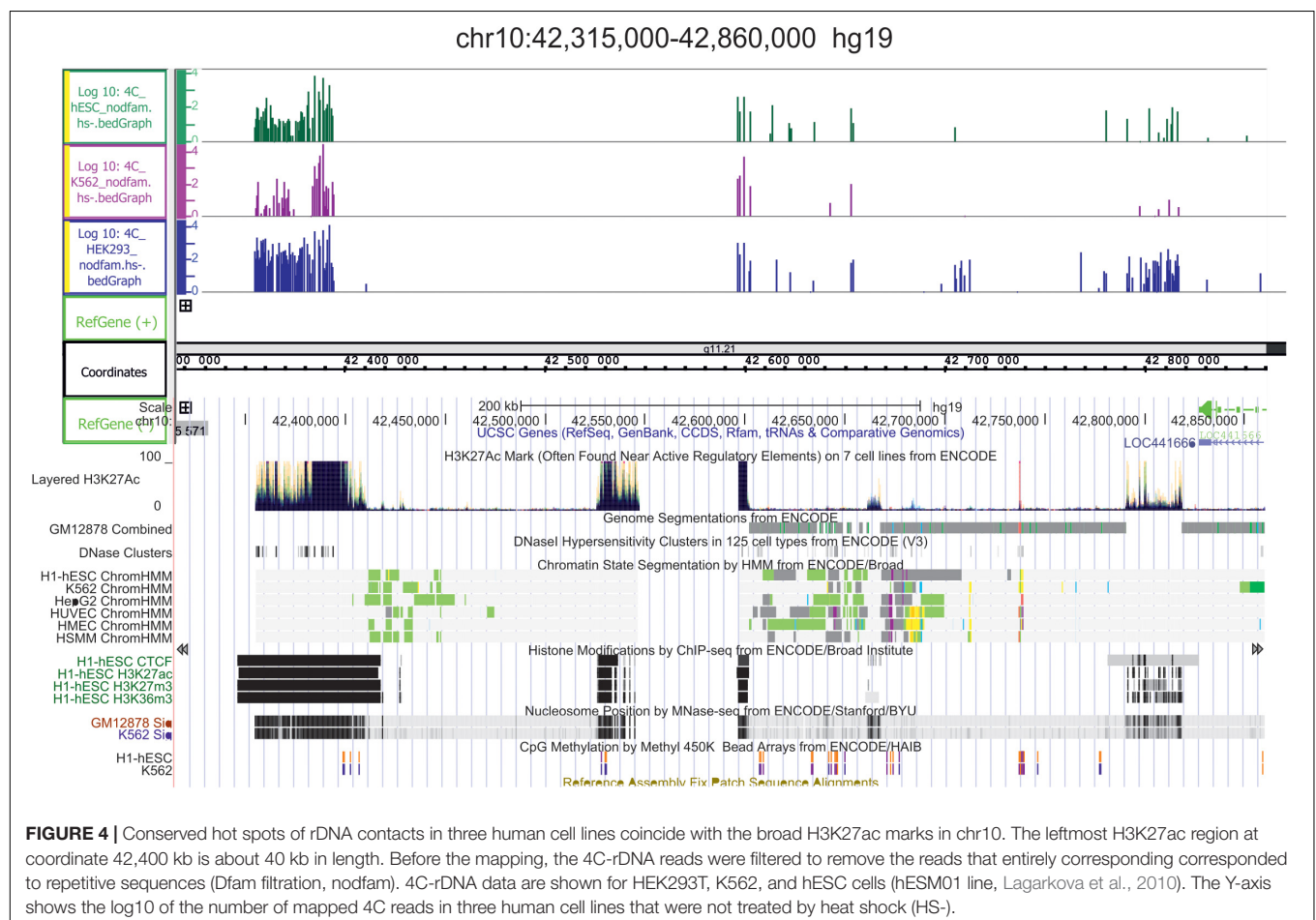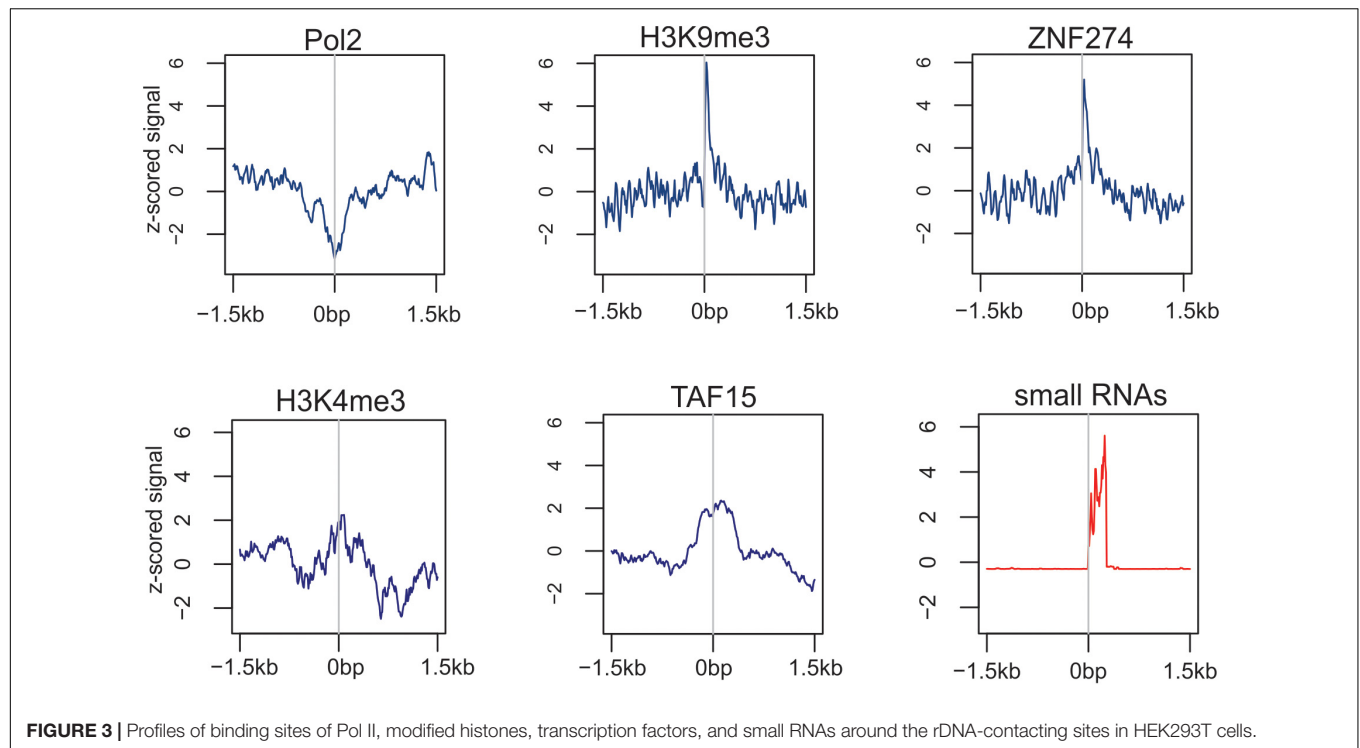
In light of the above, we conclude that both active and repressed chromosomal regions shape the contacts with nucleoli. We speculate that both active and silent rDNA clusters could spread the corresponding chromatin state to the chromosomal regions that make contact and, thus, by these mechanisms, nucleoli participate in the organization of both active and repressed hubs in nuclei during differentiation. This conclusion is supported by the detection of conspicuous rDNA contacts in different human cell lines in 5–50-kb regions marked with active H3K27ac marks that may correspond to super-enhancers (Hnisz et al., 2013; Tchurikov et al., 2013); **Figure 4** shows one example in chr10. The functional role of these regions is unknown.
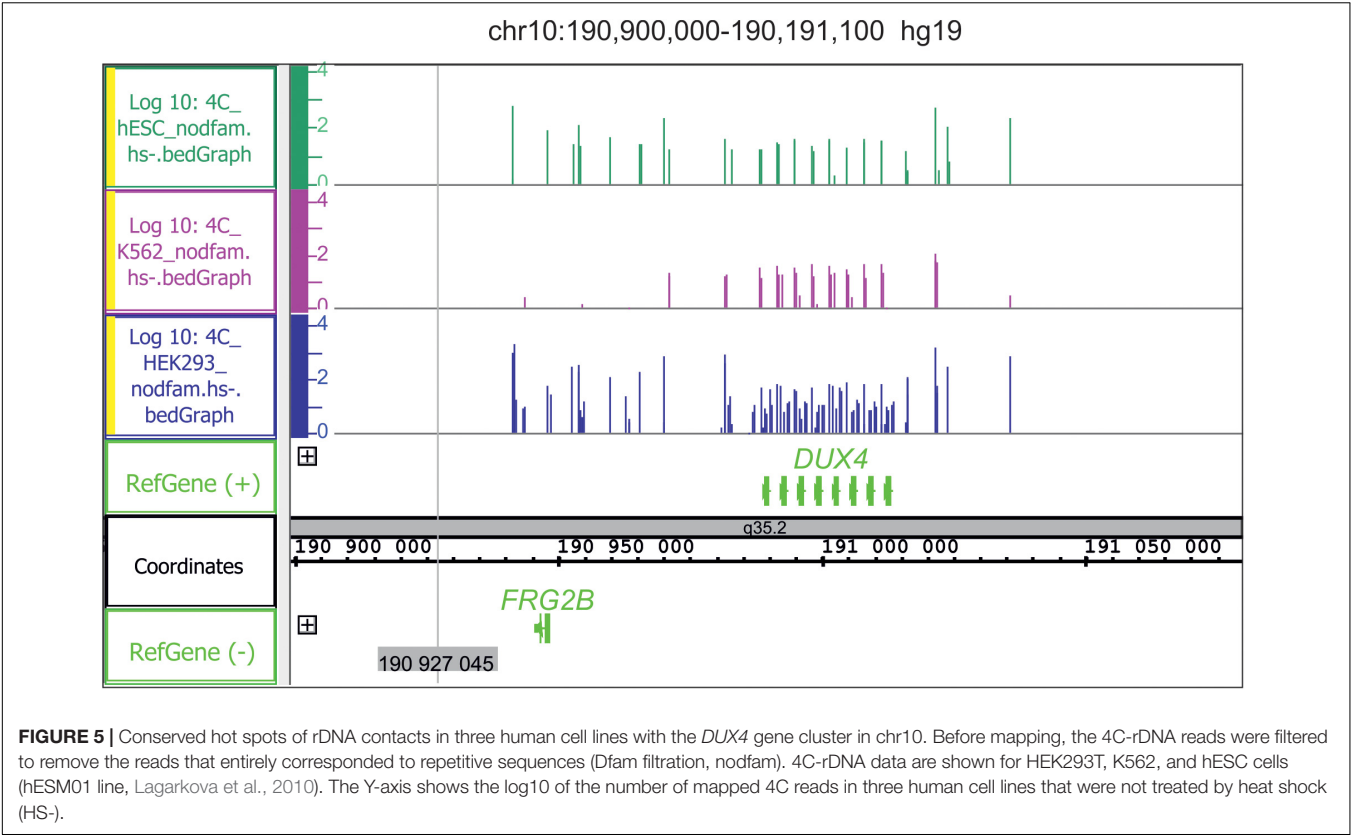
There are regions of frequent rDNA contacts that span about 100 kb in length and cover the silenced genes. One example is the *DUX4* gene cluster in the sub-telomeric region of chr4 (**Figure 5**). Heat shock treatment removes the rDNA contacts in this region (Tchurikov et al., 2020). *DUX4*, which is required at the two-cell embryo stage, specifies a transcription factor that activates hundreds of endogenous human genes (Hendrickson et al., 2017).

At later stages, the genes are repressed and their abnormal activation leads to facioscapulohumeral muscular dystrophy (De Iaco et al., 2017). The data strongly suggest both the association of rDNA contacts with the silencing of human developmental genes and the dynamic character of the contacts. The regions of rDNA contacts may correspond to the repressed chromatin. The large rDNA-contacting region precisely coincides with the repressed chromatin state in the *FANK1* gene (Kretova et al., 2020b).

Different human cell lines possess overlapping sets of rDNA-contacting genes that exhibit conserved rDNA contacts. For example, in HEK293T, K652, and hESM01 cells, the same set of about 500 genes frequently shape the contacts with rDNA (**Figure 6**).

Gene ontology searches suggest that the overlapping genes are involved in development and morphogenesis. About 100 of these genes (**Supplementary Table 2**) are highly associated with silencing by the H3K27me3 mark in several normal cell types, including bronchial epithelial cells, keratinocytes, myoblasts, monocytes, endothelial cells, and kidney epithelial cells (Tchurikov et al., 2021). Thus, a concerted silencing of a specific group of rDNA-contacting genes controlling development occurs during differentiation.

FIGURE 3 | Profiles of binding sites of Pol II, modified histones, transcription factors, and small RNAs around the rDNA-contacting sites in HEK293T cells.

FIGURE 4 | Conserved hot spots of rDNA contacts in three human cell lines coincide with the broad H3K27ac marks in chr10. The leftmost H3K27ac region at coordinate 42,400 kb is about 40 kb in length. Before the mapping, the 4C-rDNA reads were filtered to remove the reads that entirely corresponding corresponded to repetitive sequences (Dfam filtration, nodfam). 4C-rDNA data are shown for HEK293T, K562, and hESC cells (hESM01 line, Lagarkova et al., 2010). The Y-axis shows the log10 of the number of mapped 4C reads in three human cell lines that were not treated by heat shock (HS-).

FIGURE 5 | Conserved hot spots of rDNA contacts in three human cell lines with the *DUX4* gene cluster in chr10. Before mapping, the 4C-rDNA reads were filtered to remove the reads that entirely corresponded to repetitive sequences (Dfam filtration, nodfam). 4C-rDNA data are shown for HEK293T, K562, and hESC cells (hESM01 line, Lagarkova et al., 2010). The Y-axis shows the log10 of the number of mapped 4C reads in three human cell lines that were not treated by heat shock (HS-).

The association of nucleoli contacts with silenced or activated genes suggests the involvement of rDNA clusters in the global regulation of gene expression. Nevertheless, although the nucleoli

FIGURE 6 | There are the conserved rDNA contacts in human cell lines. Venn diagram showing the numbers of overlapping rDNA-contacting genes between HEK293T, K562, and hESC cells (hESM01 line). The complete list of corresponding genes is shown in **Supplementary Table 1**.

may play a major role in the regulation of gene expression regulation, they cannot work alone. Therefore, nucleoli contacts are necessary but not sufficient for such regulation. There are likely many other players in the global regulation of gene expression, e.g., for the silencing of *DUX4* genes, LINE1 transcripts are required (Percharde et al., 2018). It is conceivable that active or silent DNA units harbor hundreds of RNA and protein factors and their complexes. Furthermore, dynamic rDNA contacts may be shared with different genes and DNA regions, thus leading to an active or repressed state, or to treat the DNA breakage, and so on, delivering tools for multiple processes. The data on the presence of small RNAs at rDNA-contacting sites (**Figure 3**) confirm this supposition. However, the rDNA-mediated epigenetic players remain to be elucidated.

There are three major classes of rRNA genes in mammalian cells: silent, inactive, and active (Bersaglieri and Santoro, 2019). Silent rDNA units are characterized by DNA methylation in their promoter regions and by the presence of repressive histone marks, such as H3K9me2, H3K9me3, and deacetylated histones (Zhou et al., 2002). The active and inactive rDNA clusters do not possess DNA methylation in their promoter regions but may carry significant DNA methylation levels in, for example, the non-coding IGS (Moss et al., 2019). The active clusters are epigenetically marked by UBF and are nucleosome-free in the rDNA coding region, while inactive genes do not possess UBF marks and are packed with nucleosomes (Mars et al., 2018; Bersaglieri and Santoro, 2019). Therefore, the 4C experiments should reflect the inter-chromosomal interactions of all rDNA
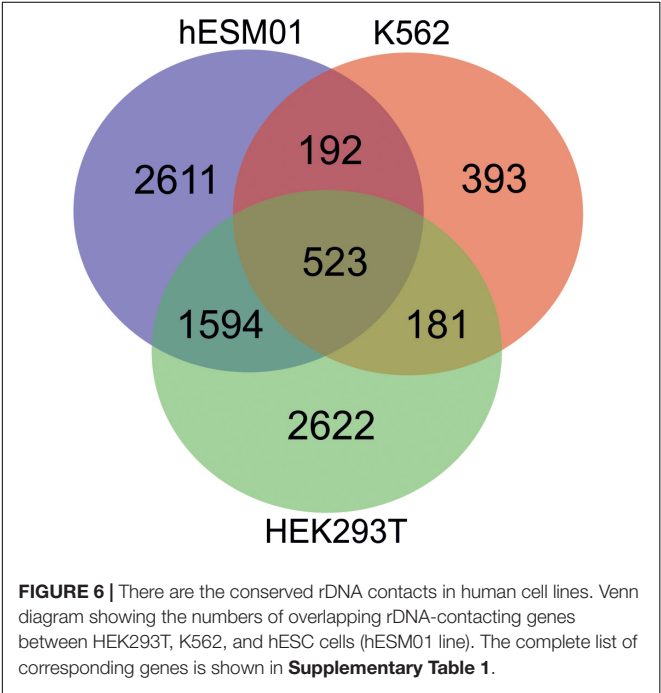
classes. We assume that ChIP-Seq experiments that aim to reveal the genome-wide localization of nucleolar proteins (UBF, fibrillarin, or nucleolin), which are present at active rDNA units, could help to discriminate the inter-chromosomal contacts of nucleoli from the contacts of the silent or inactive rDNA clusters.

Nucleoli and Phase-Separation Mechanisms

Active rDNA clusters organize nucleoli and rebuild them after each cell cycle (Hernandez-Verdun, 2011). UBF, which is required for the activation of rDNA units, is epigenetically inherited and marks the clusters that were active in the previous cell cycle and are destined to be active in the next cell cycle (O'Sullivan et al., 2002; McStay and Grummt, 2008; Dimitrova, 2011). The same is true for γ-H2AX that marks the active rDNA clusters (Tchurikov et al., 2016). Inter-chromosomal rDNA contacts are also re-established in every cell cycle. These data suggest that epigenetic mechanisms are involved in the lifecycle of nucleoli and their 3D network.

Nucleoli in their genomic contacts prefer some epigenetic marks, e.g., active H3K27ac marks (Tchurikov et al., 2015). At present, we do not know whether these marks appeared before or after the contacts were made with rDNA clusters. The H3K27ac mark is associated with super-enhancers and with phase-separation mechanisms (Hnisz et al., 2013; Sabari et al., 2018). The link between rDNA contacts, broad H3K27ac marks, and super-enhancers suggests the involvement of nucleoli in phase-separation mechanisms (Tchurikov et al., 2020). The H3K27ac mark is a characteristic of super-enhancers and was used to create a catalog of super-enhancers in different human cell and tissue samples (Hnisz et al., 2013; Tang et al., 2020). Super-enhancers have a high density of different transcription factors, which makes them a source of the nucleation event during phase separation (Mansour et al., 2014). MED1, a subunit of Mediator, and BRD4, a chromatin reader protein that recognizes and binds acetylated histones, occupy discrete nuclear bodies that occur at super-enhancers (Sabari et al., 2018). These bodies are disrupted by 1, 6-hexanediol, a drug that disrupts liquid-like condensates, possibly by disruption of hydrophobic interactions (Kroschwald et al., 2017). These results show that transcriptional coactivators form phase-separated condensates at super-enhancers. Together with the data that nucleoli frequently form contacts with the regions decorated with broad H3K27ac marks (Tchurikov et al., 2015), these data suggest that nucleoli are associated with phase-separated condensates. The idea is supported by data showing that transcriptional condensate formation contributes to long-range genomic interactions (Shrinivas et al., 2019).

In general, all genomic repeats could generate phase separations (Hall et al., 2019). The tripartite structure of the nucleolus itself, which separates the FC, DFC, and GC, also depends on the phase separation of different protein components. The large number of different factors controlling transcription and DNA repair and non-coding RNAs that accumulate at rDNA units could be the source of the nucleation event of phase separation and the formation of nucleoli in each cell cycle. Recently, novel chaperone-like properties of the

nucleolus as a phase-separated organelle associated with the refolding of misfolded proteins were described (Frottin et al., 2019). Metastable nuclear proteins that misfold after heat shock treatment could enter the nucleoli where they avoid irreversible aggregation and remain competent for HSP70-dependent refolding upon recovery from stress.

The cognate phase-separated structures on chromosomes could promote the interaction between condensates of the same nature, including nucleoli interactions. The regions of inter-chromosomal rDNA contacts may compete with the local intra-chromosomal contacts and displace them. In *Drosophila* genes, the multiple nucleoli contacts are located in the center of a bubble around which the main chromatin loops are formed (Kretova et al., 2020a; Tchurikov et al., 2020). The forces of phase-separation mechanisms of nucleoli are probably stronger than those between intra-chromosomal loops in this region.

Recently, it was demonstrated that fibrillarin, the dense fibrillar component constituent, and nucleophosmin, the scaffold protein of the granular component, are implicated in nucleation, including the tripartite organization of nucleoli (Yao et al., 2019; Lafontaine et al., 2021). In direct experiments, the 5′ end of nascent 47S pre-rRNA binds co-transcriptionally to the RNA-binding domain of fibrillarin, which diffuses to the DFC (Yao et al., 2019). In the DFC, the local self-association between glycine- and arginine-rich domains of fibrillarin shapes the phase-separated clusters that immobilize fibrillarin-interacting pre-rRNAs. In this way, the directional traffic of nascent pre-rRNAs occurs, thus facilitating pre-rRNA processing and DFC formation. *In vitro* droplet reconstitution with purified fibrillarin and nucleophosmin showed that the proteins readily form condensed liquid droplets that exhibit biophysical features similar to those of intact nucleoli (Lafontaine et al., 2021).

The nature of the nucleolus, which is made up of phase-separated compartments itself, suggests a potential role in the long-range dynamic interactions in the nucleus because liquid-liquid phase separation physically allows the rapid movement of components into and within the dense phase (Brangwynne et al., 2009). These interactions are dynamic and dependent on the differentiation state, phase of the cell cycle, and external physiological conditions. The dynamics of nucleoli correspond to the dynamic organization of chromosomes revealed by live-cell imaging data that suggest an organized motion of highly viscous droplet-like domains that can be likened to chromatin "breathing" (Latonen, 2019; Misteli, 2020; Shaban and Seeber, 2020; Feric and Misteli, 2021). The understanding of nucleoli structure and function has come a long way from the 1830s (Pederson, 2011). Novel approaches could elucidate the phase separation mechanisms underlying the structure and functions of nucleoli as the most remarkable component of nuclei.

CONCLUSION AND FUTURE PERSPECTIVES

Nucleoli play many important roles beyond the biogenesis of ribosomes, including shaping of the nuclear architecture

and regulation of DNA repair, differentiation, chaperone-like functions, RNP formation, diverse stress responses and others. The abnormal function of nucleoli leads to cancer genesis and diseases. Recently, it was demonstrated that inter-chromosomal contacts of nucleoli are involved in the regulation of global gene expression. The nature of these contacts and their role in development remains to be elucidated. It is not clear how rDNA inter-chromosomal contacts affect the local intra-chromosomal 3D domains. The contacts may be important for other functions of the nucleoli, including DNA repair and stress responses. The key areas for study in the future include determining the underlying molecular mechanisms of nucleoli function as a driver of nucleoli's role in cellular development and the response to environmental stimuli, the RNA-mediated mechanisms involved in recognizing target genes, and the phase-separation mechanisms in the formation of nucleoli and their dynamic 3D structures.

AUTHOR CONTRIBUTIONS

NT wrote the manuscript. YK developed software for image and statistical analysis. Both authors contributed to the article and approved the submitted version.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.730633/full#supplementary-material

REFERENCES

Akhmanova, A., Verkerk, T., Langeveld, A., Grosveld, F., and Galjart, N. (2000). Characterisation of transcriptionally active and inactive chromatin domains in neurons. *J. Cell Sci.* 113, 4463–4474. doi: 10.1242/jcs.113.24.4463

Ananiev, E. V., Barsky, V. E., Ilyin, Y. V., and Churikov, N. A. (1981). Localization of nucleoli in Drosophila melanogaster polytene chromosomes. *Chromosoma* 81, 619–628. doi: 10.1007/bf00285853

Antoniali, G., Lirussi, L., Poletto, M., and Tell, G. (2014). Emerging roles of the nucleolus in regulating the DNA damage response: the noncanonical DNA repair enzyme APE1/Ref-1 as a paradigmatical example. *Antioxid. Redox Signal.* 20, 621–639. doi: 10.1089/ars.2013.5491

Audas, T. E., Jacob, M. D., and Lee, S. (2012). Immobilization of proteins in the nucleolus by ribosomal intergenic spacer noncoding RNA. *Mol. Cell* 45, 147–157. doi: 10.1016/j.molcel.2011.12.012

Baßler, J., and Hurt, E. (2019). Eukaryotic Ribosome Assembly. *Annu. Rev. Biochem.* 88, 281–306. doi: 10.1146/annurev-biochem-013118-110817

Bersaglieri, C., and Santoro, R. (2019). Genome Organization in and around the Nucleolus. *Cells* 8:579. doi: 10.3390/cells8060579

Bhattacharya, D., Talwar, S., Mazumder, A., and Shivashankar, G. V. (2009). Spatio-temporal plasticity in chromatin organization in mouse cell differentiation and during Drosophila embryogenesis. *Biophys. J.* 96, 3832–3839. doi: 10.1016/j.bpj.2008.11.075

Boisvert, F. M., van Koningsbruggen, S., Navascués, J., and Lamond, A. I. (2007). The multifunctional nucleolus. *Nat. Rev. Mol. Cell Biol.* 8, 574–585. doi: 10.1038/nrm2184

Boulon, S., Westman, B. J., Hutten, S., Boisvert, F. M., and Lamond, A. I. (2010). The nucleolus under stress. *Mol. Cell* 40, 216–227. doi: 10.1016/j.molcel.2010.09.024

Brangwynne, C. P., Eckmann, C. R., Courson, D. S., Rybarska, A., Hoege, C., Gharakhani, J., et al. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* 324, 1729–1732. doi: 10.1126/science.1172046

De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., and Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* 49, 941–945. doi: 10.1038/ng.3858

Derenzini, M., Trerè, D., Pession, A., Montanaro, L., Sirri, V., and Ochs, R. L. (1998). Nucleolar function and size in cancer cells. *Am. J. Pathol.* 152, 1291–1297.

Diesch, J., Bywater, M. J., Sanij, E., Cameron, D. P., Schierding, W., Brajanovski, N., et al. (2019). Changes in long-range rDNA-genomic interactions associate with altered RNA polymerase II gene programs during malignant transformation. *Commun. Biol.* 2:39.

Dimitrova, D. C. (2011). DNA replication initiation patterns and spatial dynamics of the human ribosomal RNA gene loci. *J. Cell Sci.* 124, 2743–2752. doi: 10.1242/jcs.082230

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380. doi: 10.1038/nature11082

Feinberg, A. (2014). The Nucleolus Gets the Silent Treatment. *Stem Cell* 15, 675–676. doi: 10.1016/j.stem.2014.11.017

Feric, M., and Misteli, T. (2021). Phase Separation in Genome Organization across Evolution. *Trends Cell Biol.* 31, 671–685. doi: 10.1016/j.tcb.2021.03.001

Frottin, F., Schueder, F., Tiwary, S., Gupta, R., Körner, R., Schlichthaerle, T., et al. (2019). The nucleolus functions as a phase-separated protein quality control compartment. *Science* 365, 342–347. doi: 10.1126/science.aaw9157

Gaviraghi, M., Vivori, C., and Tonon, G. (2019). How Cancer Exploits Ribosomal RNA Biogenesis: a Journey beyond the Boundaries of rRNA Transcription. *Cells* 8:1098. doi: 10.3390/cells8091098

Granneman, S., and Baserga, S. J. (2005). Crosstalk in gene expression: coupling and co-regulation of rDNA transcription, pre-ribosome assembly and pre-rRNA processing. *Curr. Opin. Cell Biol.* 17, 281–286. doi: 10.1016/j.ceb.2005.04.001

Guetg, C., Scheifele, F., Rosenthal, F., Hottiger, M. O., and Santoro, R. (2012). Inheritance of silent rDNA chromatin is mediated by PARP1 via noncoding RNA. *Mol. Cell* 45, 790–800. doi: 10.1016/j.molcel.2012.01.024

Hall, A. C., Ostrowski, L. A., and Mekhail, K. (2019). Phase Separation as a Melting Pot for DNA Repeats. *Trends Genet.* 35, 589–600. doi: 10.1016/j.tig.2019.05.001

Hein, N., Hannan, K. M., George, A. J., Sanij, E., and Hannan, R. D. (2013). The nucleolus: an emerging target for cancer therapy. *Trends Mol. Med.* 19, 643–654. doi: 10.1016/j.molmed.2013.07.005

Hendrickson, P., Doráis, J., Grow, E., Whiddon, J. L., Lim, J.-W., Wike, C. L., et al. (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* 49, 925–934. doi: 10.1038/ng.3844

Hernandez-Verdun, D. (2006). Nucleolus: from structure to dynamics. *Histochem. Cell Biol.* 125, 127–137. doi: 10.1007/s00418-005-0046-4

Hernandez-Verdun, D. (2011). Assembly and disassembly of the nucleolus during the cell cycle. *Nucleus* 2, 189–194. doi: 10.4161/nucl.2.3.16246

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., et al. (2013). Super-enhancers in the control of cell identity and disease. *Cell* 155, 934–947. doi: 10.1016/j.cell.2013.09.053

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biol.* 14:R115.

Hutten, S., Prescott, A., James, J., Riesenberg, S., Boulon, S., Lam, Y. W., et al. (2011). An intranucleolar body associated with rDNA. *Chromosoma* 120, 481–499. doi: 10.1007/s00412-011-0327-8

Kobayashi, T. (2014). Ribosomal RNA gene repeats, their stability and cellular senescence. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* 90, 119–129. doi: 10.2183/pjab.90.119

Korsholm, L. M., Gál, Z., Lin, L., Quevedo, O., Ahmad, D. A., Dulina, E., et al. (2019). Double-strand breaks in ribosomal RNA genes activate a distinct signaling and chromatin response to facilitate nucleolar restructuring and repair. *Nucleic Acids Res.* 47, 8019–8035. doi: 10.1093/nar/gkz518

Kresoja-Rakic, J., and Santoro, R. (2019). Nucleolus and rRNA Gene Chromatin in Early Embryo Development. *Trends Genet.* 35, 868–879. doi: 10.1016/j.tig.2019.06.005

Kretova, O. V., Fedoseeva, D. M., Kravatsky, Y. V., Klushevskaya, E. S., Alembekov, I. R., Slovohotov, I. Y., et al. (2020a). Contact Sites of rDNA Clusters with *FANK1* Gene Correspond to Repressed Chromatin. *Mol. Biol.* 54, 262–266. doi: 10.31857/S002689842002007X

Kretova, O. V., Fedoseeva, D. M., Slovohotov, I. Y., Klushevskaya, E. S., Kravatsky, Y. V., and Tchurikov, N. A. (2020b). Drosophila rDNA Genes Shape the Stable Contacts with the Tlk Gene at the Expression Area of Small RNAs and Affect on Looped Domains inside the Gene. *Mol. Biol.* 54, 445–449. doi: 10.31857/S0026898420020081

Kroschwald, S., Maharana, S., and Simon, A. (2017). Hexanediol: a chemical probe to investigate the material properties of membrane-less compartments. *Matters* 3:e201702000010. doi: 10.19185/matters.201702000010

Lafontaine, D. L. J., Riback, J. A., Bascetin, R., and Brangwynne, C. P. (2021). The nucleolus as a multiphase liquid condensate. *Nat. Rev. Mol. Cell Biol.* 22, 165–182. doi: 10.1038/s41580-020-0272-6

Lagarkova, M. A., Shutova, M. V., Bogomazova, A. N., Vassina, E. M., Glazov, E. A., Zhang, P., et al. (2010). Induction of pluripotency in human endothelial cells resets epigenetic profile on genome scale. *Cell Cycle* 9, 937–946. doi: 10.4161/cc.9.5.10869

Larsen, D. H., and Stucki, M. (2016). Nucleolar responses to DNA double strand breaks. *Nucleic Acids Res.* 44, 538–544. doi: 10.1093/nar/gkv1312

Latonen, L. (2019). Phase-to-Phase With Nucleoli - Stress Responses, Protein Aggregation and Novel Roles of RNA. *Front. Cell Neurosci.* 13:151. doi: 10.3389/fncel.2019.00151

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293. doi: 10.1126/science.1181369

Lindström, M. S., Jurada, D., Bursac, S., Orsolic, I., Bartek, J., and Volarevic, S. (2018). Nucleolus as an emerging hub in maintenance of genome stability and cancer pathogenesis. *Oncogene* 37, 2351–2366. doi: 10.1038/s41388-017-0121-z

Lindström, M. S., and Latonen, L. (2013). "The nucleolus as a stress response organelle," in *Proteins of the Nucleolus. Regulation, Translocation and Biomedical Functions*, eds D. H. O'Day and A. Catalano (Berlin: Springer), 251–273.

Mansour, M. R., Abraham, B. J., Anders, L., Berezovskaya, A., Gutierrez, A., Durbin, A. D., et al. (2014). An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* 346, 1373–1377. doi: 10.1126/science.1259037

Mars, J. C., Sabourin-Felix, M., Tremblay, M. G., and Moss, T. (2018). A Deconvolution Protocol for ChIP-Seq Reveals Analogous Enhancer Structures on the Mouse and Human Ribosomal RNA Genes. *G3* 8, 303–314. doi: 10.1534/g3.117.300225

Matsuoka, S., Ballif, B. A., Smogorzewska, A., McDonald, E. R. III, Hurov, K. E., Luo, J., et al. (2007). ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* 316, 1160–1166.

McStay, B., and Grummt, I. (2008). The epigenetics of rRNA genes: from molecular to chromosome biology. *Annu. Rev. Cell Dev. Biol.* 24, 131–157.

Misteli, T. (2020). The Self-Organizing Genome: principles of Genome Architecture and Function. *Cell* 183, 28–45. doi: 10.1016/j.cell.2020.09.014

Moore, H. M., Bai, B., Boisvert, F. M., Latonen, L., Rantanen, V., Simpson, J. C., et al. (2011). Quantitative proteomics and dynamic imaging of the nucleolus reveal distinct responses to UV and ionizing radiation. *Mol. Cell. Proteomics* 10:M111.009241. doi: 10.1074/mcp.M111.009241

Moss, T., Mars, J. C., Tremblay, M. G., and Sabourin-Felix, M. (2019). The chromatin landscape of the ribosomal RNA genes in mouse and human. *Chromosome Res.* 27, 31–40. doi: 10.1007/s10577-018-09603-9

Németh, A., Conesa, A., Santoyo-Lopez, J., Medina, I., Montaner, D., Péterfia, B., et al. (2010). Initial Genomics of the Human Nucleolus. *PLoS Genet.* 6:e1000889. doi: 10.1371/journal.pgen.1000889

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385.

Ogawa, L. M., and Baserga, S. J. (2017). Crosstalk between the nucleolus and the DNA damage response. *Mol. Biosyst.* 13, 443–455.

O'Sullivan, A. C., Sullivan, G. J., and McStay, B. (2002). UBF binding *in vivo* is not restricted to regulatory sequences within the vertebrate ribosomal DNA repeat. *Mol. Cell Biol.* 22, 657–668.

Pederson, T. (2011). The nucleolus. *Cold Spring Harb. Perspect. Biol.* 3:a000638. doi: 10.1101/cshperspect.a000638

Percharde, M., Lin, C. J., Yin, Y., Guan, J., Peixoto, G. A., Bulut-Karslioglu, A., et al. (2018). A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell* 174, 391–405.e19. doi: 10.1016/j.cell.2018.05.043

Pianese, G. (1896). *Beitrag zur Histologie und Aetiologie des Carcinoms.* Jena: G. Fischer.

Quinodoz, S. A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J. M., Detmar, E., et al. (2018). Higher-order interchromosomal hubs shape 3D genome organization in the nucleus. *Cell* 174, 744–757.

Rinn, J. L., and Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* 81, 145–166.

Sabari, B. R., Dall'Agnese, A., Boija, A., Klein, I. A., Coffey, E. L., Shrinivas, K., et al. (2018). Coactivator Condensation at Super-Enhancers Links Phase Separation and Gene Control. *Science* 361:eaar3958. doi: 10.1126/science.aar3958

Santoro, R., Schmitz, K. M., Sandoval, J., and Grummt, I. (2010). Intergenic transcripts originating from a subclass of ribosomal DNA repeats silence ribosomal RNA genes in trans. *EMBO Rep.* 11, 52–58.

Savić, N., Bär, D., Leone, S., Frommel, S. C., Weber, F. A., Vollenweider, E., et al. (2014). lncRNA maturation to initiate heterochromatin formation in the nucleolus is required for exit from pluripotency in ESCs. *Cell Stem Cell.* 15, 720–734. doi: 10.1016/j.stem.2014.10.005

Shaban, H. A., and Seeber, A. (2020). Monitoring the spatio-temporal organization and dynamics of the genome. *Nucleic Acids Res.* 48, 3423–3434.

Shrinivas, K., Sabari, B. R., Coffey, E. L., Klein, I. A., Boija, A., Zamudio, A. V., et al. (2019). Enhancer Features that Drive Formation of Transcriptional Condensates. *Mol. Cell* 75, 549–561.e7. doi: 10.1016/j.molcel.2019.07.009

Sirri, V., Urcuqui-Inchima, S., Roussel, P., and Hernandez-Verdun, D. (2008). Nucleolus: the fascinating nuclear body. *Histochem. Cell Biol.* 129, 13–31. doi: 10.1007/s00418-007-0359-6

Stults, D. M., Killen, M. W., Williamson, E. P., Hourigan, J. S., Vargas, H. D., Arnold, S. M., et al. (2009). Human rRNA gene clusters are recombinational hotspots in cancer. *Cancer Res.* 69, 9096–9104.

Takeuchi, Y., Horiuchi, T., and Kobayashi, T. (2003). Transcription-dependent recombination and the role of fork collision in yeast rDNA. *Genes Dev.* 17, 1497–1506.

Tang, F., Yang, Z., Tan, Y., and Li, Y. (2020). Super-enhancer function and its application in cancer targeted therapy. *npj Precis. Onc.* 4:2. doi: 10.1038/s41698-020-0108-z

Tchurikov, N. A., Fedoseeva, D. M., Klushevskaya, E. S., Slovohotov, I. Y., Chechetkin, V. R., Kravatsky, Y. V., et al. (2019). rDNA clusters make contact with genes that are involved in differentiation and cancer and change contacts after heat shock treatment. 8:1393. doi: 10.3390/cells8111393

Tchurikov, N. A., Fedoseeva, D. M., Sosin, D. V., Snezhkina, A. V., Melnikova, N. V., Kudryavtseva, A. V., et al. (2015). Hot spots of DNA double-strand breaks and genomic contacts of human rDNA units are involved in epigenetic regulation. *J. Mol. Cell Biol.* 7, 366–382. doi: 10.1093/jmcb/mju038

Tchurikov, N. A., Klushevskaya, E. S., Kravatsky, Y. V., Kravatskaya, G. I., and Fedoseeva, D. M. (2021). Interchromosomal Contacts of rDNA Clusters in Three Human Cell Lines Are Associated with Silencing of Genes Controlling Morphogenesis. *Dokl. Biochem. Biophys.* 496, 22–26. doi: 10.1134/S1607672921010038

Tchurikov, N. A., Klushevskaya, E. S., Kravatsky, Y. V., Kravatskaya, G. I., Fedoseeva, D. M., and Kretova, O. V. (2020). Interchromosomal Contacts of rDNA Clusters with DUX Genes in Human Chromosome 4 Are Very Sensitive

to Heat Shock Treatment. *Dokl. Biochem. Biophys.* 490, 50–53. doi: 10.1134/S1607672920010032

Tchurikov, N. A., Kretova, O. V., Fedoseeva, D. M., Sosin, D. V., Grachev, S. A., Serebraykova, M. V., et al. (2013). DNA double-strand breaks coupled with PARP1 and HNRNPA2B1 binding sites flank coordinately expressed domains in human chromosomes. *PLoS Genet.* 9:e1003429. doi: 10.1371/journal.pgen.1003429

Tchurikov, N. A., Yudkin, D. V., Gorbacheva, M. A., Kulemzina, A. I., Grischenko, I. V., Fedoseeva, D. M., et al. (2016). Hot spots of DNA double-strand breaks in human rDNA units are produced in vivo. *Sci. Rep.* 10, 25866. doi: 10.1038/srep25866

Tiku, V., and Antebi, A. (2018). Nucleolar Function in Lifespan Regulation. *Trends Cell Biol.* 28, 662–672.

Tiku, V., Jain, C., Raz, Y., Nakamura, S., Heestand, B., Liu, W., et al. (2017). Small nucleoli are a cellular hallmark of longevity. *Nat. Commun.* 8:16083. doi: 10.1038/ncomms16083

van Koningsbruggen, S., Gierlinski, M., Schofield, P., Martin, D., Barton, G. J., Ariyurek, Y., et al. (2010). High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. *Mol. Biol. Cell* 21, 3735–3748.

Wang, M., and Lemos, B. (2019). Ribosomal DNA harbors an evolutionarily conserved clock of biological aging. *Genome Res.* 29, 325–333. doi: 10.1101/gr.241745.118

Yao, R. W., Xu, G., Wang, Y., Shan, L., Luan, P. F., Wang, Y., et al. (2019). Nascent Pre-rRNA Sorting via Phase Separation Drives the Assembly of Dense Fibrillar Components in the Human Nucleolus. *Mol. Cell* 76, 767–783.e11. doi: 10.1016/j.molcel.2019.08.014

Yu, S., and Lemos, B. (2018). The long-range interaction map of ribosomal DNA arrays. *PLoS Genet.* 14:e1007258. doi: 10.1371/journal.pgen.1007258

Zhang, Q., Shalaby, N. A., and Buszczak, M. (2014). Changes in rRNA transcription influence proliferation and cell fate within a stem cell lineage. *Science* 343, 298–301. doi: 10.1126/science.1246384

Zhimulev, I. F. (1998). Polytene chromosomes, heterochromatin, and position variegation. *Adv. Genet.* 37, 178–180.

Zhou, Y., Santoro, R., and Grummt, I. (2002). The chromatin remodeling complex NoRC targets HDAC1 to the ribosomal gene promoter and represses RNA polymerase I transcription. *EMBO J.* 21, 4632–4640.

Profiling of 3D Genome Organization in Nasopharyngeal Cancer Needle Biopsy Patient Samples by a Modified Hi-C Approach

Sambhavi Animesh[1†], Ruchi Choudhary[1,2†], Bertrand Jern Han Wong[2], Charlotte Tze Jia Koh[2], Xin Yi Ng[3], Joshua Kai Xun Tay[4], Wan-Qin Chong[3], Han Jian[1], Leilei Chen[1,5], Boon Cher Goh[1,3,6] and Melissa Jane Fullwood[1,2]*

[1] Cancer Science Institute of Singapore, Centre for Translational Medicine, National University of Singapore, Singapore, Singapore, [2] School of Biological Sciences, Nanyang Technological University, Singapore, Singapore, [3] Department of Haematology-Oncology, National University Cancer Institute, National University Health System, Singapore, Singapore, [4] Department of Otolaryngology - Head and Neck Surgery, National University of Singapore, Singapore, Singapore, [5] Department of Anatomy, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore, [6] Department of Pharmacology, Yong Loo Lin School of Medicine, National University Health System, Singapore, Singapore

Nasopharyngeal cancer (NPC), a cancer derived from epithelial cells in the nasopharynx, is a cancer common in China, Southeast Asia, and Africa. The three-dimensional (3D) genome organization of nasopharyngeal cancer is poorly understood. A major challenge in understanding the 3D genome organization of cancer samples is the lack of a method for the characterization of chromatin interactions in solid cancer needle biopsy samples. Here, we developed Biop-C, a modified *in situ* Hi-C method using solid cancer needle biopsy samples. We applied Biop-C to characterize three nasopharyngeal cancer solid cancer needle biopsy patient samples. We identified topologically associated domains (TADs), chromatin interaction loops, and frequently interacting regions (FIREs) at key oncogenes in nasopharyngeal cancer from the Biop-C heatmaps. We observed that the genomic features are shared at some important oncogenes, but the patients also display extensive heterogeneity at certain genomic loci. On analyzing the super enhancer landscape in nasopharyngeal cancer cell lines, we found that the super enhancers are associated with FIREs and can be linked to distal genes *via* chromatin loops in NPC. Taken together, our results demonstrate the utility of our Biop-C method in investigating 3D genome organization in solid cancers.

Keywords: nasopharyngeal cancer, chromatin organization, Hi-C, Biop-C, 3D genome organization

INTRODUCTION

The three-dimensional (3D) genome organization of the nucleus plays a vital role in the regulation of transcription (Hnisz et al., 2016, 2018). Alterations in 3D genome organization structures including topologically associated domain boundaries and chromatin loops have been shown to lead to oncogene expression and cancer progression (Fudenberg et al., 2011; Flavahan et al., 2016; Hnisz et al., 2016; Valton and Dekker, 2016; Li et al., 2019).

High-throughput chromosome conformation capture technologies such as Hi-C have been used to investigate 3D chromatin conformation (Lieberman-Aiden et al., 2009; van Berkum et al., 2010). The standard Hi-C approach generally requires approximately one million cells. Consequently, most previous analyses of cancer samples have been restricted to human cancer cell lines (Rao et al., 2014; Darrow et al., 2016; Haarhuis et al., 2017), but recently, Hi-C has been conducted on clinical samples from liquid cancers such as T-cell acute lymphoblastic leukemia (T-ALL) (Kloetgen et al., 2020) and diffuse large B-cell lymphoma (DLBCL) (Díaz et al., 2018) and one solid cancer—gastric cancer (Ooi et al., 2020). For these cancers, it is possible to obtain one million cells—for example, gastric cancers can grow to a large size. However, there are many cancers for which only needle biopsies are available (Lin et al., 2017).

To allow interrogation of samples with more limited quantities of starting materials, there have been several efforts to reduce the number of cells required to just 1K or 500 cells using modified protocols such as small-scale *in situ* Hi-C (sisHi-C) (Du et al., 2017), easy Hi-C (Lu et al., 2020), and Low-C (Díaz et al., 2018) (**Supplementary Table 1**). However, when dealing with solid cancers, a second challenge is that the tissue requires special preparation in order to dissociate the tissue into single cells for Hi-C analysis. The core needle biopsies pose the challenge of both limited cell numbers as well as the requirement for tissue dissociation, which might lead to further loss or degradation of chromatin for analysis. Solid cancers represent approximately 90% of adult human cancers (American Cancer Society, 2020); therefore, an easy-to-use method for preparing Hi-C libraries from needle biopsy cancer samples would advance our understanding of how chromatin organization contributes to cancer pathogenesis in solid cancers.

Here, we present Biop-C, a modified *in situ* Hi-C method for the chromatin analysis in solid cancer tissues from needle biopsy samples. The Biop-C method has been designed to be used on small tissue samples obtained from needle biopsies. To demonstrate the utility of this method, we analyzed three nasopharyngeal cancer (NPC) patient samples. NPC is an epithelial malignancy of the nasopharyngeal mucosa and is an aggressive subtype of head and neck cancers. NPC is highly prevalent in Southeast China, Southeast Asia, North Africa, Middle East, and the Arctic regions, but rare in most other parts of the world (Yu and Yuan, 2002). Multiple factors, including predisposing genetic factors, environmental carcinogens, and Epstein–Barr virus (EBV) infection, contribute to the etiology of NPC (Li et al., 2014; Dai et al., 2016). NPCs are further subdivided into three subtypes, viz. non-keratinizing undifferentiated carcinoma, non-keratinizing differentiated carcinoma, and keratinizing squamous cell carcinoma. Depending on the treatment given, the stage of the cancer, and the site where the cancer presents at, NPCs can be small and analyzed using fine needle biopsies.

It has been established that NPC has a comparatively low mutational burden, and oncogenicity is driven by epigenetic regulation. Typically, NPCs associated with EBV are characterized as having comparatively low DNA mutation rates but widespread DNA hypermethylation and overexpression or mutation of DNA methylation enzymes, histone modification enzymes, and chromatin remodeling enzymes (Dai et al., 2015, 2016; Tsang et al., 2020). Hence, unraveling the 3D conformational structure will provide further insight into the epigenetic regulatory mechanisms that promulgate the NPC phenotype. Here, we obtained a comprehensive understanding of the 3D genome organization of nasopharyngeal cancer through Biop-C analysis, which revealed complex 3D genome organization patterns at oncogenes important in NPC.

Moreover, as cancers are known to be heterogeneous but chromatin interaction heterogeneity in patient samples is poorly understood, we investigated chromatin interactions in three different nasopharyngeal cancer samples. We found that while there were similar chromatin interactions in all three samples, there were also chromatin interactions that were heterogeneous. We also prepared Hi-C libraries from an NPC cell line, HK1, and found differences between chromatin interactions in the patient samples as compared with the cell line, indicating the necessity of interrogating chromatin interactions in actual patient cancers, which Biop-C enables. Taken together, our results demonstrate the utility and importance of Biop-C as a method for understanding cancer 3D genome organization.

Additionally, we analyzed how NPC chromatin interactions differ from that of a near-normal nasopharyngeal cell line, NP-69, and we analyzed how THZ1, a drug that inhibits super enhancers, leads to changes in chromatin interactions in the NPC cell line, HK1. These functional data suggest that there are differences in the chromatin interaction landscape between nasopharyngeal cancer and normal nasopharyngeal cell line, and targeting super enhancers by THZ1 can modulate the chromatin interactome and lead to losses in chromatin interactions, suggesting that epigenetic drugs may be able to affect chromatin interactions that are altered in nasopharyngeal cancer.

MATERIALS AND METHODS

Cell Culture
Nasopharyngeal cancer HK1 cells were cultured in Roswell Park Memorial Institute (RPMI) 1640 media (Hyclone) supplemented with 10% heat-inactivated fetal bovine serum (FBS; Hyclone) and 1% penicillin/streptomycin (Hyclone). NP-69 cells were cultured in keratinocyte SFM media. All cell lines were grown in an incubator at 5% CO_2 and 37°C. The HK1 and NP-69 cell lines were a kind gift from Prof. Goh Boon Cher, Cancer Science Institute, National University of Singapore.

Biopsy Samples
Nasopharyngeal cancer patient samples were obtained from the National University Health System (NUHS) with patient consent, under Institute Review Board number 2018/00947-SRF0002. The clinical samples were collected by needle biopsy in a 1.5-μl microtube by trained clinicians. The needle type varies depending on the size and the location of the tumor. All clinical samples were obtained from the National University Hospital Singapore and collected according to the Human Biomedical Research Act

requirements. Informed consent was obtained for all clinical samples used in the study. The clinical details of the patients are listed in **Supplementary Table 2**.

The needle biopsy samples were flash frozen in liquid nitrogen immediately and stored at −80°C until further use.

Pulverization of the Needle Biopsy Sample

The liquid nitrogen-cooled mini mortar and pestle (SP Scienceware, United States) were used for the pulverization of the fine needle biopsy sample of nasopharyngeal patients. The stainless-steel mortar and pestle were cooled on dry ice before use. The liquid nitrogen was poured into the steel cavity up to the mark indicated.

The samples were taken out of the freezer and immediately placed in the fixture of the mortar. The sample was kept frozen throughout the pulverization. A small amount of liquid nitrogen (up to half of the microtube) was carefully poured into the microtube. Then, the liquid nitrogen was allowed to evaporate just enough for the small biopsy sample to stay submerged. The sample was pulverized in the mortar using the precooled pestle. The above steps were repeated until the sample resembled a fine powder without visible chunks. Generally, it took three reiterations to pulverize a needle biopsy sample to a fine powder.

Finally, a cooled spatula was used to transfer any remaining pulverized tissue from the pestle into the 1.5-µl microtube. The microtube was submerged in dry ice to keep all the pulverized tissue frozen.

Biop-C Library Preparation and Sequencing

Next, the Biop-C library was generated using the Arima Hi-C kit, according to the protocols of the manufacturer with slight modifications.

Specifically, the microtube was then removed from the dry ice, and the powder was mixed with 500 µl of 1× PBS. The chromatin was cross-linked by adding 50 µl of freshly prepared TC buffer (Arima kit; sodium chloride: 100 mM, EDTA: 1 mM, EGTA: 0.5 mM, HEPES pH 8.0: 50 mM, formaldehyde: 2%, water).

The samples were then mixed well by inverting and incubated at room temperature for 20 min. Finally, the reaction was stopped by adding stop solution (Arima kit). The pellet obtained after centrifugation was stored at −80°C until further use.

The fixed cells were permeabilized using a lysis buffer supplied in Arima Hi-C kit and then digested with a restriction enzyme cocktail. The resulting overhangs were filled in with biotinylated nucleotides followed by ligation. After ligation, crosslinks were reversed, and the DNA was purified from protein. Purified DNA was treated to remove biotin that was not internal to ligated fragments. The purified proximity-ligated DNA was sheared using a Covaris ME220 ultrasonicator.

The DNA fragments were size selected from ∼200 to 600 bp using DNA purification beads (AMPure XP beads). The size-selected fragments were then enriched for biotin and converted into Illumina-compatible sequencing libraries using low input Swift Biosciences Accel-NGS 2S Plus DNA Library Kit (Cat #

21024) and Swift Biosciences Indexing Kit (Cat # 26148). After adapter ligation, DNA was PCR amplified and purified using AMPure XP beads. The purified DNA underwent standard QC according to Arima Hi-C kit instructions such as qPCR and Agilent Bioanalyzer. Finally, the libraries were sequenced 150 bases paired-end on the Illumina HiSeq 4000 following the protocols of the manufacturer.

Analysis of HiC Sequencing Data Using Juicer and FitHiC2 Pipelines

Hi-C data were processed using Juicer (version 1.5.7) (Rao et al., 2014) with default parameters. The reference genome used for mapping the reads was hg19. Reads with mapping quality under 30 were discarded. The quality data and statistics for Hi-C analysis can be found in **Supplementary Table 3**.

TADs were called using the Juicer tool called Arrowhead with 10 kb resolution. The normalization used for TAD calling is Knight-Ruiz (KR). The list of TADs called for each patient sample and HK1 cell line can be found in **Supplementary Table 4**.

Chromatin loops were identified using the Juicer tool called HiCCUPS. Loops were called for three resolutions: 5, 10, and 25 kb. The normalization used is KR. The list of loops called for each sample and HK1 can be found in **Supplementary Table 5**.

FitHiC2 was additionally used for the estimation of contact significance (Kaul et al., 2020). Resultant hic files produced by Juicer were converted to FitHiC2-compatible sparse matrices. Biases were estimated from the matrix, excluding 20% of the lowest contact frequency bins. Sparse matrices and bias files were subsequently passed to FitHiC2.

THZ1 Treatment

The CDK7 inhibitor THZ1 (A8882) was purchased from ApexBio. The HK1 cells were grown overnight and then treated with THZ1 at 200, 500, and 1,000 nM concentrations for 24 h. The RNA was isolated from all three time points for real-time quantitative PCR (RT-qPCR). The Hi-C was performed from the cells treated at 500 nM for 24 h.

RNA Extraction, cDNA Synthesis, and Real-Time Quantitative PCR

Total RNA was extracted from the THZ1-treated and DMSO-treated HK1 cells using RNeasy Mini Kit (Qiagen) with on-column DNase digestion (Qiagen) according to the instructions of the manufacturer. The cDNA was synthesized using the qScript™ cDNA Synthesis Kit (Quanta Biosciences, MA, United States). RT-qPCR was conducted with GoTaq DNA Polymerase Master Mix (Promega, United States).

FIRE Calling

FIRE calling was carried out using the FIREcaller software (Crowley et al., 2021). The resolution for FIRE calling was 10 kb. Dense Hic matrix for each chromosome was created using Juicer dump with KR normalization and was then converted to mcool format using script from FIREcaller. Frequently interacting regions (FIREs) which were either overlapping or book-ended were merged together using bedtools merge.

Structural Variants Calling

Structural variants were called using the HiNT software (Wang et al., 2020). HiNT pre was used for preprocessing of Hi-C data and normalization followed by HiNT TL which is used for calling translocations and breakpoint detection. HiNT TL outputs candidate translocated chromosomal pairs and the exact location of the breakpoint (**Supplementary Table 8**). In case of an unmappable region, HiNT TL provides a 100-kb interval for the breakpoint.

AB Compartment Analysis

The AB compartments were identified using FAN-C 0.9.1 (Kruse et al., 2020) at 500 kb resolution. Fanc compartments function was used for calling the AB domains and eigenvector values for each domain. Genome file was provided using -g command for assignment of domains. The eigenvector is oriented in such a way that negative entries correspond to "B" (low GC content, inactive compartment) and positive entries to "A" (high GC content, active compartment).

For analyzing the changes between A/B compartments of two samples, we calculated Pearson correlation coefficient.

Super Enhancer and Enhancer Calling

ChIP-Seq data were aligned to reference genome hg19 using bwa mem (Li, 2013) with default parameters. PCR duplicated and blacklisted regions that fall under ENCODE consensus were removed using samtools markdup and bedtools intersect. Narrow peaks were then called using MACS2 (version 2.1.2) (Zhang et al., 2008). The modified version of ROSE package (Lunardon et al., 2014; Cao et al., 2017) was used for calling super enhancer stitched at 12 kb stitching distance. For calling individual enhancers, we overlapped stitched peaks with H3k27ac narrow peaks and the intersection was called as an enhancer. Super enhancers (SE) and enhancers from all the three cell lines were merged together using bedtools merge with a cutoff of 1 bp overlap. SE and enhancers present in all the three cell lines were identified as "common."

Comparison of TADs and Loops Between the Patients

Jaccard index was calculated using bedtools Jaccard for all the three patient samples as well as HK1 samples followed by comparison using bedtools overlap to identify the exact number of TADs which are common in two samples. TAD overlap of 80% was used as cutoff for characterizing two TADs as the same. The compare lists analysis of Juicer pipeline was used to compare the chromatin loops between the samples.

Associating SE and Enhancers With Chromatin Loops and FIREs

Loop anchors within 15 kb distance from a super enhancer or an enhancer were recognized as an associated chromatin loop and the gene within 15 kb distance from the other anchor of the same loop was called as an associated gene. Similarly, SE and enhancers were also associated with FIREs *via* chromatin

loops. Bedtools closest was used to identify features within 15 kb distance from SE.

Comparison of Replicates and Other Hi-C Data

Hi-C data from sample replicates as well as different samples were compared using hicrep (Yang et al., 2017) with a resolution of 100 kb. For the analysis, cool files were generated using hic2cool[1] and was run with default settings. Hicrep outputs stratum-adjusted correlation coefficient (SCC) as a similarity measure of Hi-C interaction matrices. SCC shares a similar range and interpretation as the standard correlation coefficients and the value lies between 1 and (−1).

Enhancer–Promoter Interaction Analysis

For a given set of loops, we identified paired features (enhancers on one anchor and promoters on the other), by searching for features within 15 kb of the loop anchor centroids. The transcription start site was used to approximate the promoter position.

Gene Set Overrepresentation Analysis

For the analysis of the NPC samples, we considered the intersection of genes looping to super enhancers, with promoters within 15 kb of loop anchor centroids, that were identified in all three NPC samples (S009, S010, and S012). For the analysis of HK1 following the THZ1 treatment paradigm, we considered the set of genes with promoters within 15 kb of loop anchor centroids, that showed a net loss in the number of associated loops following THZ1 treatment. For a given set of genes, significantly overrepresented Biological Process Gene Ontology (GO) terms were identified using the PANTHER 16.0 GSOA webtool (Mi et al., 2021). Statistical significance was estimated using a binomial test with false discovery rate (FDR) correction. Resultant enriched GO terms were used for subsequent GO set and network analysis using the NaviGo webtool (Wei et al., 2017).

RESULTS AND DISCUSSION

A Genome-Wide Map of 3D Genome Organization in Nasopharyngeal Cancer

We applied "Biop-C" to analyze three NPC tissue samples, i.e., "S009," "S010," and "S012." The tumor cores were collected by needle biopsies and immediately flash-frozen in liquid nitrogen (**Supplementary Figure 1**). The tumor cores were approximately weighed in the range 3–10 mg, and the clinical characteristics of the samples are listed in **Supplementary Table 2**. To prepare the tissue for analysis of 3D genome organization, we used a liquid nitrogen-cooled mini mortar and pestle for the pulverization of tissue within a microtube. This approach kept the biopsy sample frozen and reduced the risk of sample degradation. Additionally, this approach also provided the flexibility of

---

[1] https://github.com/4dn-dcic/hic2cool

performing the workflow from sample acquisition to Hi-C library preparation in a single microtube, which further minimized potential sample losses.

After pulverization, we processed the samples with a commercially available Arima *in situ* Hi-C kit. Briefly, chromatin was fixed with formaldehyde in the nucleus and digested with a restriction enzyme. Then, overhangs were filled in with biotinylated nucleotides followed by proximity ligation. After ligation, crosslinks were reversed, and the DNA was purified from protein. Furthermore, the purified DNA was sheared to ~350 bp mean fragment size. Finally, the sequencing libraries were generated using low input swift bioscience Illumina-compatible adapters (**Figure 1A**). The usage of the mini mortar and pestle followed by Hi-C is the key innovation of the Biop-C method. While this improvement in sample processing is a small change, it is highly effective in generating high-quality chromatin interaction libraries from needle biopsy clinical samples (**Figure 1C**).

Finally, we sequenced Biop-C libraries deeply by Illumina next-generation sequencing using a HiSeq4000 machine. Each library contained between 450 and 922 M contacts (**Supplementary Table 3**). We obtained more than 200 million mapped/valid junction reads (>50% of total read pairs) for each library, reflecting that our Biop-C datasets are adequately complex (Lajoie et al., 2015) (**Figure 1B** and **Supplementary Table 3**). Furthermore, the low ratios of the number of *trans* to *cis* contacts indicate high library quality for all samples (Nagano et al., 2015) (**Supplementary Table 3**). Notably, in some samples (e.g., S009), only one lane of HiSeq4000 sequencing was sufficient to obtain a high-quality Biop-C library.

Additionally, for comparison with a typical Hi-C library, we generated two replicates of HK1 NPC cell line by traditional *in situ* Hi-C Arima kit (Dixon et al., 2012; Rao et al., 2014). Moreover, for comparison with normal nasopharynx, we generated two replicates of NP-69 near-normal nasopharynx cell line by traditional *in situ* Hi-C.

We used Juicer for processing the resulting data, and the package Arrowhead was used to annotate TADs genome-wide, while the package HiCCUPS was used to call loops (Rao et al., 2014). Heatmaps were visualized with Juicebox (Durand et al., 2016). We were able to successfully call TADs and loops from our libraries (**Figure 1B** and **Supplementary Tables 3**, **4**), permitting comprehensive mapping of putative super and typical enhancer–promoter interactions in these samples (**Supplementary Table 6**). The TADs were clearly identifiable at a resolution of 10 kb. A high number of significant chromatin interactions could also be called using FitHiC2 (Kaul et al., 2020) at a resolution of 50 kb (**Supplementary Table 7**).

Patient S009 had 1,309 TADs, while patient S010 had 1,453 TADs and patient S012 had 1,516 TADs (**Figure 1B**). Loops could be identified at 5, 10, and 25 kb resolutions in all datasets, and these loops were all merged together for subsequent analyses. Patient S009 had 4,730 merged loops, while patient S010 had 6,539 merged loops, and patient S012 had 2,546 merged loops (**Figure 1B**).

We also called TADs and loops for NP-69 and found 5,227 TADs and 11,415 loops (**Supplementary Figure 4**). To better understand how NPC chromatin loops differ from normal loops found in a near-normal nasopharyngeal cell line, we compared chromatin loops in NP-69 with the ones called in NPC Biop-C samples. For comparison, we aggregated loops from the three samples (S009, S010, and S012) and compared them with NP-69 loops. We found that 4,026 loops (39.6% of aggregated NPC loops) were similar in NPC samples and NP-69. By contrast, 6,158 (60.4% of aggregated NPC loops) were specific to NPC, and 7,388 (72.5% of NP-69 loops) loops were specific to NP-69. We then associated these loops with gene promoters and found that the NP-69-specific loops (7,388) were associated with 11,167 gene promoters and the NPC-specific loops (6,158) were associated with 6,303 gene promoters (**Supplementary Figure 4**). We also found some examples of genes which were only associated with loops in NPC like *MMP3*, *CASP3*, and *ULK1* (**Supplementary Figure 4**). Overall, our results indicate that while some loops are similar between NP-69 near-normal cell line and NPC samples, there are also many loops that differ between them, suggesting that these may be cancer-specific loops that regulate oncogenes.

We looked for structural variants in biopsy samples S009, S010, and S012 as well as cell lines HK1 and NP-69. In S009, five translocations were identified out of which four can be seen in Hi-C heatmaps as well (**Supplementary Table 8** and **Supplementary Figure 5A**). In S010, no translocation was recognized, and in S012, we found six translocation incidences and all of these can also be seen in Hi-C heatmap (**Supplementary Table 8** and **Supplementary Figures 5B,C**). In cell lines, we found 63 translocation incidences in HK1 and 23 in NP-69 (**Supplementary Table 8**). Many of these translocations can be very clearly seen in the Hi-C heatmaps (**Supplementary Figures 5D,E**).

We could also recognize A/B compartments in our Biop-C samples (**Supplementary Table 9** and **Supplementary Figures 6A–C**) as well as in NP-69 cell line (**Supplementary Figure 6D**). We then wanted to compare these A/B compartments within each sample as well as NP-69 by calculating the Pearson correlation coefficient (Pearson's *r*) of the eigenvalues called with 500 kb resolution (**Supplementary Table 9**). The Pearson's *r* for S009 compared with S010 is 0.83, for S009 and S012 0.74, and for S010 and S012 0.85. We then compared the eigenvalues between NP-69 and biopsy samples. The Pearson's *r* for NP-69 compared with S009, S010, and S012 is 0.60, 0.64, and 0.54, respectively (**Supplementary Figures 6E–J**). These results show that the correlation between A/B compartments within biopsy samples is higher as compared with NP-69.

Moreover, because "FIREs" are a new type of chromatin interaction landmark associated with super enhancers and tissue-specific chromatin interactions (Schmitt et al., 2016), we used FIREcaller R package (Crowley et al., 2019) to call FIREs. We identified 2,783 FIREs in NPC sample S009, 1,393 FIREs in sample S010, and 2,906 FIREs in sample S012 from our Biop-C data. We also called FIREs from the NPC cell line HK1 and identified 3,585 FIREs (**Supplementary Table 10** and **Figure 1B**).

Next, we examined the chromatin interactions around important oncogenes in NPC, such as *MYC* (Yu et al., 2003) (**Figure 2A**) and epidermal growth factor receptor (*EGFR*) (**Supplementary Figure 2A**) (Fujii et al., 2002; Chua et al., 2004;
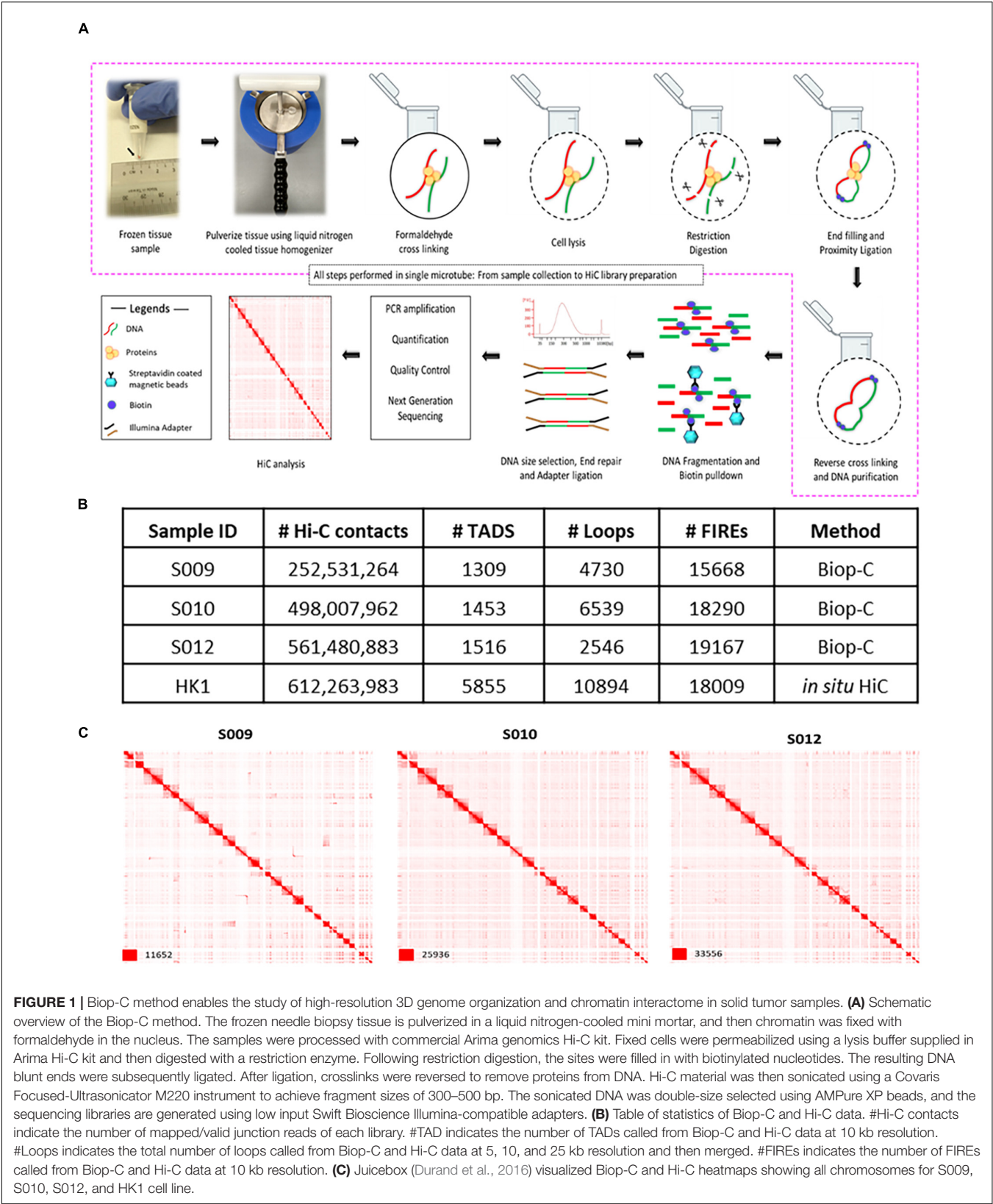
FIGURE 1 | Biop-C method enables the study of high-resolution 3D genome organization and chromatin interactome in solid tumor samples. **(A)** Schematic overview of the Biop-C method. The frozen needle biopsy tissue is pulverized in a liquid nitrogen-cooled mini mortar, and then chromatin was fixed with formaldehyde in the nucleus. The samples were processed with commercial Arima genomics Hi-C kit. Fixed cells were permeabilized using a lysis buffer supplied in Arima Hi-C kit and then digested with a restriction enzyme. Following restriction digestion, the sites were filled in with biotinylated nucleotides. The resulting DNA blunt ends were subsequently ligated. After ligation, crosslinks were reversed to remove proteins from DNA. Hi-C material was then sonicated using a Covaris Focused-Ultrasonicator M220 instrument to achieve fragment sizes of 300–500 bp. The sonicated DNA was double-size selected using AMPure XP beads, and the sequencing libraries are generated using low input Swift Bioscience Illumina-compatible adapters. **(B)** Table of statistics of Biop-C and Hi-C data. #Hi-C contacts indicate the number of mapped/valid junction reads of each library. #TAD indicates the number of TADs called from Biop-C and Hi-C data at 10 kb resolution. #Loops indicates the total number of loops called from Biop-C and Hi-C data at 5, 10, and 25 kb resolution and then merged. #FIREs indicates the number of FIREs called from Biop-C and Hi-C data at 10 kb resolution. **(C)** Juicebox (Durand et al., 2016) visualized Biop-C and Hi-C heatmaps showing all chromosomes for S009, S010, S012, and HK1 cell line.
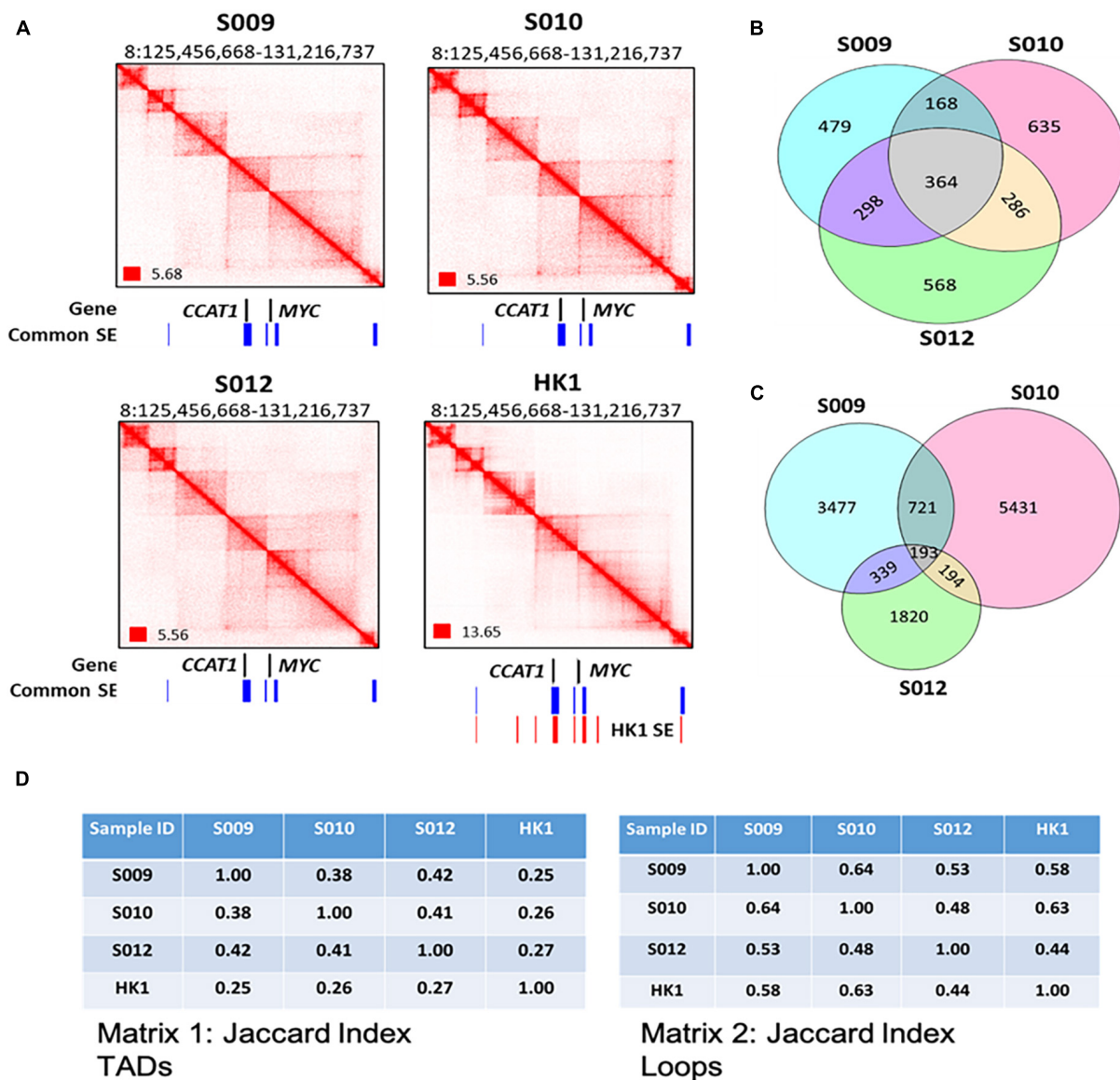
FIGURE 2 | A genome-wide map of 3D genome organization in nasopharyngeal cancer. **(A)** Biop-C heatmaps can detect various conformational genomic features such as TADs, loops, and FIREs in S009, S010, and S012 at *CCAT1* and *MYC*. Coverage normalization was used to visualize the Biop-C and Hi-C heatmaps in Juicebox (Durand et al., 2016). Genes are indicated in black color. The "common super enhancers," shown in blue color, indicate the super enhancers present in all three NPC cell lines—HK1, C66-1, and HNE1 cell lines. The super enhancers present in HK1 cell lines are indicated in red color. The super enhancer datasets are obtained from Ke et al. (2017). **(B)** Venn diagram showing the overlap of TAD boundaries between the patients inferred using the Arrowhead algorithm (Rao et al., 2014) at 10 kb resolution. **(C)** Venn diagram depicting the overlap of loops between the patients. The loops were called using the HiCCUPS algorithm (Rao et al., 2014) at 5, 10, and 25 kb resolution. The loops of the different resolutions were merged for this analysis. **(D)** Jaccard index. Matrix 1: Jaccard indices for TADs in NPC samples S009, S010, and S012 and cell line HK1; each cell in the matrix indicates the Jaccard index value for the column sample and row sample. Matrix 2: Jaccard indices for loops in NPC samples S009, S010, and S012 and cell line HK1 where each cell of the matrix indicates the Jaccard index value for the column and row sample.

Ruan et al., 2011). *MYC* is overexpressed in 76% of the NPC patients. The patients with *MYC*-positive tumors had a longer disease-free period (Yu et al., 2003). *EGFR* is highly expressed in nearly 85% of NPC patients and associated with a significantly poorer prognosis in patients with advanced nasopharyngeal cancer than in patients without *EGFR* overexpression (Fujii et al., 2002; Chua et al., 2004). We found that *EGFR* is marked by three super enhancers in the HK1 cell line: two of these three SE are localized upstream, i.e., near the start transcription site, and the third one is located in an intron (**Supplementary Figure 2B**).

We further examined if there were any significant single nucleotide polymorphisms (SNPs) that were associated with loops in our NPC Biop-C samples. We observed an overlap of a loop anchor with a region at the CDKN2B(-AS1) locus in

samples S009 and S010 (**Supplementary Table 11**), which was previously identified by a GWAS study on NPC in patients of Chinese descent (Bei et al., 2016).

Additionally, for comparison with a typical Hi-C library, we examined the two replicates of HK1 NPC cell line that we generated by traditional *in situ* Hi-C (Arima kit) (Dixon et al., 2012; Rao et al., 2014). Visual inspection of coverage normalized Biop-C heatmaps of three NPC tissue samples, and Hi-C maps of the HK1 cell line showed that the libraries were largely similar to each other, although we also noted that certain loci contained differences suggesting patient heterogeneity which we explore further in the next section of this manuscript (**Figures 3A,B** and **Supplementary Figures 3A,B**).

To determine the reproducibility of our Biop-C method, we took one NPC patient sample, and cut it into half before performing the Biop-C method on the two separate halves. These are referred to as S024_R1 and S024_R2. We performed Juicer pipeline for calling TADs (247 in replicate 1 and 8 TADs in replicate 2) and loops (5,094 in replicate 1 and 2,199 in replicate 2) (**Supplementary Figure 7A**). We also called FIREs and found 1,707 FIREs in S024_R1 and 2,893 FIREs in S024_R2 (**Supplementary Figure 7A**). We used the HiCRep software to calculate the similarity between these replicates and found that 16 out of 24 chromosomes have high correlation (SCC scores above 0.8), while the other 8 chromosomes have SCC scores between 0.6 and 0.8, which shows that the two replicates are highly correlated (**Supplementary Table 12**). We also wanted to compare the similarity between S024 sample replicates with NPC cell line and a cell line from a different cancer. We found that the NPC samples showed higher similarity with HK1 (NPC cell line) as compared with T47D (breast cancer cell line) (**Supplementary Figure 7B**). We also did A/B compartment analysis for the two replicates and found a very high correlation between them (Pearson's $r = 0.95$) (**Supplementary Table 9** and **Supplementary Figures 7C–E**). These above results show that our Biop-C method is indeed reproducible.

Overall, our successful detection of TADs, loops, structural variants, A/B compartments, and FIREs suggests that our Biop-C method can generate high-quality genome-wide chromatin conformation maps from the solid tumor needle biopsy samples. Our evaluation using the S024 tumor cut in half and analyzed by two separate Biop-C libraries indicates that our Biop-C method is reproducible.

Patient Heterogeneity in Chromatin Interactions

The question of patient heterogeneity is relatively unexplored in chromatin interaction analyses. In our previous research investigating chromatin interactions at the *TP53* and *MYC* loci, we observed that some chromatin interactions at MYC and TP53 could be detected in bone marrow and peripheral blood samples, but not all chromatin interactions that were observed in K562 cells were detected in clinical samples (Cao et al., 2017).

To investigate potential patient heterogeneity, we compared the TADs and loops between the Biop-C heatmaps of patients directly (**Figures 2B,C**) and using the Jaccard index. We

calculated Jaccard's similarity coefficient (Jaccard index, JI) to measure the overlap between the called TADs and loops in three Biop-C matrices. The resulting JI value indicates the fraction of shared TAD boundaries and loops between the patients. We observed that 38–42% of the TADs and 53–64% of loops are shared in the three patients (**Figure 2D**). We can conclude that the NPC samples show chromatin interaction heterogeneity. But since our samples are clinical samples, we cannot rule out the possibility that the patient chromatin interactome differences are due to the tumor heterogeneity and/or surrounding normal cells in the tumor.

Next, we examined individual specific chromatin interactions. We found that chromatin interactions for genomic locations *EGFR*, *PTPN1*, *DDIT4*, *MIR205HG*, *PDGFA*, *MALAT1*, *CAV2*, *NOTCH1*, *TEAD1*, *TP63*, *RUNX1*, *CCAT1*, *MYC*, and *YAP1* (**Supplementary Figure 2A**) are similar and *FOXA1*, *MIPOL1*, *SP4*, *SGCZ*, *MROH9*, *FMO1*, *FMO2*, *FMO1*, *FMO4*, and *FMO6P* gene are different (**Figures 3A,B** and **Supplementary Figures 3A,B**). In one example, we observed that two loops, i.e., loop 1 and loop 2 are present near *FOXA1* and *MIPOL1* genes in S009 and S012, which are thought to be tumor suppressors in nasopharyngeal cancer (Kwok et al., 2020; Leong et al., 2020). However, these loops were absent in S010 (**Figures 3A,B**). In another example, we observed a loop only in S009 and absent in S010 and S012 near miR383, which is considered an excellent diagnostic biomarker for head and neck cancers (Liu et al., 2019) (**Supplementary Figure 3A**). We also observed extensive chromatin looping near *MROH9*, *FMO1*, *FMO2*, *FMO1*, *FMO4*, and *FMO6P* locus in S009, which was absent in S010 and S012 (**Supplementary Figure 3B**).

Next, as we had observed patient-specific chromatin interactions, we investigated the tissue specificity of these chromatin interactions. Thus, we characterized the similarities and differences between the NPC landscape and other tissue types. We compared the Biop-C heatmaps and the Hi-C heatmap from HK1 with the previously published Hi-C heatmaps in human cell lines such as K562 (chronic myelogenous leukemia cell line) (Rao et al., 2014), HAP1 (near-haploid cell line) (Haarhuis et al., 2017), IMR90 (fetal lung fibroblast cell line) (Rao et al., 2014), KBM7 (chronic myelogenous leukemia) (Rao et al., 2014), HUVEC (human umbilical vein endothelial cell line) (Rao et al., 2014), RPE1 (retinal pigment epithelium cell line) (Darrow et al., 2016), GM12878 (lymphoblastoid cell line) (Rao et al., 2014), NHEK (normal human epidermal keratinocytes) (Rao et al., 2014), HeLa (human cervical carcinoma cell line) (Rao et al., 2014), HCT116 (colon cancer cell line) (Rao et al., 2014), and HMEC (mammary epithelial cell line) (Rao et al., 2014) (**Supplementary Figures 2C–F**).

We observed that the Hi-C heatmaps of K562, HAP1, IMR90, KBM7, HUVEC, and RPE1 show a similar pattern as the Biop-C heatmaps of S009, S010, and S012 and HK1 Hi-C heatmap for the genomic locations around *MYC* and *CCAT1*, but differences in the profile were observed in GM12878, NHEK, HeLa, HCT116, and HMEC (**Supplementary Figure 2C**). The genome organization at the *RUNX1* locus was similar between the Biop-C heatmaps of our patient in K562, HAP1, IMR90, RPE1, NHEK1, HeLa, HCT116, and HMEC but different in
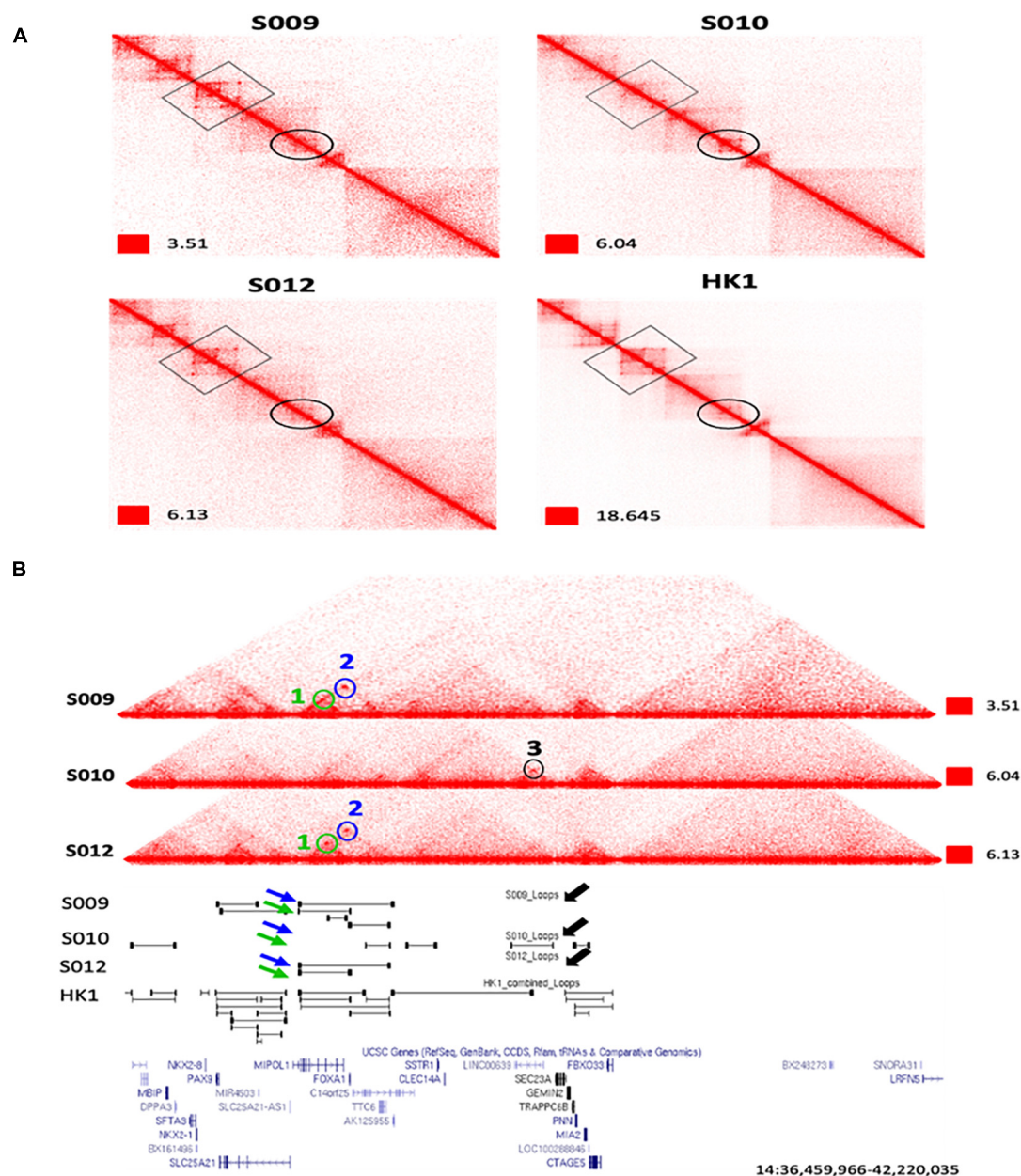
FIGURE 3 | An example of extensive heterogeneity between patients in the chromatin interactome in nasopharyngeal cancer at the *FOXA1* and *MIPOL1* genes. **(A)** Zoomed view of Biop-C Juicebox-visualized heatmaps of S009, S010, and S012 showing patient heterogeneity in chromatin interaction at the *FOXA1* and *MIPOL1* genes. **(B)** UCSC genome browser (Kent et al., 2002) screenshots of genomic coordinates chr14: 36,459,966–42,220,035. Loops 1 and 2 are present in S009 and S012, while loop 3 is present only in S010. Loop 1, loop 2, and loop 3 are represented in green, blue, and black colors, respectively.

KBM7, HUVEC, and GM12878 (**Supplementary Figure 2D**). However, the genomic patterns for the *PTPN1* locus were found to be similar in all the Hi-C and Biop-C heatmaps of the patient (**Supplementary Figure 2E**). On the other hand, the genomic patterns for the *MALAT1* locus [*MALAT1* is known to promote cell proliferation in gastric cancer (Wang et al., 2014)] did not show any similarity and were different in all the Hi-C and Biop-C heatmaps (**Supplementary Figure 2F**). We conclude that certain chromatin interactions in nasopharyngeal cancer are

common across tissue types (*PTPN1*) and certain regions are tissue specific (*MALAT1*).

The observation of patient heterogeneity and tissue specificity in TADs appears to contradict earlier observations that TADs are primarily conserved across different human cell types and possibly even across different species (Dixon et al., 2012; Jin et al., 2013; Rao et al., 2014). However, Sauerwald analyzed 137 Hi-C samples from nine studies and observed significant TAD variations across human cell and tissue types
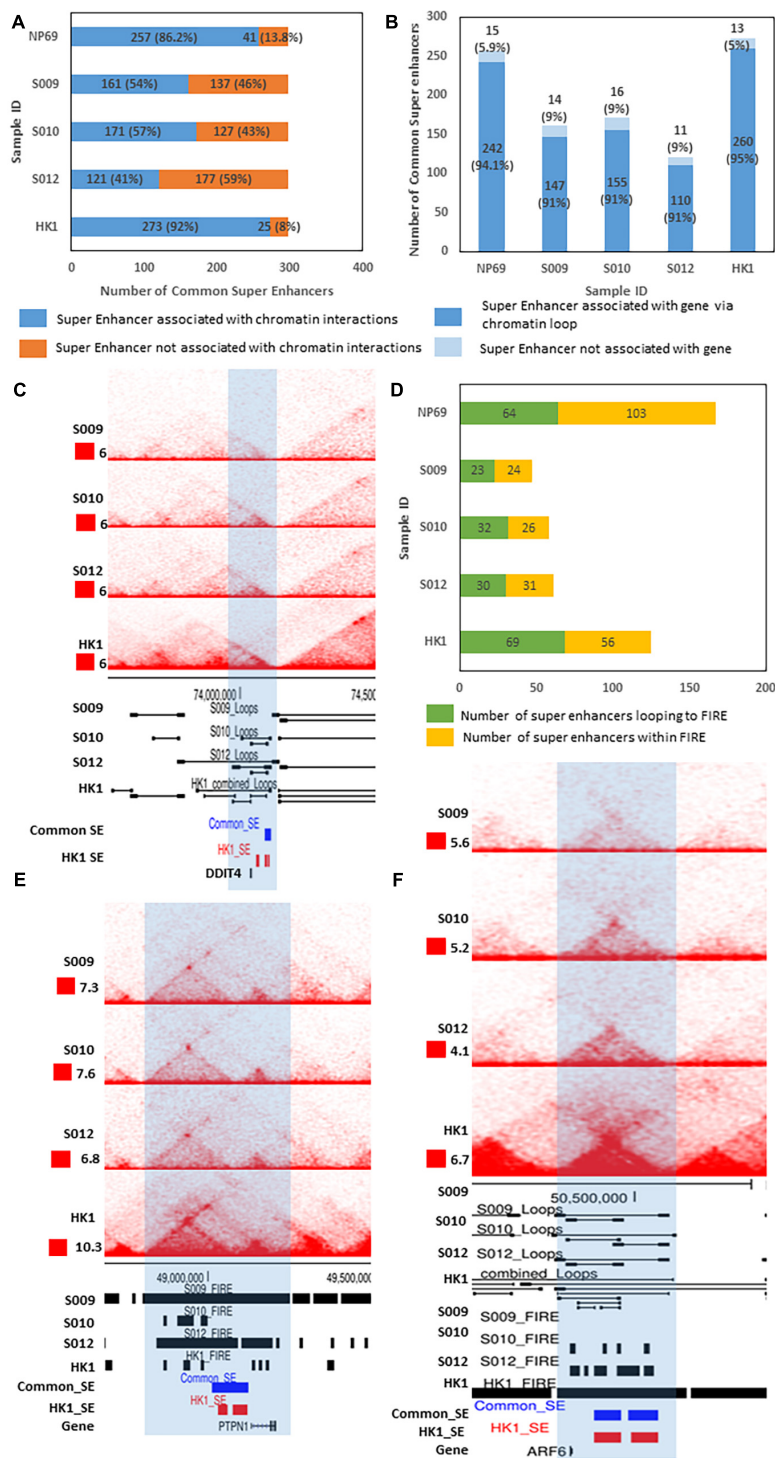
FIGURE 4 | Association of super enhancers with chromatin interactions and genes. **(A)** Graph showing the number of common SE associated with chromatin interactions (blue) and the number of SE which are not associated with chromatin interactions (orange). **(B)** Graph showing the number of looping SE (chromatin interaction-associated SE) associated with distant genes *via* chromatin loops (dark blue) and the number of SE (chromatin interaction-associated SE) which do not link to distant genes *via* chromatin loops (light blue). **(C)** Biop-C heatmaps for samples S009, S010, and S012 and Hi-C heatmap for the cell line HK1 for the DDIT4 gene locus (zoomed-in view of genomic locus chr10:72974237–75095236) where a SE links to the gene through a chromatin loop. **(D)** Graph showing the number of SE associated with FIREs *via* chromatin loop (green) and the number of SE which are within FIRE (yellow). **(E)** Biop-C heatmaps for samples S009, S010, and S012 and Hi-C heatmap for the cell line HK1 (zoomed-in view of locus chr20: 48422029–49905948) showing a SE (blue: common SE, red: HK1 SE) and PTPN1 gene within the same FIRE. **(F)** Biop-C heatmaps for samples S009, S010, and S012 and Hi-C heatmap for HK1 (zoomed-in view of locus chr14: 49,497,144–51,494,634) showing a SE looping to FIRE at the ARF6 gene locus.
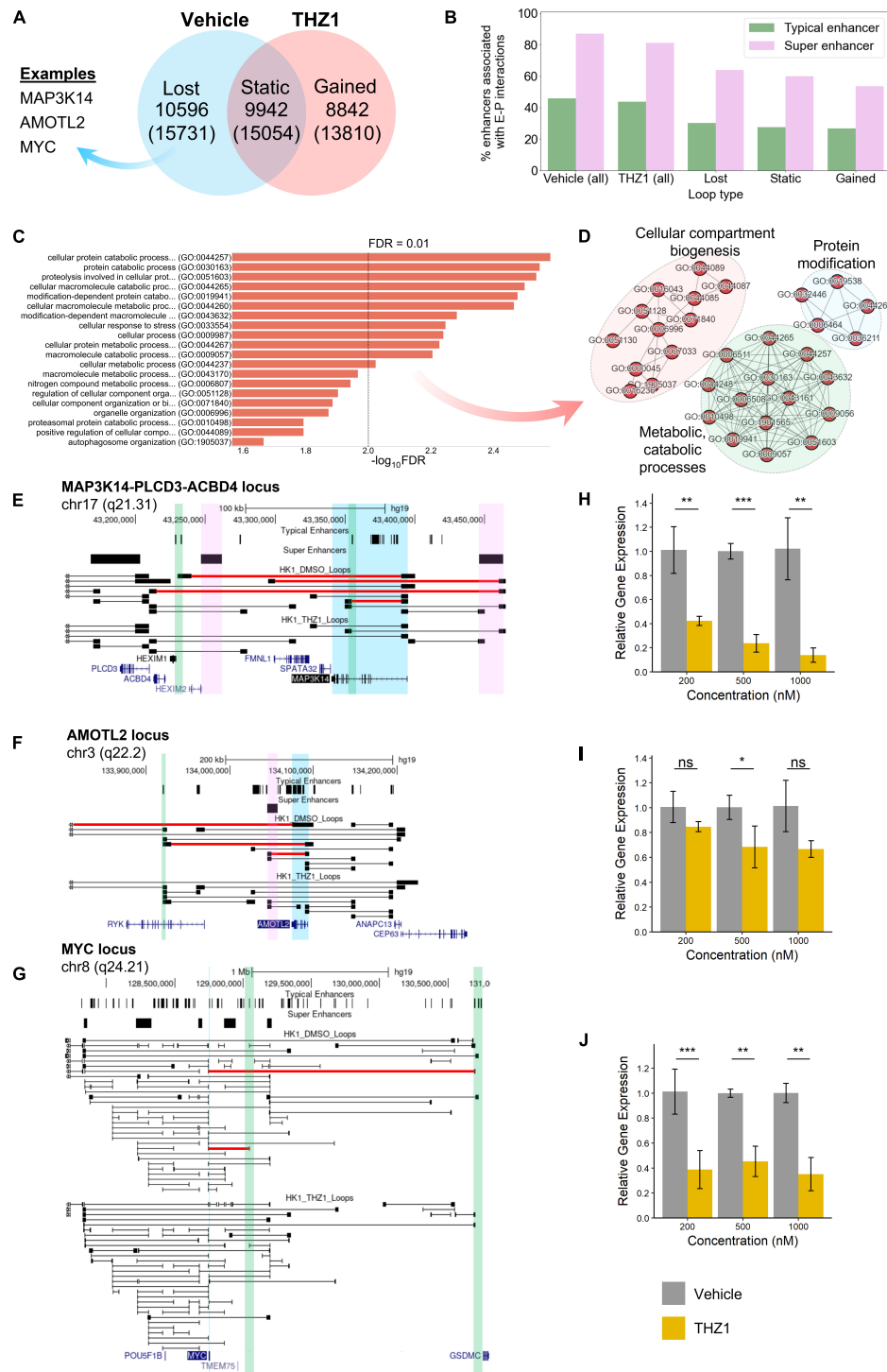
FIGURE 5 | (A) Functional analysis of HK1 chromatin interactome following treatment with CDK7 inhibitor THZ1. (A) Venn diagram representing the number of loops lost after THZ1 treatment compared with DMSO (vehicle) treatment, unchanged between treatments, and gained after treatment. The number of genes with promoter regions within 15 kb of the respective loop anchors is shown in the parentheses below. (B) Anchor occupancy characteristics of super enhancers and typical enhancers predicted from H3K27ac ChIP-Seq. A higher percentage of super enhancers is associated with enhancer–promoter loops, compared with typical enhancers. (C) Top 20 overrepresented GO-Slim Biological Process terms by false discovery rate (FDR), derived from the set of protein-coding genes with net lost loops following THZ1 treatment. (D) Network representation of predominantly enriched GO term clusters. (E–G) Looping structure of MAP3K14, AMOTL2, and MYC loci. Lost loops following THZ1 treatment are shown in red. Gene of interest is highlighted in blue. Select super enhancers with lost loops to promoters are highlighted in pink. Select typical enhancers with lost loops to promoters are highlighted in green. (H–J) RT-qPCR results for MAP3K14, AMOTL2, and MYC, respectively, following 24 h of vehicle or THZ1 treatment (*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$).

(Sauerwald et al., 2020), suggesting that while there are common TADs and loops, there are also TADs and loops that vary across patient samples and tissue types.

Super Enhancers Are Associated With Frequently Interacting Regions and Loop to Genes

Super enhancers are regions of the DNA which enhances the transcription of target genes. These are comprised of a group of enhancers which are at close proximity to each other and are marked by high enrichment of H3K27ac histone modification (Pott and Lieb, 2015). In previous research, we and others have shown that super enhancers can regulate distant genes *via* long-range chromatin interactions (Cao et al., 2017; See et al., 2019). Moreover, FIREs have been reported to form at genomic regions also enriched by super enhancers (Schmitt et al., 2016). Consequently, we wished to understand the relationship between super enhancers and chromatin interactions in nasopharyngeal cancer.

As the biopsy samples were too small for us to obtain both H3K27ac ChIP-Seq data as well as Biop-C data, we identified SE from NPC cell lines HK1, C666-1, and HNE1 using the ChIP-Seq data from Ke et al. (2017) (**Supplementary Table 13**) and we found that 298 SE were common in all the three cell lines. We reasoned that these "common" super enhancers that are present in all cell lines examined will most likely also be present in the patient samples examined and hence used them for downstream analysis. We then associated these common super enhancers with chromatin loops obtained from the Biop-C data of the patient samples as well as from the Hi-C data of HK1 and NP-69 cell lines. As a result, we found that these SE are highly associated with chromatin loops. In samples S009, S010, and S012, we found 54, 57, and 41% (respectively) SE which were associated with chromatin loop anchors within 15 kb distance ("looping SE") (**Figure 4A**).

We further looked for associations between these looping SE and genes (**Supplementary Figure 8A**). We observed that more than 90% of looping SE loop to distant genes (**Supplementary Figure 8A**). In S009, 95% of looping SE are linked to distant genes *via* these loops. In S010 as well as in S012, 91% of looping SE are linked to distant genes (**Figure 4B** and **Supplementary Table 13**). We also associated these SE to chromatin loops in cell line HK1. We found that 92% of common super enhancers are associated with chromatin loops in HK1 and 91% of these looping SE are linked to distant genes (**Figures 4A,B** and **Supplementary Table 15**). For example, the *DDIT4* gene (one of the top ranked SE-associated gene in the HK1 cell line) (Ke et al., 2017) is also associated with a distant SE in patient samples S010 and S012 and NPC cell line HK1 (**Figure 4C**). We repeated the above analysis with cell line NP-69 and found that 86.2% of SE are associated with chromatin interactions and 94.1% of looping SE are linked to distal genes (**Figures 4A,B**). This coincides with our previous results that the number of chromatin loops lost is more than the acquired loops in NPC, and hence, we

can see more SE association with loops in NP-69 as compared with NPC samples.

We also repeated this analysis with common enhancers within cell lines HK1, C661, and HNE1. In samples S009, S010, and S012, the association of enhancers with chromatin loops was 24, 29, and 18%, respectively, and about 60% (63, 65, and 64% in S009, S010, and S012, respectively) of these chromatin interaction-associated enhancers loop to distal genes (**Supplementary Figures 8D,E**). We also repeated this analysis with cell lines HK1 and NP-69 and found that 44.3 and 43.8% SE are associated with loops, respectively, and about 56% (56.2% in NP-69 and 55.6% in HK1) of these chromatin interaction-associated enhancers loop to distal genes (**Supplementary Figures 8D,E**). From these results, we conclude that the association of enhancers with chromatin loops is much less as compared with SE in NPC patients. In NP-69, the association of enhancers with chromatin loops is higher compared with NPC samples but less as compared with SE in NP-69.

Next, we categorized common SE based on their proximity to a gene: the ones which were at close proximity (less than 15 kb from a gene) to a gene were called "proximal" SE and the ones which are away from the gene and associated *via* chromatin loops were called distal SE. Out of 298 common SE tested, we found 27 proximal SE in all the Biop-C samples and 147 distal SE in S009, 155 distal SE in S010, 110 distal SE in S012, and 260 distal SE in HK1. We observed that 48 genes were associated with proximal SE (**Supplementary Table 14**), while 356 genes in S009, 421 genes in S010, 291 in S012, and 944 genes in HK1 were associated with distal SE (**Supplementary Table 16**). There were also some genes that have both proximal and distal SE: in S009, we found 10 genes; in S010 as well as S012, we found 6 genes, and in HK1, we found 18 such genes (**Supplementary Table 16**). For example, the *MACF1* gene in all the three NPC samples as well as HK1 cell line has distal as well as proximal SE (**Supplementary Table 16**). Gene set overrepresentation analyses (GSOA) on the set of genes associated with SE showed significant overrepresentation of Biological Process Gene Ontology (GO) terms such as metabolic process and protein modification/phosphorylation and regulation (**Supplementary Figure 9**).

Subsequently, we wanted to associate these SE with FIREs called (**Supplementary Table 10**) from the Biop-C data from patient samples as well as HK1 and NP-69 Hi-C data. We looked for two types of associations between SE and FIREs: SE within FIRE and SE that loop to FIRE (**Supplementary Figures 8B,C**). We could recognize 24 common SE (8% of SE) in S009, 26 (9% of SE) in S010, 31 (11% of SE) in S012, 56 (19% of SE) in HK1, and 103 (35% of SE) in NP-69 which are within a FIRE (**Figure 4D**). Upon combining the chromatin loops data with FIRE calling, we found 23 (8% of SE) in S009, 32 (11% of SE) in sample S010, 30 (10% of SE) in S012, 69 (23% of SE) in HK1, and 64 (22% of SE) in NP-69 cell line that loops to a FIRE (**Figure 4D** and **Supplementary Table 15**). We also found some examples of SE associated with genes *via* FIRE. For example, SE which falls within the same FIRE as the gene *PTPN1* (**Figure 4E**) and SE which loops to a distant FIRE containing the *ARF6* gene whose overexpression can be

correlated with metastasis and invasion in several cancers (Li et al., 2017) (**Figure 4F**).

We also repeated this analysis with enhancers and found 512 (3.8% of total enhancers), 270 (2% of total enhancers), and 524 (4% of total enhancers) enhancers within FIRE in S009, S010, and S012, respectively, and 242 (1.8% of total enhancers), 285 (2.1% of total enhancers), 189 (1.4% of total enhancers) enhancers looping to FIRE in samples S009, S010, and S012 (**Supplementary Figure 8F**). In the cell lines, we found 783 (5.8% of total enhancers) and 868 (6.5% of total enhancers) enhancers within FIRE in HK1 and NP-69, respectively; 539 (4% of total enhancers) and 868 (6.4% of total enhancers) enhancers looped to FIRE in HK1 and NP-69 (**Supplementary Figure 8F**). Based on these findings, we can conclude that SE are associated with FIREs in NPC. However, the SE association with FIRE is higher in NP-69 as compared with the NPC patient sample.

THZ1 Treatment Leads to Loss of Specific Super Enhancer– and Typical Enhancer–Promoter Chromatin Interactions at Key Oncogenic Loci

To investigate the functional significance of SE–promoter loops in NPC, we examined changes in the chromatin interactome *via* conventional *in situ* Hi-C, following the treatment of HK1 cells with the CDK7 inhibitor THZ1. In several cancers, THZ1 treatment leads to decreased cell viability (Chipumuro et al., 2014; Kwiatkowski et al., 2014). The binding of THZ1 to CDK7 was previously shown to lead to inhibited CDK7-mediated phosphorylation of RNA Pol II, which coincides with a loss of transcription factor binding at distal sites with high H3K27ac (Sampathi et al., 2019), representing putative super and typical enhancer loci (Hnisz et al., 2013). SE-associated genes are exceptionally sensitive to perturbation by THZ1 treatment (Ke et al., 2017). However, alterations in the chromatin interactome upon CDK7 inhibition in NPC cells have not been previously explored.

To investigate this, we performed *in situ* Hi-C on THZ1 and DMSO (vehicle)-treated HK1 cells at 500 nM concentration for 24 h. THZ1 treatment led to clear changes in the chromatin interactome: we observed that 51.6% (10,596) of the loops present under vehicle treatment were lost following THZ1 treatment, whereas 48.4% (9,942) were conserved. Interestingly, 8,842 loops were gained following THZ1 treatment (**Figure 5A**).

As with NPC Biop-C samples, we observed that approximately 80% of all super enhancers identified by HK1 H3K27ac ChIP-Seq were associated with long-range enhancer–promoter interactions, compared with only 40% of typical enhancers, regardless of vehicle or THZ1 treatment (**Figure 5B** and **Supplementary Table 16**). Gene set overrepresentation analyses on the genes with net lost loops revealed a significant enrichment of GO terms associated with cellular stress response, compartment/organelle biogenesis, metabolic/catabolic processes, and protein modification (**Figures 5C,D**). Clusters relating to metabolic processes and protein modification were also observed for the SE-associated genes identified in the three NPC Biop-C samples (**Supplementary Figure 8**), further

suggesting that SE are a common dependency-mediating expression at the pathway level.

Subsequently, to determine if lost loops to super or typical enhancers could explain THZ1-mediated loss of viability and gene expression, we performed a close examination of the looping behavior in three candidate genes, namely, *MAP3K14*, *AMOTL2*, and *MYC*, which are known to be involved in NPC pathogenesis (**Figures 5E–G**). At the *MAP3K14* and *AMOTL2* loci, we observed the loss of super and typical enhancer-associated loops to these promoters, whereas at the *MYC* locus, we only observed the loss of typical enhancer-associated loops. *Via* RT-qPCR, we additionally observed decreased expression of all three genes following THZ1 treatment, compared with vehicle (**Figures 5H–J**).

We also compared the A/B compartments of HK1 (control) and HK1–THZ1 treated and found a high correlation between them (Pearson's $r = 0.95$) suggesting that the A/B compartments remain conserved upon THZ1 treatment (**Supplementary Figure 10**).

As a whole, these results suggest that THZ1 treatment leads to specific perturbations of the chromatin interactome. Some of the lost loops correspond to specific SE– and typical enhancer–promoter interactions that are involved in the control of expression at these loci. THZ1 treatment leads to downregulation of expression at several of these loci, further suggesting the regulatory role of these SE- and typical enhancer-associated loops.

CONCLUSION

Taken together, our new Biop-C method is suitable for interrogating needle biopsy patient samples and, more generally, situations of limited tumor sampling when surgical biopsies may be technically difficult. Using Biop-C, we examined chromatin interactions in three nasopharyngeal cancer patient samples, which allowed us to identify super enhancers associated with FIREs and which loop to important oncogenes. We also demonstrated patient heterogeneity in chromatin interactions in these patient samples, as well as tissue specificity. Hi-C libraries from an NPC cell line, HK1, showed differences compared with chromatin interactions in the patient samples. These differences could arise due to different subtypes of NPC.

Upon comparison with near-normal nasopharynx cell line NP-69, we found that while there were some loops that were similar between NPC samples and NP-69, there were also loops that were different, suggesting that there may be NPC-specific loops that could potentially regulate NPC oncogenes. To test the reproducibility of our method, we also performed Biop-C on two replicates of the same patient sample and found high correlation between the two replicates in most of the chromosomes (16 out of 24). Upon comparison with cell lines, we found that NPC patient replicates were highly correlated with the NPC cell line HK1 as compared with the breast cancer cell line T57D. We also observed that SE are much more associated with chromatin interactions and FIREs as compared with enhancers; however, these associations are higher in

NP-69 as compared with NPC samples, which is consistent with our previous findings that more loops are lost in NPC as compared with those acquired. Our results indicate the necessity of interrogating chromatin interactions in actual patient cancers, which Biop-C enables. Our results demonstrate the utility and importance of Biop-C as a method for understanding cancer 3D genome organization. Additionally, our results suggest that THZ1 may be able to modulate chromatin interactions. In the future, we anticipate that the versatility of Biop-C will also allow us to interrogate perturbations of chromatin gene regulation in patients undergoing therapeutic interventions.

DATA AVAILABILITY STATEMENT

The datasets of Hi-C are available in GEO under accession number GSE166570 (reviewer token: yhizaucglrghlmp).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institute Review Board, number 2018/00947-SRF0002. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SA and MF conceived the research idea. MF, SA, and RC contributed to the study design. SA and MF developed the Biop-C method. SA, HJ, and LC prepared Biop-C and HiC libraries. SA and RC performed Hi-C analysis of the Biop-C and Hi-C samples. SA performed manual curation of the Hi-C data. RC performed FIRE calling, Structural Variants calling, A/B compartment analysis, replicate comparison, and ChIP-Seq analysis on published data to identify SEs and their looping patterns in NPC. BW performed Fithic2 calling, enhancer promoter association, loops comparison between NPC and NP-69 samples, and Gene ontology. SA and CK performed THZ1 treatment. BW performed RT-qPCR and data analysis. XY, JT, W-QC, and BG provided NPC clinical samples. SA, RC, and MF reviewed the data.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.673530/full#supplementary-material

REFERENCES

American Cancer Society (2020). *American Cancer Society. Cancer Facts & Figures 2020*. Atlanta. GA: American Cancer Society, 1–52.

Bei, J. X., Su, W. H., Ng, C. C., Yu, K., Chin, Y. M., Lou, P. J., et al. (2016). A GWAS meta-analysis and replication study identifies a novel locus within CLPTM1L/TERT associated with nasopharyngeal carcinoma in individuals of Chinese ancestry. *Cancer Epidemiol. Biomarkers Prev.* 25, 188–192. doi: 10.1158/1055-9965.EPI-15-0144

Cao, F., Fang, Y., Tan, H. K., Goh, Y., Choy, J. Y. H., Koh, B. T. H., et al. (2017). Super-enhancers and broad h3k4me3 domains form complex gene regulatory circuits involving chromatin interactions. *Sci. Rep.* 7:2186. doi: 10.1038/s41598-017-02257-3

Chipumuro, E., Marco, E., Christensen, C. L., Kwiatkowski, N., Zhang, T., Hatheway, C. M., et al. (2014). CDK7 inhibition suppresses super-enhancer-linked oncogenictranscription in MYCN-drivencancer. *Cell* 159:1126. doi: 10.1016/J.CELL.2014.10.024

Chua, D. T. T., Nicholls, J. M., Sham, J. S. T., and Au, G. K. H. (2004). Prognostic value of epidermal growth factor receptor expression in patients with advanced stage nasopharyngeal carcinoma treated with induction chemotherapy and radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* 59, 11–20. doi: 10.1016/j.ijrobp.2003.10.038

Crowley, C., Yang, Y., Qiu, Y., Hu, B., Abnousi, A., Lipiński, J., et al. (2021). FIREcaller: detecting frequently interacting regions from Hi-C data. *Comput. Struct. Biotechnol. J.* 19, 355–362. doi: 10.1016/j.csbj.2020.12.026

Crowley, C., Yang, Y., Qiu, Y., Hu, B., Won, H., Ren, B., et al. (2019). FIREcaller: an R package for detecting frequently interacting regions from Hi-C data. *bioRxiv* [Preprint] doi: 10.1101/619288 bioRxiv 619288,

Dai, W., Cheung, A. K. L., Ko, J. M. Y., Cheng, Y., Zheng, H., Ngan, R. K. C., et al. (2015). Comparative methylome analysis in solid tumors reveals aberrant methylation at chromosome 6p in nasopharyngeal carcinoma. *Cancer Med.* 4, 1079–1090. doi: 10.1002/cam4.451

Dai, W., Zheng, H., Cheung, A. K. L., and Lung, M. L. (2016). Genetic and epigenetic landscape of nasopharyngeal carcinoma. *Chin. Clin. Oncol.* 5:4. doi: 10.21037/CCO.V0I0.9478

Darrow, E. M., Huntley, M. H., Dudchenko, O., Stamenova, E. K., Durand, N. C., Sun, Z., et al. (2016). Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc. Natl. Acad. Sci. U.S.A.* 113, E4504–E4512. doi: 10.1073/pnas.1609643113

Díaz, N., Kruse, K., Erdmann, T., Staiger, A. M., Ott, G., Lenz, G., et al. (2018). Chromatin conformation analysis of primary patient tissue using a low input Hi-C method. *Nat. Commun.* 9:4938. doi: 10.1038/s41467-018-06961-0

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380. doi: 10.1038/s41467-018-06961-0

Du, Z., Zheng, H., Huang, B., Ma, R., Wu, J., Zhang, X., et al. (2017). Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* 547, 232–235. doi: 10.1038/nature23263

Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., et al. (2016). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* 3, 99–101. doi: 10.1016/j.cels.2015.07.012

Flavahan, W. A., Drier, Y., Liau, B. B., Gillespie, S. M., Venteicher, A. S., Stemmer-Rachamimov, A. O., et al. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* 529, 110–114. doi: 10.1038/nature16490

Fudenberg, G., Getz, G., Meyerson, M., and Mirny, L. A. (2011). High order chromatin architecture shapes the landscape of chromosomal alterations in cancer. *Nat. Biotechnol.* 29, 1109–1113. doi: 10.1038/nbt.2049

Fujii, M., Yamashita, T., Ishiguro, R., Tashiro, M., and Kameyama, K. (2002). Significance of epidermal growth factor receptor and tumor associated tissue eosinophilia in the prognosis of patients with nasopharyngeal carcinoma. *Auris Nasus Larynx* 29, 175–181. doi: 10.1016/S0385-8146(01)00135-3

Haarhuis, J. H. I., van der Weide, R. H., Blomen, V. A., Yáñez-Cuna, J. O., Amendola, M., van Ruiten, M. S., et al. (2017). The cohesin release factor WAPL restricts chromatin loop extension. *Cell* 169, 693–707.e14. doi: 10.1016/j.cell.2017.04.013

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., et al. (2013). XSuper-enhancers in the control of cell identity and disease. *Cell* 155:934. doi: 10.1016/j.cell.2013.09.053

Hnisz, D., Schuijers, J., Li, C. H., and Young, R. A. (2018). Regulation and dysregulation of chromosome structure in cancer. *Annu. Rev. Cancer Biol.* 2, 21–40. doi: 10.1146/annurev-cancerbio-030617-050134

Hnisz, D., Weintrau, A. S., Day, D. S., Valton, A. L., Bak, R. O., Li, C. H., et al. (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science (80-.).* 351, 1454–1458. doi: 10.1126/science.aad9024

Jin, F., Li, Y., Dixon, J. R., Selvaraj, S., Ye, Z., Lee, A. Y., et al. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* 503, 290–294. doi: 10.1038/nature12644

Kaul, A., Bhattacharyya, S., and Ay, F. (2020). Identifying statistically significant chromatin contacts from Hi-C data with FitHiC2. *Nat. Protoc.* 15, 991–1012. doi: 10.1038/s41596-019-0273-0

Ke, L., Zhou, H., Wang, C., Xiong, G., Xiang, Y., Ling, Y., et al. (2017). Super-enhancers promote transcriptional dysregulation in nasopharyngeal carcinoma. *Proc. Natl. Acad. Sci. U.S.A.* 77, 6614–6626. doi: 10.1158/0008-5472.CAN-17-1143

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., et al. (2002). The human genome browser at UCSC. *Genome Res.* 12, 996–1006. doi: 10.1101/gr.229102

Kloetgen, A., Thandapani, P., Ntziachristos, P., Ghebrechristos, Y., Nomikou, S., Lazaris, C., et al. (2020). Three-dimensional chromatin landscapes in T cell acute lymphoblastic leukemia. *Nat. Genet.* 52, 388–400. doi: 10.1038/s41588-020-0602-9

Kruse, K., Hug, C. B., and Vaquerizas, J. M. (2020). FAN-C: a feature-rich framework for the analysis and visualisation of chromosome conformation capture data. *Genome Biol.* 21:303. doi: 10.1186/S13059-020-02215-9

Kwiatkowski, N., Zhang, T., Rahl, P. B., Abraham, B. J., Reddy, J., Ficarro, S. B., et al. (2014). Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature* 511, 616–620. doi: 10.1038/NATURE13393

Kwok, A., Cheung, L., Lung, H. L., Mun, J., Ko, Y., and Cheng, Y. (2020). *Chromosome 14 Transfer And Functional Studies Identify A Candidate Tumor Suppressor Gene, Mirror Image Polydactyly 1, In Nasopharyngeal Carcinoma*. Available online at: www.pnas.org/cgi/content/full/ [accessed November 25, 2020].

Lajoie, B. R., Dekker, J., and Kaplan, N. (2015). The Hitchhiker's guide to Hi-C analysis: practical guidelines. *Methods* 72, 65–75. doi: 10.1016/j.ymeth.2014.10.031

Leong, M. M. L., Cheung, A. K. L., Kwok, T. C. T., and Lung, M. L. (2020). Functional characterization of a candidate tumor suppressor gene, Mirror Image Polydactyl 1, in nasopharyngeal carcinoma. *Int. J. Cancer* 146, 2891–2900. doi: 10.1002/ijc.32732

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv: Genomics* 1–3. arXiv:1303.3997,

Li, L., Barth, N. K. H., Pilarsky, C., and Taher, L. (2019). Cancer is associated with alterations in the three-dimensional organization of the genome. *Cancers (Basel)* 11:1886. doi: 10.3390/cancers11121886

Li, M., Cheung, A. K. L., Ko, J. M., Lung, H. L., Cheng, Y., and Dai, W. (2014). The interplay of host genetic factors and Epstein-Barr virus in the development of nasopharyngeal carcinoma. *Chin. J. Cancer* 33, 566–568. doi: 10.5732/cjc.014.10170

Li, R., Peng, C., Zhang, X., Wu, Y., Pan, S., and Xiao, Y. (2017). Roles of Arf6 in cancer cell invasion, metastasis and proliferation. *Life Sci.* 182, 80–84.

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293. doi: 10.1126/science.1181369

Lin, F., Zhang, J., and Liu, H. (2017). *Handbook of Practical Fine Needle Aspiration And Small Tissue Biopsies*. Berlin: Springer, 1–496. doi: 10.1007/978-3-319-57386-1

Liu, C., Yu, Z., Huang, S., Zhao, Q., Sun, Z., Fletcher, C., et al. (2019). Combined identification of three miRNAs in serum as effective diagnostic biomarkers for HNSCC. *EBioMedicine* 50, 135–143. doi: 10.1016/j.ebiom.2019.11.016

Lu, L., Liu, X., Huang, W. K., Giusti-Rodríguez, P., Cui, J., Zhang, S., et al. (2020). Robust Hi-C maps of enhancer-promoter interactions reveal the function of non-coding genome in neural development and diseases. *Mol. Cell* 79, 521–534.e15. doi: 10.1016/J.MOLCEL.2020.06.007

Lunardon, N., Menardi, G., and Torelli, N. (2014). ROSE: a package for binary imbalanced learning. *R J.* 6, 79–89. doi: 10.32614/RJ-2014-008

Mi, H., Ebert, D., Muruganujan, A., Mills, C., Albou, L. P., Mushayamaha, T., et al. (2021). PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* 49, D394–D403. doi: 10.1093/nar/gkaa1106

Nagano, T., Várnai, C., Schoenfelder, S., Javierre, B. M., Wingett, S. W., and Fraser, P. (2015). Comparison of Hi-C results using in-solution versus in-nucleus ligation. *Genome Biol.* 16:175. doi: 10.1186/s13059-015-0753-7

Ooi, W. F., Nargund, A. M., Lim, K. J., Zhang, S., Xing, M., Mandoli, A., et al. (2020). Integrated paired-end enhancer profiling and whole-genome sequencing reveals recurrent CCNE1 and IGF2 enhancer hijacking in primary gastric adenocarcinoma. *Gut* 69, 1039–1052. doi: 10.1136/gutjnl-2018-317612

Pott, S., and Lieb, J. D. (2015). What are super-enhancers? *Nat Genet.* 47, 8–12. doi: 10.1038/ng.3167

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680. doi: 10.1016/j.cell.2014.11.021

Ruan, L., Li, X. H., Wan, X. X., Yi, H., Li, C., Li, M. Y., et al. (2011). Analysis of EGFR signaling pathway in nasopharyngeal carcinoma cells by quantitative phosphoproteomics. *Proteome Sci.* 9:35. doi: 10.1186/1477-5956-9-35

Sampathi, S., Acharya, P., Zhao, Y., Wang, J., Stengel, K. R., Liu, Q., et al. (2019). The CDK7 inhibitor THZ1 alters RNA polymerase dynamics at the 5′ and 3′ ends of genes. *Nucleic Acids Res.* 47, 3921–3936. doi: 10.1093/nar/gkz127

Sauerwald, N., Singhal, A., and Kingsford, C. (2020). Analysis of the structural variability of topologically associated domains as revealed by Hi-C. *NAR Genomics Bioinforma.* 2, 1–13. doi: 10.1093/nargab/lqz008

See, Y. X., Wang, B. Z., and Fullwood, M. J. (2019). Chromatin Interactions and Regulatory Elements in Cancer: From Bench to Bedside. *Trends Genet.* 35, 145–158. doi: 10.1016/j.tig.2018.11.007

Schmitt, A. D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C. L., et al. (2016). A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* 17, 2042–2059. doi: 10.1016/j.celrep.2016.10.061

Tsang, C. M., Lui, V. W. Y., Bruce, J. P., Pugh, T. J., and Lo, K. W. (2020). Translational genomics of nasopharyngeal cancer. *Semin. Cancer Biol.* 61, 84–100. doi: 10.1016/j.semcancer.2019.09.006

Valton, A. L., and Dekker, J. (2016). TAD disruption as oncogenic driver. *Curr. Opin. Genet. Dev.* 36, 34–40. doi: 10.1016/j.gde.2016.03.008

van Berkum, N. L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L. A., et al. (2010). Hi-C: a method to study the three-dimensional architecture of genomes. *J. Vis. Exp.* 6, 1869. doi: 10.3791/1869

Wang, J., Su, L., Chen, X., Li, P., Cai, Q., Yu, B., et al. (2014). MALAT1 promotes cell proliferation in gastric cancer by recruiting SF2/ASF. *Biomed. Pharmacother.* 68, 557–564. doi: 10.1016/J.BIOPHA.2014.04.007

Wang, S., Lee, S., Chu, C., Jain, D., Kerpedjiev, P., Nelson, G. M., et al. (2020). HiNT: a computational method for detecting copy number variations and

translocations from Hi-C data. *Genome Biol.* 21:73. doi: 10.1186/S13059-020-01986-5

Wei, Q., Khan, I. K., Ding, Z., Yerneni, S., and Kihara, D. (2017). NaviGO: interactive tool for visualization and functional similarity and coherence analysis with gene ontology. *BMC Bioinforma.* 18:177. doi: 10.1186/S12859-017-1600-5

Yang, T., Zhang, F., Yardımci, G. G., Song, F., Hardison, R. C., Noble, W. S., et al. (2017). HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res.* 27, 1939–1949. doi: 10.1101/gr.220640.117

Yu, M. C., and Yuan, J. M. (2002). Epidemiology of nasopharyngeal carcinoma. *Semin. Cancer Biol.* 12, 421–429. doi: 10.1016/S1044579X02000858

Yu, Y., Dong, W., Li, X., Yu, E., Zhou, X., and Li, S. (2003). Significance of c-Myc and Bcl-2 protein expression in nasopharyngeal carcinoma. *Arch. Otolaryngol. Head Neck Surg.* 129, 1322–1326. doi: 10.1001/archotol.129.12.1322

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9:R137. doi: 10.1186/gb-2008-9-9-r137

Investigation of the Basic Steps in the Chromosome Conformation Capture Procedure

Oleg V. Bylino[1], Airat N. Ibragimov[1,2], Anna E. Pravednikova[1] and Yulii V. Shidlovskii[1,3]*

[1]Department of Gene Expression Regulation in Development, Institute of Gene Biology, Russian Academy of Sciences, Moscow, Russia, [2]Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Gene Biology, Russian Academy of Sciences, Moscow, Russia, [3]Department of Biology and General Genetics, I.M. Sechenov First Moscow State Medical University, Moscow, Russia

A constellation of chromosome conformation capture methods (C-methods) are an important tool for biochemical analysis of the spatial interactions between DNA regions that are separated in the primary sequence. All these methods are based on the long sequence of basic steps of treating cells, nuclei, chromatin, and finally DNA, thus representing a significant technical challenge. Here, we present an in-depth study of the basic steps in the chromatin conformation capture procedure (3C), which was performed using Drosophila Schneider 2 cells as a model. We investigated the steps of cell lysis, nuclei washing, nucleoplasm extraction, chromatin treatment with SDS/Triton X-100, restriction enzyme digestion, chromatin ligation, reversion of cross-links, DNA extraction, treatment of a 3C library with RNases, and purification of the 3C library. Several options were studied, and optimal conditions were found. Our work contributes to the understanding of the 3C basic steps and provides a useful guide to the 3C procedure.

Keywords: chromatin conformation capture, chromosome conformation capture, chromatin, distal interaction, DNA, Drosophila

INTRODUCTION

A chromatin conformation capture (3C) method is probably one of the most complex protocols in molecular biology, mainly due to its multistep nature (**Table 1**). The steps should be done in proper order and require careful execution. Incorrect implementation of the steps leads to poor restriction enzyme digestion, ineffective ligation, degradation, and/or loss of DNA. Importantly, a mistake made at any of the stages becomes known only at the very end of the procedure. Numerous controls therefore used, which are selected at different time points of protocol execution (Dekker, 2006).

The experimental literature on chromosome conformation capture methods (C-methods) is quite extensive. However, the literature lacks systematic analysis of how exactly the basic steps of the protocol work. At the same time, there are a lot of studies where individual selected steps were investigated (Gavrilov and Razin, 2009; Comet et al., 2011; Gavrilov et al., 2013b, 2015; Nagano et al., 2015b; Oksuz et al., 2020; Gridina et al., 2021). In this article, we try to fill this gap and explore in detail all the basic steps of the 3C protocol at once. A methodology for determining whether remote DNA regions can interact with each other in nuclei was first proposed by Cullen et al. (1993) and was called the nuclear ligation assay (NLA). In NLA,

Step	Characteristic of the step
1	Cell fixation, formaldehyde (FA) inactivation, and storage of nuclei
2	Cell lysis
3	Nucleoplasm release and chromatin treatment with heat (nuclei heating in the presence of SDS and Triton X-100/SDS sequestration with Triton X-100)
4	Digestion of DNA in nuclei
5	Ligation of DNA in nuclei
6	Reversion of cross-links and isolation of a 3C library
7	Treatment of the 3C library with RNases
8	Purification of the 3C library on magnetic beads

ligation of restriction endonuclease (RE)-digested chromatin was carried out in mammalian isolated intact nuclei using T4 DNA ligase and the frequency of ligation between regulatory elements was estimated by PCR (Cullen et al., 1993). In 2002, the methodology was finalized by Dekker et al. who supplemented the structure of the method with a fundamental step of cell fixation with formaldehyde (FA) and its subsequent quenching with glycine before chromatin digestion and ligation. The cross-linking of chromatin preserved the nuclear structure intact throughout the procedure without affecting its flexibility (Dekker et al., 2002). Data processing into genome-wide chromatin interaction maps was also proposed, suggesting that the 3C approach implemented on a yeast model can be applied to determine the spatial organization of whole genomes from bacteria to humans (Dekker et al., 2002). In 2009, a genome-wide 3C method was proposed and termed Hi-C. The method allows measuring the contact frequencies of all chromatin interactions that occur in the nucleus in a single experiment (Lieberman-Aiden et al., 2009). This area of research, called 3D genomics (Razin et al., 2019; Zhang and Li, 2020), is currently developing rapidly.
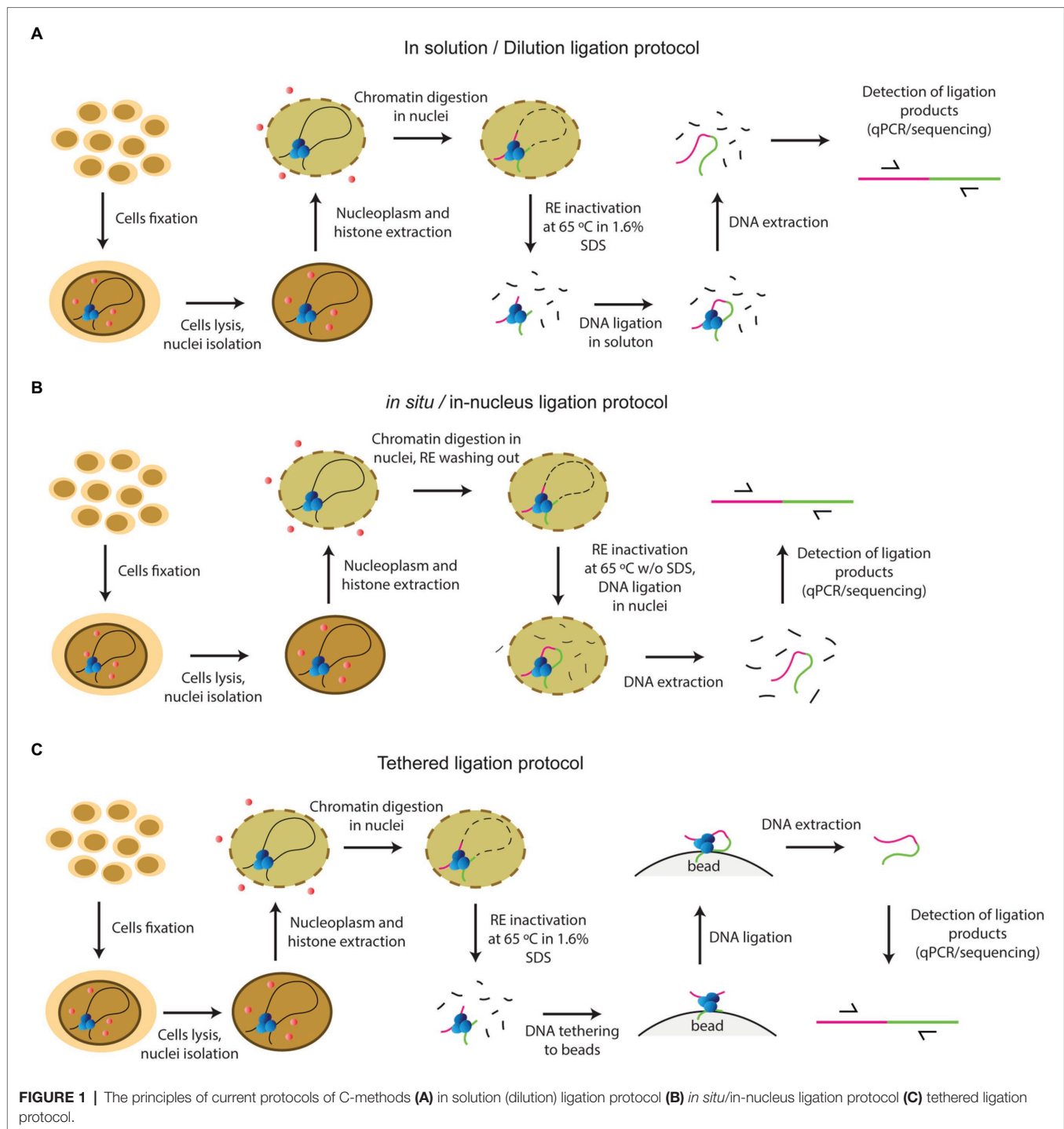
More recently, the basic principles of the method have been reassessed and an alternative strategy has been proposed, wherein a combination of two fixing agent (formaldehyde (FA) and disuccinimidyl glutarate (DSG)) and of two frequently cutting REs is used. This strategy, entitled Hi-C 3.0, should, in future, complement or even substitute the standard approach that is based on fixation with FA only and utilizes one frequently cutting RE (Oksuz et al., 2020).

At the same time, there remained a considerable lack of clarity regarding the steps of the classical version of the procedure. Since 2002, several add-ons and variations have been introduced into the basic 3C protocol by different working groups. Adding extra steps yielded a whole panel of the so-called C-methods, i.e., 4C, 5C, Hi-C, and so on, all making it possible to determine different aspects of the 3D genome organization in the cell nucleus (Zhang and Li, 2020). Various C-techniques have been standardized in the framework of the international 4D nucleome program (Dekker et al., 2017). Despite the phenomenal variety of existing C-methods, they all utilize the same basic steps, including fixation of cells with FA, cell lysis, nucleoplasm extraction, chromatin endonuclease digestion, ligation of the resulting DNA fragments, reversion of DNA-DNA

and DNA-protein links by heating, and subsequent isolation and analysis of contact frequencies between all or specific fragments. Initial stages of the development of the method since 2002 have employed an in-solution (dilution) ligation protocol; i.e., solubilized chromatin fragments were ligated in a highly diluted solution (Dekker et al., 2002; Tolhuis et al., 2002; **Figure 1A**). In 2012, a tethered Hi-C protocol was proposed (**Figure 1C**). The protocol also involved solubilization of chromatin fragments, like the dilution protocol. The difference was that cross-linked DNA fragments were ligated not in a large volume of solution, but on the surface of the solid phase of streptavidin-coated beads. Bound chromatin fragments were spatially separated from each other on the solid phase, which enhanced the signal-to-noise ratio as needed for detecting chromatin loops, and helped to avoid trans-ligation events, ensuring that ligations in these libraries are between DNA fragments that are cross-linked to each other (Kalhor et al., 2012; Gabdank et al., 2016). A Dounce homogenizer was used to isolate nuclei in this protocol even with cultured cells, and the library was processed with bovine RNase A (Kalhor et al., 2012; Gabdank et al., 2016).

In 2011, a key work in terms of the development of the C-methodology was carried out with a *Drosophila* model. Comet et al. (2011) showed for the first time that the same 3C profile was obtained with a complete non-fractionated sample and an insoluble chromatin fraction, while a 3C signal of a sufficient intensity was not detected with the supernatant fraction (Comet et al., 2011). The same result was demonstrated a couple of years later with mammalian cells, confirming the universality of this observation (Gavrilov et al., 2013b). The focus has changed since that time, and an *in situ* protocol came into use in 2014, involving cross-ligation of fragments within the nucleus (Rao et al., 2014; **Figure 1B**). This approach prevents random ligation between chromatin fragments released in solution by utilizing ligation *in situ*, i.e., in intact nuclei. The *in situ* protocol made it possible to obtain maps of very high quality and resolution under a sub-kilobase scale due to several distinctive features: a Dounce homogenizer was not used to obtain cell nuclei from cultured cells; a four-base pair RE was used instead of a six-base pair cutter; RE was inactivated by heating at 65°C for 20 min without adding 1.6% SDS, unlike in the dilution protocol; and ligation was performed with the nuclear fraction without large sample dilution (in ~1 ml). These circumstances reduced the frequency of non-specific contacts due to random ligation in dilute solution, – as was evidenced by a lower frequency of junctions between mitochondrial and nuclear DNAs and a higher frequency of random ligations observed when the supernatant fraction was sequenced (Rao et al., 2014). In addition, Rao's protocol did not include a processing of the library with RNase A.

In fact, proposals to avoid SDS treatment by inactivating the RE by heating at 65°C for 20 min and to use SDS only to inactivate heat insensitive enzymes have been expressed earlier (Splinter et al., 2012). However, the idea that ligation occurs mainly in the nuclear fraction and the structure of the nucleus should therefore be kept as intact as possible came only after the works of (Comet et al. (2011) and

FIGURE 1 | The principles of current protocols of C-methods **(A)** in solution (dilution) ligation protocol **(B)** *in situ*/in-nucleus ligation protocol **(C)** tethered ligation protocol.

Gavrilov et al. (2013b) and crystallized in Rao et al. (2014) as an *in situ* protocol.

In 2013, single-cell Hi-C was done for the first time (Nagano et al., 2013). In 2015, an in-nucleus ligation protocol was proposed in the development of this approach (**Figure 1B**). The in-nucleus ligation protocol is similar to the *in situ* protocol but utilizes even milder chromatin processing procedures. The step of RE inactivation by heating at 65°C

in the presence of SDS was completely eliminated from the protocol and ligation was thus performed in nuclei preserved to a maximal possible extent. It was found that the in-nucleus ligation protocol results in consistently lower levels of spurious ligation events manifested in trans-chromosomal contacts, reduces the experimental noise, and eliminates restriction fragment length bias found with the in-solution ligation protocol (Nagano et al., 2015a,b, 2017). Also, this protocol

does not employ a Dounce homogenizer to prepare a suspension of isolated nuclei.

In 2012, the Hi-C method was first applied to *Drosophila* cell culture (Hou et al., 2012), a model that we used in this study, and late embryos (Sexton et al., 2012). The two studies showed the principles of *Drosophila* genome organization into well-demarcated physical domains. In 2015, a work was published to demonstrate for the first time that a reconfiguration of topologically associated domains (TADs) occurs in response to cell stress was observed (Li et al., 2015). These early works all used the dilution protocol. The *in situ* protocol was employed in later works, which investigated the role of architectural proteins in enhancer-promoter interactions and TADs structure (Cubeñas-Potts et al., 2017; Rowley et al., 2017; Chathoth and Zabet, 2019). The *in situ* protocol was also used in single-nucleus Hi-C with cultured *Drosophila* cells (Ulianov et al., 2021). In 2017, a variant of the in-nucleus ligation protocol was applied to study the long-range interactions in *Drosophila* embryos (Stadler et al., 2017). During the experimental procedure, the RE was inactivated without using SDS or higher temperature by washing out from the nuclei, but embryos were homogenized using a Dounce homogenizer to prepare the nuclei (Stadler et al., 2017). Despite the general shift to the *in situ*/in-nucleus ligation protocol, the dilution protocol is still used in studies of the *Drosophila* genome organization (Ulianov et al., 2016, 2019; El-Sharnouby et al., 2017; Lo Sardo, 2021). The tethered ligation protocol was also successfully applied in two works with *Drosophila* (Eagen et al., 2015, 2017).

Thus, several protocols based on different principles of ligation are used in parallel today. The protocols differ in the conditions of basic steps: FA fixation and FA inactivation, cell lysis, the buffers used, the procedure of isolation and washing of nuclei, the severity of chromatin treatment, the conditions of chromatin restriction, RE inactivation, DNA ligation, and isolation and processing of the library with RNases.

Here we describe our study of the basic steps of the 3C procedure; offer our own options, which were found to be optimal in our hands; and provide the protocol suitable for S2 cells. We validate our results by showing with a model locus that efficient chromatin digestion and ligation occurs at an acceptable level, making it possible to distinguish the products of target ligation from the background of uncleaved DNA. Our procedure demonstrated its efficiency not only with cultured cells, but also with living objects, namely, – *Drosophila* larvae (Shidlovskii et al., 2021). Our work contributes to the understanding of the basic steps of the 3C procedure.

CELLS FIXATION, FA INACTIVATION, AND STORAGE OF NUCLEI

Several modes of cells fixation with FA and FA inactivation with glycine have been proposed in the literature for the 3C procedure. The amount of FA used for fixation varies from 1 to 3%, while the amount of glycine used for inactivation of FA can be classically 0.125 M (Dekker et al., 2002; Tolhuis et al., 2002), equimolar or in excess to FA (Comet et al., 2011; Sexton et al., 2012). In turn, the FA concentration is

influenced by the composition of the buffer in which fixation takes place. For example, the buffer used to isolate and fix the nuclei should not include Tris because Tris contains reactive amines, which cross-link FA to Tris, leaving less FA to fix DNA and proteins (Louwers et al., 2009), although original Dekker's protocol for yeast cells utilizes Tris. In the case of mammalian cells, it was originally proposed to add FA directly into DMEM supplemented with 10% FCS (Tolhuis et al., 2002), and DMEM contains many amino acids with reactive amines that bind FA, as Tris does. Therefore, cells were fixed in 1X PBS at room temperature (RT) in our experiments. For fixation, a 2X formulation of PBS was mixed in equal proportions with an aqueous solution of FA of a necessary concentration. Alternative fixation buffers (15 mM HEPES-KOH pH 7.6, 60 mM KCl, 15 mM NaCl, 4 mM MgCl$_2$, 0.1% Triton X-100, 0.5 mM DTT, and protease inhibitor cocktail) can be used as well (Comet et al., 2011; Sexton et al., 2012).

The first thing we observed was that DNA from cells fixed with 2% FA for 10 min was poorly isolated by phenol/chloroform (Ph/Chl) extraction immediately after fixation (FA was inactivated with equimolar glycine concentration; e.g., **Supplementary Figure S4D**, lane 16). Attempts at cross-links reversion of such DNA (overnight incubation at 50°C; **Supplementary Figure S4B**, lane 8) or at keeping such cells after fixation at −20°C overnight (**Supplementary Figure S4H**) unexpectedly aggravated the situation: DNA was not isolated by Ph/Chl extraction at all. Moreover, after keeping such cells at −20°C for 2 weeks, the DNA seemed to be fixed very firmly and was not isolated, even in much more severe isolation conditions using a protocol for FA fixed tissues (**Supplementary Figure S4I**; Campos and Gilbert, 2012). It was concluded that cells treated for 10 min with 2% FA were almost over-fixed and their DNA could not be isolated after further incubation or storage at −20°C. After fixation at other concentrations of FA (0.5–1.5% FA) for 10 min, DNA could be successfully isolated (FA was inactivated with equimolar glycine concentration; **Supplementary Figure S4B**, lanes 5–7; **Supplementary Figure S4F**, lanes 5, 6). Thus, we have established 1.5% as an upper limit for the concentration of FA at which DNA is not fixed too firmly in 10 min.

These results suggest that the storage of cells fixed with FA may be detrimental. We assume that cell fixation continues until the cells are completely frozen at −20°C, especially when glycine is used in a strong molar deficiency (0.125 M). We recommend to avoid storing the fixed material (or flash freezing using liquid nitrogen) and to immediately proceed to cell lysis after washing fixed cells.

The 0.125 M glycine concentration was taken for 3C protocols from ChIP protocols (Orlando et al., 1997) in which this concentration is the standard (Kim and Dekker, 2018a,b). As it was previously pointed out, quenching is not likely to be complete in the presence of 0.125 M glycine because glycine is not in excess over FA (Splinter et al., 2012). Over-fixation of the material, in this case, may therefore occur during freezing at −20°C or even thawing of cells. If inactivation was done with 0.125 M glycine to favor reproducibility between experiments, it was recommend immediately to proceed to the next step after quenching (Splinter et al., 2012).

Thus, we also do not recommend inactivating FA with glycine in a strong molar deficiency (0.125 M) and storing

nuclei inactivated in this way at −20°C. Instead, we recommend using glycine in an equimolar amount or in slight excess to FA (Comet et al., 2011; Sexton et al., 2012), keeping in mind two reactive groups of FA vs. one group of glycine. Thus, we used glycine at 666 (equimolar) and 800 (slight excess) mm for 1% of FA (**Supplementary Table S1**).

Empirically, we chose 10 min to fix cells properly. Shorter incubation times will result in lower detection signals of chromatin interactions, whereas longer incubation times will cause too many DNA-protein cross-links, resulting in a reduced digestion efficiency (van Berkum and Dekker, 2009). We observed that 25-min fixation leads to cell over-fixation. Over-fixed cells withstand digestion with Proteinase K (PrK) in the presence of 1% SDS, and DNA from such cells is not extracted with Ph/Chl.

We found that glycine used in a slight (about 20%) excess to FA does not change the PCR signal. A larger excess of glycine (for example, 800 mM glycine vs. 333 mM FA) leads to the formation of a sticky pellet that does not go down along the wall of the tube. The signal intensity of such samples in PCR is usually 3–4 times lower than in the case of equimolar inactivation of FA with glycine.

It was proposed to incubate cells on ice for 15 min after 5-min quenching with glycine at RT in order to stop the cross-linking completely (van Berkum and Dekker, 2009). The same processing step was included in the *in situ* protocol (Rao et al., 2014). We combined the stage of complete cessation of cross-links on ice with the stage of cell lysis in our procedure (see "Cell Lysis").

CELL LYSIS

The composition of the cell lysis buffer was the next important issue that we investigated. The 3C literature describes several fundamentally different compositions of lysis buffers. The two main buffers that are currently in use are a hypotonic buffer with a low concentration of a non-ionic detergent (10 mM Tris-HCl pH 8.0, 10 mM NaCl, and 0.2% NP-40), which is a classical buffer and was first introduced in the work (Tolhuis et al., 2002), and an isotonic buffer with a higher concentration of non-ionic detergents (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.5% NP-40, 1% Triton X-100, and 5 mM EDTA), which was first introduced in the works (Splinter et al., 2012; van de Werken et al., 2012). The first buffer inherited the detergent content from the pioneering work by Cullen et al. (1993). The second buffer was proposed as an alternative to the first one, contains an increased amount of non-ionic detergents, and is designed to ensure effective cell lysis and easy release of nuclei without the use of a Dounce homogenizer (Splinter et al., 2012). Several alternative lysis buffers have also been proposed in the literature. Examples include a hypotonic buffer that contains Mg^{2+}, is completely devoid of detergents (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 5 mM MgCl2, and 0.1 mM EGTA), and requires the use of a Dounce homogenizer (Hagège et al., 2007), and a medium salt buffer that contains Mg^{2+} and has a low content of non-ionic detergent (15 mM HEPES-KOH pH 7.6, 60 mM KCl, 15 mM NaCl, 4 mM $MgCl_2$, 0.1% Triton

X-100, and 0.5 mM DTT; Comet et al., 2011), in which the Na^+ and K^+ concentrations were taken from the work (Cullen et al., 1993).

After reviewing the variety of existing lysis buffers, we decided to test the contributions of individual components of the lysis buffer and composed 11 buffers, which differed in major component composition (**Table 2**).

Of these, buffers 1, 2, 3, 9, 10, and 11 appeared to provide the best results in terms of maintaining DNA integrity and were further tested under more variable chromatin treatment conditions. The results of a representative experiment are shown in **Figure 2**. The composition of the lysis buffer affects the integrity of the resulting DNA under different regimens of chromatin treatment (**Figures 2A–C**). When nuclei were subsequently treated with SDS at 65°C, lysis buffers with NP-40 were more preferable. DNA obtained after lysis in hypo- or isotonic conditions without NP-40 was poorer quality than after lysis in the same buffers, but with NP-40 (**Figure 2A**). When nuclei were subsequently processed with SDS at 37°C, a hypotonic lysis buffer with a maximum content of NP-40 was the only lysis buffer that stabilized DNA during lysis and ensured the absence of its degradation (**Figure 2B**).

Thus, a combination of hypotonic conditions of the classical buffer of Tolhuis et al. (2002) (10 mM Tris-HCl pH 8.0, 10 mM NaCl, and 0.2% NP-40) and the amount of detergents proposed by Splinter et al. (2012), van de Werken et al. (2012) (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, and 1% Triton X-100) makes it possible to preserve DNA integrity under different regimens of chromatin treatment. We additionally removed EDTA from the buffer since the presence of EDTA in the lysis buffer is incompatible with the presence of Mg^{2+} ions, while Mg^{2+} is required for maintaining the RNA structure and stabilizing chromatin (Louwers et al., 2009) and RNA is an integral architectural component of the nucleus, nuclear organelles, and heterochromatin (Caudron-Herger and Rippe, 2012; Hall and Lawrence, 2016; Ding et al., 2019; Michieletto and Gilbert, 2019; Thakur and Henikoff, 2020). The pH of the buffer was slightly shifted from 7.5 to 8.0 because pH 8.0 is most often used for DNA buffers. Thus, our lysis buffer has the following composition: 50 mM Tris-HCl, pH 8.0, 10 mM NaCl, 0.5% NP-40, and 1% Triton X-100. A similar buffer was used in a very delicate procedure to process mouse oocytes and zygotes in single-nucleus Hi-C (Flyamer et al., 2017). We believe that nuclei are more complete released from the cytoplasm remnants in the presence of a higher content of non-ionic detergents in the buffer. Also the buffer makes it possible to avoid the harsh impact of a Dounce homogenizer, ensuring the intactness of nuclei, while a low ionic strength creates hypotonic conditions in which extraction of the nucleoplasm is most efficient (Méndez and Stillman, 2000; Golov et al., 2015).

It has been reported that a cell lysate should not be viscous since its higher viscosity indicates insufficient cross-linking due to use of FA that is too old (van Berkum and Dekker, 2009). However, we never observed viscous lysates, even when cells were fixed with 1% FA that was stored at +4°C for several months before fixation. According to our observations, viscous

TABLE 2 | The lysis buffers studied.

Buffer	50 mM Tris-HCl, pH 8.0	10 mM Tris-HCl, pH 8.0	150 mM NaCl	10 mM NaCl	0.5% NP-40	0.2% NP-40	1% Triton X-100	5 mM MgCl$_2$	0.1 mM EGTA
1	+		+				+		
2	+		+			+	+		
3	+		+		+		+		
4	+		+				+	+	
5		+		+		+			
6		+		+					
7		+		+				+	
8		+		+				+	+
9	+			+			+		
10	+			+		+	+		
11	+			+	+		+		

lysates rather indicate destruction of nuclei and a release of DNA from them, and DNA always appeared to be degraded in such samples (not shown).

NUCLEOPLASM RELEASE AND CHROMATIN TREATMENT WITH HEAT

SDS and Triton X-100 treatment of nuclei fixed with FA were found to be necessary for any digestion of chromatin to occur (Splinter et al., 2004). The treatment removes the proteins that have not been cross-linked from DNA after fixation and partly denatures cross-linked proteins (Naumova et al., 2012). Accordingly, two fundamentally different regimens of chromatin processing have been proposed for the 3C procedure in the literature. The first regimen, at 37°C, was proposed in the original protocol by Dekker et al. (2002) and was developed in the work by Tolhuis et al. (2002), who increased the durations of consecutive chromatin treatments with SDS and then with Triton X-100 from 10 min up to 1 h each. Subsequently, as the dilution protocol in the work by Miele et al. (2006), the step of prolonged incubation of nuclei with SDS at 37°C was replaced with a step of short incubation with SDS at 65°C. This step was found to be essential for template generation (Miele et al., 2006) and dramatically increased DNA accessibility, by opening chromatin (Naumova et al., 2012), but, in fact, exactly how the effect of the step is achieved was not demonstrated in either of these two works. Actually, the step was properly studied in a field of plant biology and incubation at 65°C in the presence of SDS was shown to be necessary for inactivating endogenous nucleases (Louwers et al., 2009). The authors found that a nucleosome pattern was obtained when maize nuclei were permeabilized at 37°C with SDS regardless whether RE was added or not, indicating DNA degradation. However, DNA was still intact when isolated from an aliquot taken before the SDS incubation. The authors concluded that degradation occurred during incubation of nuclei with SDS at 37°C. Degradation was completely prevented when plant nuclei were first incubated at 65°C for 60 min before adding SDS (Louwers et al., 2009).

We also studied the effect of this step and found that heating chromatin at 65°C results in a some extent of DNA preservation during subsequent isolation (**Figure 2D**). This is supported by

predominant DNA degradation that was observed after processing chromatin with SDS at 37°C (**Figure 2B**) and eradicated by chromatin treatment at 65°C (**Figure 2A**). The results indicate that DNA degradation in S2 cells can also be due to the presence of endogenous nucleases, as it was previously described for plants (Louwers et al., 2009), while the enzymes are at least partly inactivated at 65°C.

As follows from **Figure 2A**, lysis under hypotonic conditions appears to be more efficient in terms of maintaining DNA integrity since the nucleoplasm, containing different nucleases with DNase I activity (Yang, 2011), is efficiently released in these conditions (Méndez and Stillman, 2000; Golov et al., 2015). We assume that a structural disruption of fixed nuclei may precede DNA degradation and is due to long-term treatment at 65°C in the presence of SDS. Therefore, after performing lysis in isotonic conditions, we titrated the duration of exposing nuclei to heat at 65°C and found that incubation of fixed nuclei for more than 40 min results in DNA degradation, presumably indicating disintegration of the nuclei and a release of nucleases (**Figure 2E**). We concluded that heating nuclei at 65°C should not exceed 40 min in case of S2 cells.

Thus, these data suggest that, on the one hand, heat treatment of fixed nuclei in the presence of SDS at 65°C partly reverses the cross-links and inactivates the remaining nucleases of a cytoplasmic origin and, in part, nuclear nucleases. On the other hand, the treatment leads to disintegration of the fixed nuclei and the release of residual nuclear nucleases, causing subsequent DNA degradation (see also the section Reversion of Cross-links and Isolation of the 3C Library section). From this point on, we used a hypotonic lysis buffer.

The effect of DNA degradation depending on the time of incubation at 65°C (**Figure 2E**) was unexpected but fits into the standard 3C protocol schemes. The time of chromatin processing at 65°C usually does not exceed 5–10 min in different types of 3C protocols (Miele et al., 2006; Kalhor et al., 2012; Rao et al., 2014). The results of the experiment (**Figure 2E**) imply that a brief chromatin treatment at 65°C already damage nuclei. This damage might partly release genomic DNA into solution. To study this issue, we analyzed 10 independent samples wherein chromatin was heated for 10 min at 65°C in the presence of 0.3% SDS. After heating, the samples were

FIGURE 2 | Cell lysis buffer composition and temperature regimens of chromatin treatment affect DNA integrity. **(A-C)** Effect of the lysis buffer composition on the integrity of the resulting DNA. 10 mg (~12 mln of cells) of S2 cells was fixed with 1% FA in 1X PBS at RT for 10 min, quenched equimolar with glycine for 5 min at RT, washed twice with 1X ice-cold PBS, and then resuspended in ice-cold lysis buffers #1, 2, 3, 9, 10, and 11 according to **Table 2**. Lane 1 – Buffer #1; Lane 2, #2; Lane 3, #3; Lane 4, #9; Lane 5, #10; and Lane 6, #11. Cells were incubated in different lysis buffers on ice for 15 min and were then washed once with ice-cold 1X PBS and resuspended in 1X RB for DpnII (50 mM Bis-Tris-HCl, 100 mM NaCl, 10 mM MgCl$_2$, 1 mM DTT, and pH 6.0 at 25°C). Then, SDS was added up to 0.1% **(A,B)** and the nuclei were treated with heat at 65°C for 10 min + 37°C for 50 min **(A)** or at 37°C for 1 h **(B)** and then Triton X-100 was added up to 1.8% **(A,B)** and the nuclei were incubated at 37°C for 1 h **(A,B)**. Thereafter, 500 μl of EB (50 mM Tris-HCl pH 8.0, 50 mM NaCl, and 1% SDS; Hug et al., 2017) containing 30 mM EDTA and 0.2 mg/ml PrK was added. The cross-links were reversed in EB at 65°C overnight (O/N) and the DNA was extracted with Ph/Chl and subsequently with Chl only, precipitated with 3 V of 96% ethanol, 0.3 M NaOAc pH 5.2, and 100 μg of glycogen, washed with 70% ethanol three times, dissolved in 25 μl of 10 mM Tris-HCl pH 8.0, and treated with 0.2–0.4 mg/ml of bovine RNase A for 30 min at RT. The amount of DNA subjected to electrophoresis was 1 μl. For **(C)**, the cells were treated in the same way as for **A and B**, but the step of chromatin heat treatment in the presence of SDS and Triton X-100 was omitted and instead, immediately after resuspension of nuclei in 1X RB, 500 μl of EB was added, supplemented with the same amount of SDS and Triton X-100 as for **A and B**. The reversion of cross-links and DNA extraction was performed in the same way. One of the three replicate experiments is shown. **(D)** Chromatin heat treatment at 65°C results in some DNA preservation during subsequent isolation. Cells were fixed, quenched, washed after quenching with 1X PBS as in **A-C**, and lysed in isotonic lysis buffer #3 (see **Table 2**; Tolhuis et al., 2002; lanes 1, 2) or in hypotonic lysis buffer #10 (Splinter et al., 2012; van de Werken et al., 2012; lanes 3, 4). Cells were incubated in lysis buffers on ice for 15 min and then nuclei were centrifuged, supernatant was discarded, nuclei were resuspended in 1X RB, SDS was added up to 0.3%, and nuclei were incubated at 65°C for 10 min (lanes 2, 4) or left untreated (lanes 1, 3). Thereafter, 500 μl of EB was added, the cross-links were reversed as in **A-C**, and the DNA was extracted using the GeneJET Genomic DNA purification kit (Thermo), according to the manufacturer's instructions. **(E)** DNA degradation upon chromatin treatment at 65°C. Cells were fixed, quenched, washed after quenching with 1X PBS as in **A-C**, and then lysed in isotonic lysis buffer #3 (see **Table 2**). Cells were incubated in lysis buffer on ice for 15 min and then nuclei were washed once with 1X ice-cold PBS, resuspended in 1X RB for DpnII, and then SDS was added up to 0.1%. Then, nuclei were treated with heat at 65°C for 10, 20, 30, 40, 50, 60, and 90 min in the presence of 0.1% SDS. After that, the nuclei were incubated for up to 1 h at 37°C for 50, 40, 30, 20, and 10 min (lanes 1–6, respectively). Then, Triton X-100 was added up to 1.8% and the nuclei were incubated at 37°C for 1 h. In the case of lane 7, after incubation in 0.1% SDS at 65°C for 90 min, Triton X-100 was added immediately up to 1.8% and nuclei were incubated at 37°C for 1 h. After incubation with SDS/Triton X-100 500 μl of EB was added to all samples, the cross-links were reversed and the DNA was extracted, precipitated, dissolved, and treated with bovine RNase A and subjected to electrophoresis as in **A-C**. One of the three replicate experiments is shown.

divided into the supernatant and pellet fractions and the supernatant fractions were analyzed. Release of the genomic DNA from nuclei into the supernatant (as well as a release of processed ribosomal RNA) was clearly detectable in 8 out

of 10 samples after nuclei were heated at 65°C for 10 min, apparently indicating damage to the fixed nuclei (**Figure 3A**).

To exclude that the nuclei were damaged at the previous stage of the procedure, we carried out the same experiment

FIGURE 3 | DNA and RNA release from nuclei upon their treatment with heat in the presence of SDS. **(A)** Chromatin treatment at 65°C causes DNA release from nuclei. Cells were fixed, quenched, and washed after quenching with 1X PBS as in **Figures 2A–C** and then lysed in hypotonic lysis buffer #11 (see **Table 1**), incubated on ice for 15 min, washed once with ice-cold 1X PBS, and resuspended in 1X RB for DpnII. Then, SDS was added up to 0.3% and the nuclei were treated with heat at 65°C for 10 min + 37°C for 50 min. After that, Triton X-100 was added up to 1.8% and the nuclei were incubated at 37°C for 1 h. Thereafter, the samples were separated into the supernatant (100 μl) and the pellet and 500 μl of EB was added to the supernatant fractions (lanes 1–10, 10 replicates of the experiment). Then, the cross-links were reversed and the DNA was extracted, precipitated, dissolved, and treated with bovine RNase A (lower panel) as in **Figures 2A–C** or left untreated (upper panel). The amount of the DNA subjected to electrophoresis was 2 μl. gDNA – genomic DNA, rRNA – processed high-molecular-weight RNA, and RNA – low-molecular-weight RNA (tRNA + degraded mRNA). **(B)** The content of nucleic acids in the supernatant after the stage of cell lysis. Cells were lysed on ice as in **A** and the samples were separated into the supernatant containing cytoplasm (100 μl) and the pellet fractions. EB was added to the supernatant fractions as in **A**. Then, the cross-links were reversed and the DNA was extracted, precipitated, dissolved, and treated with bovine RNase A (lower panel) or left untreated (upper panel) and subjected to electrophoresis as in **Figures 2A–C**. All other designations are as in **A**.

at the cell lysis stage. Analysis of the supernatant after cell lysis showed that only rRNA and low-molecular-weight RNA came out into solution during cell lysis and that there was no release of genomic DNA into solution (**Figure 3B**).
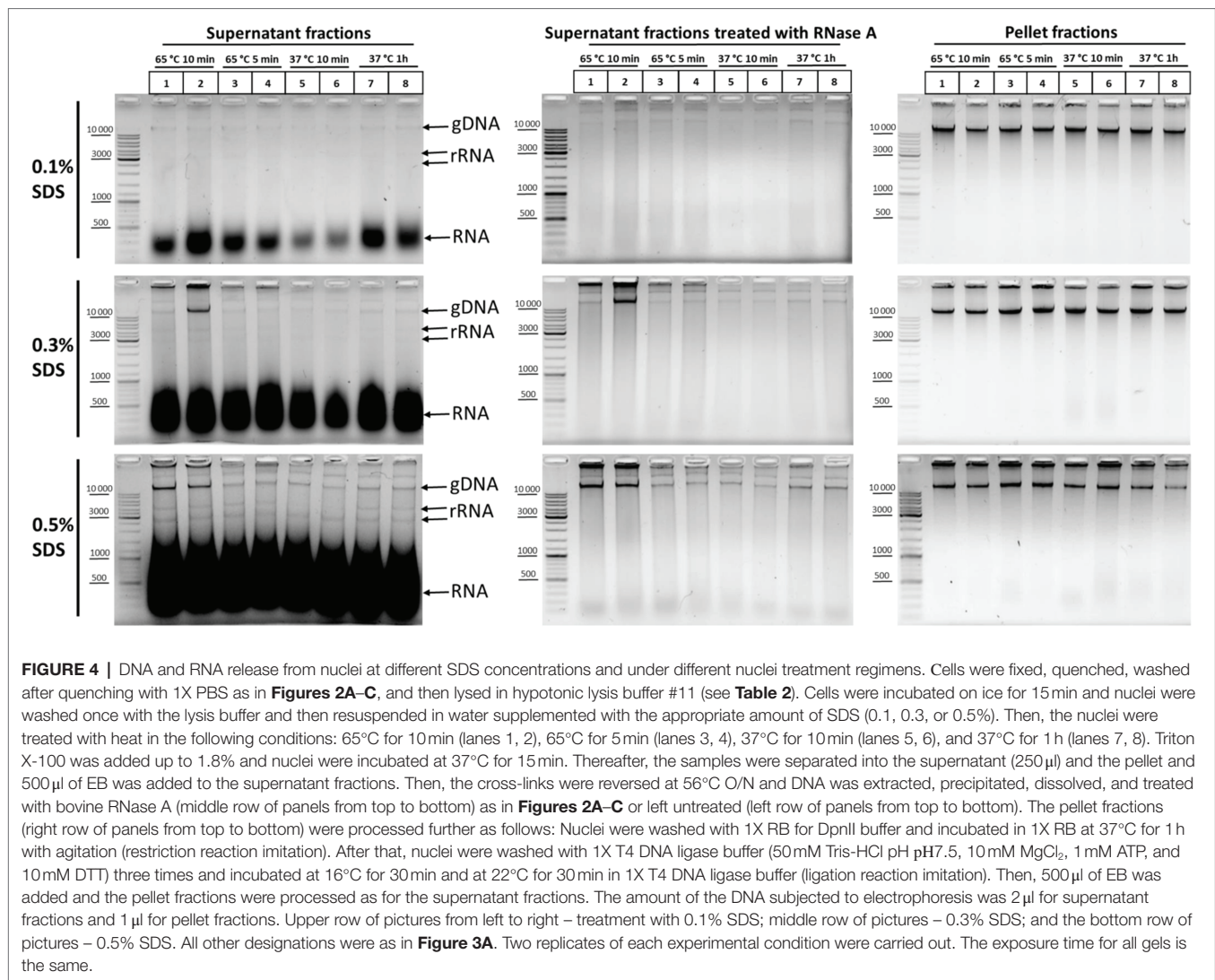
Thus, even a short incubation at 65°C apparently leads to damage to nuclear structures and a partial release of genomic DNA from nuclei into solution.

In order to understand which particular processing regimens of nuclei are the most harmful for nuclear integrity, we studied the release of DNA and rRNA into solution at different concentrations of SDS upon treatment of nuclei at 65°C for 10 and 5 min, as well as at 37°C for 10 min and 1 h. We found that, at 0.1% SDS, all processing regimens were apparently benign for nuclei and the release of DNA and RNA into solution was minimal (**Figure 4**, upper row of panels, from left to right). At 0.3% SDS, regimens with 65°C for 5 min, 37°C for 10 min, and 37°C for 1 h appeared to be equivalent in terms of maintaining nuclear integrity and provided minimal DNA release, but treatment at 65°C for 10 min provoked a noticeable DNA release in one replicate (**Figure 4**, middle row of panels, from left to right). Besides, a more pronounced release of low-molecular-weight RNA from nuclei was observed at 0.3% SDS compared with 0.1% SDS. At 0.5% SDS, treatment at 65°C for 10 min provoked a noticeable DNA release in both replicates and a more pronounced DNA release was observed

with all other regimens (**Figure 4**, bottom row of panels, from left to right). Besides, a huge amount of low-molecular-weight RNA and an appreciable amount high-molecular-weight RNA were released from nuclei into solution in the presence of 0.5% SDS.

We concluded that 0.3% is the maximum possible SDS concentration that does not cause nuclear damage. At this SDS concentration, it is possible to process nuclei at 65°C for 5 min to inactivate the residual nuclease activity without significantly compromising their integrity. Pronounced solubilization of histones from the nuclear fraction into the solution was also observed at this concentration (Gavrilov, 2016). However, the 0.1% SDS concentration appears to be the most sparing and makes it possible to preserve the maximum amount of nuclear RNA.

A scenario is also likely that the treatment of nuclei with heat and SDS does not cause partial damage to all nuclei, but rather a complete disintegration of some of the nuclei occurs. However, our observations testify against this scenario; i.e., we did not observe any decrease in the mass of the nuclear pellet even when nuclei were treated at 65°C for 10 min in the presence of 0.5% SDS. We noticed that the higher SDS concentration, the more transparent was the pellet of nuclear. When nuclei were treated with 0.1% SDS, the pellet of S2 nuclei was light gray. When nuclei were treated with 0.5%

FIGURE 4 | DNA and RNA release from nuclei at different SDS concentrations and under different nuclei treatment regimens. Cells were fixed, quenched, washed after quenching with 1X PBS as in **Figures 2A–C**, and lysed in hypotonic lysis buffer #11 (see **Table 2**). Cells were incubated on ice for 15 min and nuclei were washed once with the lysis buffer and then resuspended in water supplemented with the appropriate amount of SDS (0.1, 0.3, or 0.5%). Then, the nuclei were treated with heat in the following conditions: 65°C for 10 min (lanes 1, 2), 65°C for 5 min (lanes 3, 4), 37°C for 10 min (lanes 5, 6), and 37°C for 1 h (lanes 7, 8). Triton X-100 was added up to 1.8% and nuclei were incubated at 37°C for 15 min. Thereafter, the samples were separated into the supernatant (250 μl) and the pellet and 500 μl of EB was added to the supernatant fractions. Then, the cross-links were reversed at 56°C O/N and DNA was extracted, precipitated, dissolved, and treated with bovine RNase A (middle row of panels from top to bottom) as in **Figures 2A–C** or left untreated (left row of panels from top to bottom). The pellet fractions (right row of panels from top to bottom) were processed further as follows: Nuclei were washed with 1X RB for DpnII buffer and incubated in 1X RB at 37°C for 1 h with agitation (restriction reaction imitation). After that, nuclei were washed with 1X T4 DNA ligase buffer (50 mM Tris-HCl pH pH7.5, 10 mM MgCl₂, 1 mM ATP, and 10 mM DTT) three times and incubated at 16°C for 30 min and at 22°C for 30 min in 1X T4 DNA ligase buffer (ligation reaction imitation). Then, 500 μl of EB was added and the pellet fractions were processed as for the supernatant fractions. The amount of the DNA subjected to electrophoresis was 2 μl for supernatant fractions and 1 μl for pellet fractions. Upper row of pictures from left to right – treatment with 0.1% SDS; middle row of pictures – 0.3% SDS; and the bottom row of pictures – 0.5% SDS. All other designations were as in **Figure 3A**. Two replicates of each experimental condition were carried out. The exposure time for all gels is the same.

SDS, the pellet became "glassy" and difficult to work with at subsequent stages. A nuclear pellet obtained after treatment with 0.3% SDS had an intermediate transparency. We also noticed that cell nuclei acquire the same "glassy" appearance when the RE is inactivated after the restriction reaction in the presence of 1.3–1.6% SDS in the dilution protocol.

In addition, we found that not only high-molecular-weight DNA was detectable in solution after RNase treatment of the supernatant fractions, but also a significant amount of DNA molecules of different lengths (a DNA smear from more than 10,000 to 100 bp; **Figure 4**, middle row of panels, from top to bottom). The severity of chromatin treatment correlated with the intensity of the smear, which was minimal in the case of the regimen of 37°C for 10 min. The finding, together with the intactness of the high-molecular-weight DNA band, indicates that this fraction is a result of the processing of nuclei with SDS, rather than DNA degradation in the process of DNA isolation/cross-links reversion. We do not know about the nature of this DNA fraction, but it is possible that this

DNA may contribute to the elevated frequency of spurious contacts due to random ligation in dilute solution and might represent a source of experimental noise as described in Rao et al. (2014), Nagano et al. (2015b), Downes et al. (2021). Our data agree with the observations by Downes et al. (2021), who showed by separating the *in situ* 3C sample into intact nuclei and soluble DNA that ~25% of *in situ* 3C libraries come from disrupted nuclei.

Thus, taken together, the results suggest that the longer the incubation of nuclei at a higher temperature and the greater the SDS concentration used to extract nuclear proteins, the more genomic DNA and RNA passed into the supernatant fraction, indicating progressive damage to the nuclear structure with the increase in temperature and SDS concentration. Incubation at 37°C provides milder conditions of chromatin treatment. On the other hand, heat treatment of nuclei at 65°C is necessary to inactivate nucleases and may be useful for better DNA preservation at subsequent stages. However, the duration of chromatin treatment with heat, as well as the

SDS concentration, must be kept as low as possible to preserve the integrity of nuclei.

RESTRICTION OF DNA IN NUCLEI

In this part, we studied RE digestion of chromatin in nuclei. Protein complexes cross-linked to DNA may block restriction sites and reduce the efficiency of restriction digestion (Naumova et al., 2012). In turn, the efficiency of protein complex cross-linking can be influenced by the amount of FA used for fixation. Besides, the SDS/Triton X-100 ratio in the restriction reaction mixture, the duration of digestion and concentration of RE, and the conditions of chromatin treatment before chromatin digestion may exert an effect on the digestion efficiency. It was proposed to optimize the digestion efficiency by varying the SDS/Triton X-100 amounts before digestion, increasing Triton X-100 upon restriction digestion, and lowering the FA concentration (Hagège et al., 2007; Stadhouders et al., 2013). Since we selected the conditions for washing of nuclei from SDS/Triton X-100 (**Supplementary Figure S1C**), we therefore focused on the influence of chromatin treatment conditions and FA concentration on the restriction digestion efficiency.

First, to understand how the chromatin processing conditions affect the efficiency of DNA digestion in nuclei, we processed nuclei with SDS in two regimens, at 65 and 37°C, and then sequestrated SDS with Triton X-100. The DNA restriction pattern was not found to change depending on the temperature and duration of chromatin treatment (**Figures 5A–D**). Instead, unexpectedly, we observed more efficient ligation (denser and sharper at the top of the smear after DNA ligation) when chromatin was processed with SDS at 37°C (**Figure 5**, compare panels A and B with C and D).

Thus, more stringent conditions of chromatin treatment do not result in more efficient digestion of chromatin with RE but may instead influence the ligation efficiency.

Next, we investigated how the FA concentration affects the efficiency of chromatin digestion in nuclei. The formation of DNA fragments of higher molecular weights (MWs) is indicative of incomplete fragmentation of chromatin after its restriction digestion with a 4-bp cutter in nuclei (**Figures 5A–D**). This pattern corresponds to the pattern of DNA digested in solution with a 6-bp cutter (**Figure 5E**) rather than of DNA digested in solution with a 4-bp cutter (**Figure 5E**). Hence, we concluded that chromatin was not fully digested in nuclei. There are indications that the FA concentration may be directly related to the efficiency of chromatin digestion (Splinter et al., 2004; Dekker, 2007; Comet et al., 2011; van de Werken et al., 2012). The 1% FA concentration, which we used in experiments illustrated in **Figures 5A–D**, might be too high for efficient digestion. Therefore, we lowered the FA concentration to 0.5% and observed much more efficient digestion of cross-linked chromatin with DpnII (**Figure 5F**). The restriction pattern obtained at 0.5% FA was more similar to that observed after restriction in solution (**Figure 5E**) and did not shift down with a decrease in the FA concentration to 0.25% or even to 0.1% (**Figure 5F**).

Moreover, we noticed that the results related to chromatin accessibility to a RE may depend on the FA source. For example, a 1% FA solution prepared from PFA powder provides a higher fixation strength than a similar solution prepared from a 37% ready-to-use commercial solution. We estimate that 0.5% FA made from PFA and 1% FA made from a 37% commercial solution show comparable fixation efficiencies.

Other important issues are the duration of digestion and the RE concentration in the restriction reaction. The most common incubation time with a RE is 12–16 h (overnight incubation; Louwers et al., 2009; Nagano et al., 2013, 2015b, 2017; Flyamer et al., 2017). Digestion for 2–4 h was also suggested (Rao et al., 2014; Golov et al., 2020; Vermeulen et al., 2020; Ulianov et al., 2021). We determined that desired result is achieved within 3 h, although overnight incubation is convenient (**Figure 5F**). Regarding the RE concentration in the restriction reaction, a concentration of 2 U/μl is sufficient for efficient digestion of chromatin, in the case of using 10 mg of starting material as described in **Figures 2A–C**. The DpnII concentration we used was slightly higher than in recent works (0.66–1.66 U/μl; Golov et al., 2020; Vermeulen et al., 2020; Ulianov et al., 2021) and more similar to that used to digest yeast chromatin (2.07 U/μl; Schalbetter et al., 2019).

Thus, our data suggest that the efficiency of chromatin digestion depends mostly on the cross-linking agent concentration and is independent of the conditions of SDS/Triton X-100 chromatin treatment before digestion or the digestion time. Cells fixation with 0.5% FA for 10 min is sufficient for efficient digestion of chromatin with DpnII for 3 h at a concentration of RE of 2 U/μl.

LIGATION OF DNA IN NUCLEI

Before studying the peculiarities of ligation in nuclei, we checked how exactly T4 DNA ligase concentration affects the ligation efficiency. Activities of T4 DNA ligases were investigated in solution using genomic DNA cut with DpnII. The DNA ligation pattern in solution was not found to vary when T4 DNA ligase was used at 1, 5, or 10 U per reaction (**Figure 6A**). However, highly concentrated T4 DNA ligase (10 U/μl) is convenient to use since it prevents large amounts of glycerol from entering the reaction.

The pattern obtained after DNA digestion and ligation in solution (**Figure 5E**) differed from ligation patterns obtained after chromatin ligation in nuclei (**Figures 5A–D,F**). We assumed that DNA incubation at 16°C for 30 min followed by 22°C for 30 min (**Figures 5A–D,F**) may be insufficient for efficient ligation of cross-linked chromatin in nuclei. We therefore performed the experiments where non-cross-linked chromatin was ligated using the same incubation time (1 h) or cross-linked chromatin was ligated for incubation time extended from 1 h to overnight. It was observed that uncross-linked chromatin was more readily ligated in nuclei than cross-linked chromatin within 1 h (compare lane 3 in **Figures 5A–D,F** and lane 3 in **Figure 6B**). The prolongation of the ligation time for cross-linked chromatin from 1 h to overnight had a positive effect on the ligation pattern (made
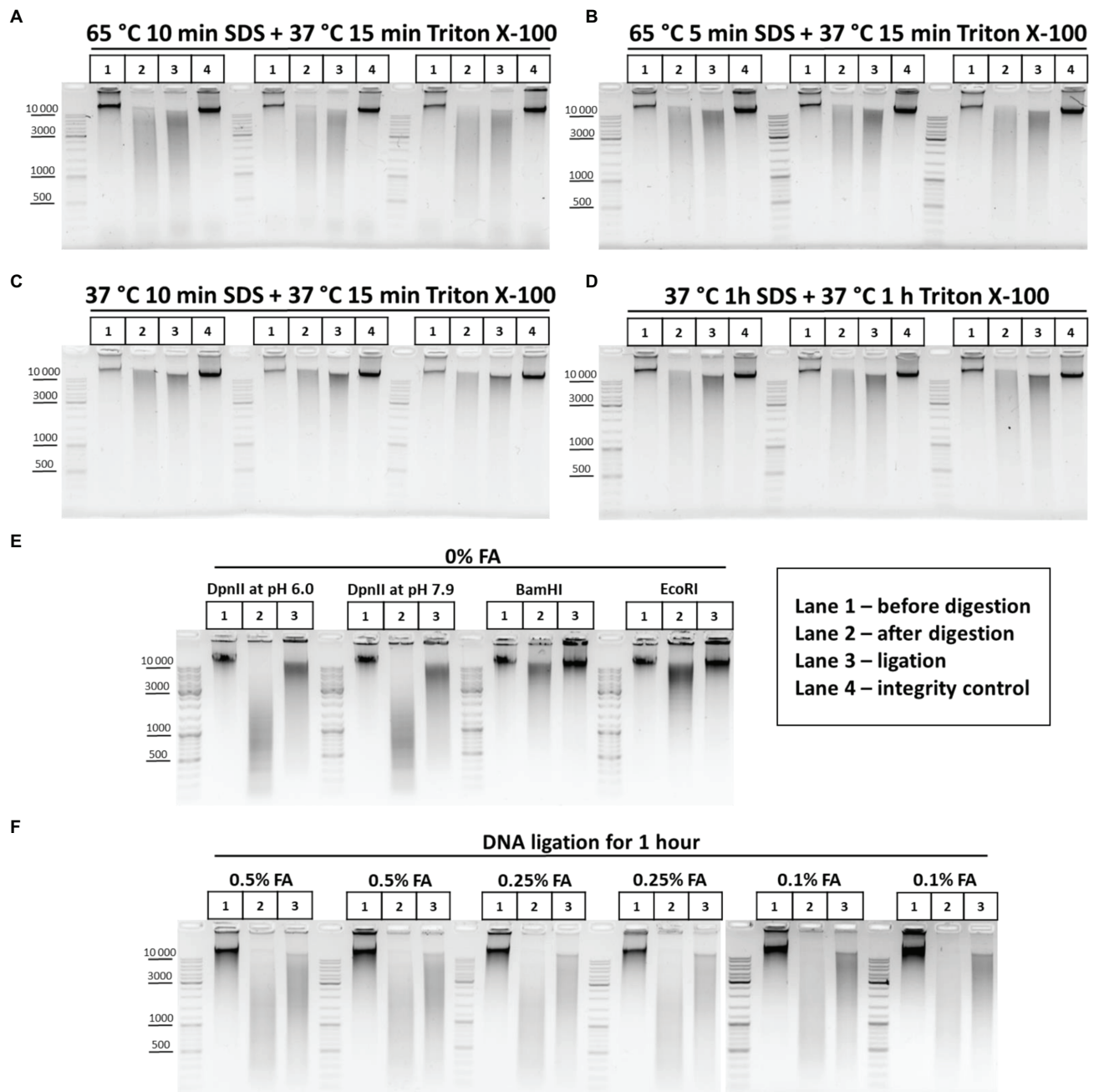
FIGURE 5 | Restriction digestion of DNA in nuclei and in solution. **(A–D)** Endonuclease digestion of the chromatin in nuclei after different regimens of chromatin treatment. Cells were fixed, quenched, washed after quenching with 1X PBS as in **Figures 2A–C** and then lysed, incubated in lysis buffer, and washed as in **Supplementary Figure S1A**. Then, nuclei were resuspended in autoclaved water supplemented with 0.3% SDS and were treated with the SDS and subsequently with Triton X-100 (1.8% final) in the following conditions, as indicated in the figure: (A) 65°C for 10 min with SDS + 37°C for 15 min with SDS/Triton X-100; (B) 65°C for 5 min with SDS + 37°C for 15 min with SDS/Triton X-100; (C) 37°C for 10 min with SDS + 37°C for 15 min with SDS/Triton X-100; and (D) 37°C for 1 h with SDS + 37°C for 1 h with SDS/Triton X-100. Then, nuclei were centrifuged, supernatant containing SDS/Triton-X100 was removed and nuclei were washed three times with 1X RB (DpnII buffer), and then resuspended in 1X RB and incubated with 2 U/μl of DpnII for 3 h at 37°C with agitation. Then, nuclei were washed three times with 1X T4 DNA ligase buffer, resuspended in 1X T4 DNA ligase buffer, and incubated with 0.25 U/μl of T4 DNA ligase (Sibenzyme) for 30 min at 16°C and for 30 min at 22°C with agitation. Aliquots of 1/10 of the sample volume were taken after chromatin treatment with SDS/Triton X-100, followed by washing of the nuclei with 1X RB (Lane 1 – undigested chromatin, chromatin integrity control before RE digestion, control #1) and, after restriction reaction, followed by washing with 1X T4 DNA ligase buffer (Lane 2 – digested chromatin, chromatin restriction control after RE digestion, control #2). The volume of samples was adjusted, with 1X T4 DNA ligase buffer, to 250 μl and PrK, SDS and EDTA were added, as in **Supplementary Figure S1A**, to the control #1, control #2, and ligation (lane 3) and control #3 (lane 4 – a control of DNA integrity throughout all stages of the procedure – sample that went through the entire procedure without addition of RE and T4 DNA ligase). Cross-links were reversed as in **Supplementary Figure S1A** and the DNA was extracted, precipitated as in **Figures 2A–C**, dissolved in 25 μl for control #1
(continued)

the shmear to shift up; compare lane 3 in **Figures 5A–D,F** and lane 3 in **Figure 6C**).

Thus, cross-linking apparently imposes certain spatial restrictions on the rate of chromatin ligation in nuclei and these can be overcome by a longer ligation duration.

Since SDS dramatically reduces the ligation efficiency (Louwers et al., 2009), chromatin ligation is usually performed in the 3C procedure after strong dilution of the restriction reaction mixture containing a high amount of SDS sequestered with Triton X-100. In this case, T4 DNA ligase is added to a buffer containing SDS diluted to 0.1% and sequestered with 1% of Triton X-100 (Dekker et al., 2002; Tolhuis et al., 2002; Lieberman-Aiden et al., 2009; Comet et al., 2011; Stadhouders et al., 2013; Vermeulen et al., 2020). However, as in the case of restriction digestion, the ligation reaction in the presence of 0.1% SDS and 1% Triton X-100 may be far from optimal. Therefore, we investigated the issue of whether the presence of 0.1% SDS and 1% Triton X-100 in the ligation reaction affects the efficiency of ligation in nuclei. We found that 0.1% SDS present in the reaction mixture alone, or surprisingly, in combination with 1% Triton X-100 (fresh stock and fresh working solution were prepared) negatively affects the ligation efficiency, whereas 1% Triton X-100 alone does not affect the ligation pattern (**Figure 6D**).

Thus, the washing of nuclei with the 1X T4 DNA ligase buffer after restriction digestion as proposed by Flyamer et al. (2017), Golov et al. (2020) helps to ensure efficient ligation by washing out SDS.

When the ligation reaction is carried out in the presence of 0.1% SDS and 1% Triton X-100, then complete sequestration of SDS by Triton X-100 is an important factor, since trace amounts of SDS will inhibit T4 DNA ligase. It was proposed to prepare a new Triton X-100 working solution every 1–2 months, since an old Triton X-100 solution has a notable negative effect on the digestion efficiency, probably due to inefficient sequestration of SDS with Triton X-100 decayed by light (Louwers et al., 2009). We did not observe any difference between a freshly prepared stock solution (20%) and an old one, which was kept protected from light at +4°C for at least 1 year (not shown). T4 DNA ligase worked inefficiently in nuclei in the presence of SDS regardless of whether the new
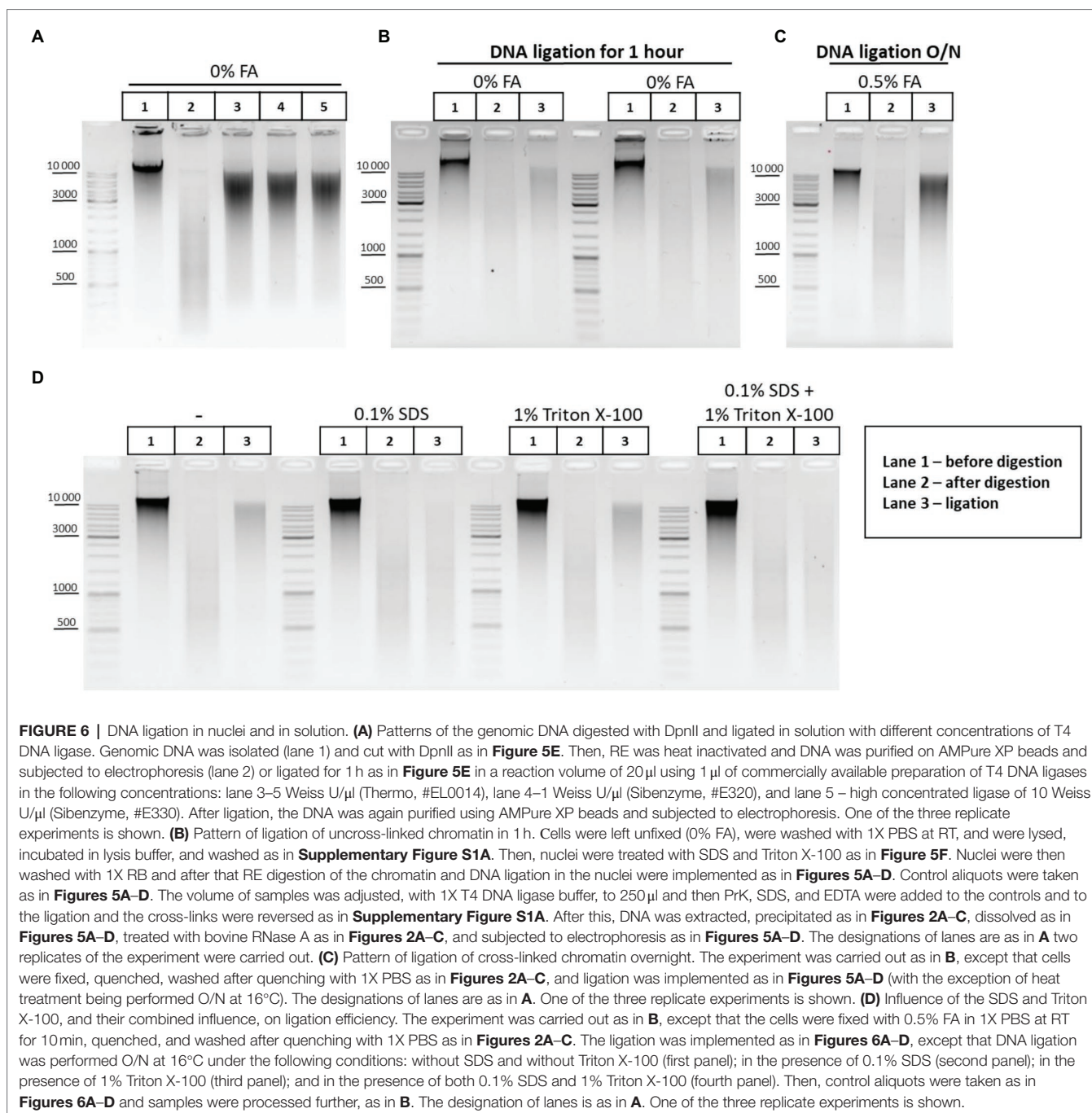
(**Figure 6D**) or old (not shown) Triton X-100 solution was used for sequestration.

It was shown that 0.1% SDS in combination with 1% Triton X-100 does not reduce the efficiency of DNA ligation in solution (plasmid DNA digested with a 6-bp cutter; Gavrilov et al., 2013a). However, 0.1% SDS in combination with 1% Triton X-100 does affect nuclear ligation (chromatin digested with a 4-bp cutter) according to our results. This effect may be due to the different times periods required for completing the reactions. Ligation of plasmid DNA fragments in solution takes place within just 10 min (Gavrilov et al., 2013a), whereas efficient ligation of fixed chromatin in nuclei requires at least several hours of incubation according to our results (compare **Figure 5F** lane 3 and **Figure 6C** lane 3). During this long incubation time, T4 DNA ligase is possibly inactivated by 0.1% SDS. Besides, 4-bp protruding ends are generally less efficiently ligated even in solution as it follows from the **Figure 6E**.

The efficiency of restriction digestion and ligation can be determined by performing a PCR spanning a specific genomic restriction site (Gavrilov, 2016). In order to quantitatively estimate the efficiency of ligation (regeneration of the DpnII site) in our conditions, the amount of the intact site (uncut and religated) was measured before and after ligation in the *RpII* locus by PCR-stop analysis (Comet et al., 2011). Loss of the amplicon signal after RE treatment is indicative of digestion efficiency (Belton et al., 2012). An increase in amplicon signal after ligation above the level of the uncut site indicates a ligation event, suggestion regeneration of the intact original restriction site (Gavrilov, 2016).

The experimental design of the system used to estimate the digestion and ligation efficiency in the *RpII* locus is shown in **Supplementary Figures S2A,B**. Before experiments, we validated the presence of a DpnII restriction site at the required position of the *RpII* locus by sequencing (**Supplementary Figure S2C**) and then optimized the Taq-man PCR conditions to achieve maximum sensitivity (**Supplementary Figure S2D; Supplementary Table S2**).

To estimate the digestion efficiency, one-third of the sample after overnight digestion (10–12 mln of starting nuclei) was recommended to take (Louwers et al., 2009), but our experience showed that up to half of the digested sample is required to

FIGURE 6 | DNA ligation in nuclei and in solution. **(A)** Patterns of the genomic DNA digested with DpnII and ligated in solution with different concentrations of T4 DNA ligase. Genomic DNA was isolated (lane 1) and cut with DpnII as in **Figure 5E**. Then, RE was heat inactivated and DNA was purified on AMPure XP beads and subjected to electrophoresis (lane 2) or ligated for 1 h as in **Figure 5E** in a reaction volume of 20 µl using 1 µl of commercially available preparation of T4 DNA ligases in the following concentrations: lane 3–5 Weiss U/µl (Thermo, #EL0014), lane 4–1 Weiss U/µl (Sibenzyme, #E320), and lane 5 – high concentrated ligase of 10 Weiss U/µl (Sibenzyme, #E330). After ligation, the DNA was again purified using AMPure XP beads and subjected to electrophoresis. One of the three replicate experiments is shown. **(B)** Pattern of ligation of uncross-linked chromatin in 1 h. Cells were left unfixed (0% FA), were washed with 1X PBS at RT, and were lysed, incubated in lysis buffer, and washed as in **Supplementary Figure S1A**. Then, nuclei were treated with SDS and Triton X-100 as in **Figure 5F**. Nuclei were then washed with 1X RB and after that RE digestion of the chromatin and DNA ligation in the nuclei were implemented as in **Figures 5A–D**. Control aliquots were taken as in **Figures 5A–D**. The volume of samples was adjusted, with 1X T4 DNA ligase buffer, to 250 µl and then PrK, SDS, and EDTA were added to the controls and to the ligation and the cross-links were reversed as in **Supplementary Figure S1A**. After this, DNA was extracted, precipitated as in **Figures 2A–C**, dissolved as in **Figures 5A–D**, treated with bovine RNase A as in **Figures 2A–C**, and subjected to electrophoresis as in **Figures 5A–D**. The designations of lanes are as in **A** two replicates of the experiment were carried out. **(C)** Pattern of ligation of cross-linked chromatin overnight. The experiment was carried out as in **B**, except that cells were fixed, quenched, washed after quenching with 1X PBS as in **Figures 2A–C**, and ligation was implemented as in **Figures 5A–D** (with the exception of heat treatment being performed O/N at 16°C). The designations of lanes are as in **A**. One of the three replicate experiments is shown. **(D)** Influence of the SDS and Triton X-100, and their combined influence, on ligation efficiency. The experiment was carried out as in **B**, except that the cells were fixed with 0.5% FA in 1X PBS at RT for 10 min, quenched, and washed after quenching with 1X PBS as in **Figures 2A–C**. The ligation was implemented as in **Figures 6A–D**, except that DNA ligation was performed O/N at 16°C under the following conditions: without SDS and without Triton X-100 (first panel); in the presence of 0.1% SDS (second panel); in the presence of 1% Triton X-100 (third panel); and in the presence of both 0.1% SDS and 1% Triton X-100 (fourth panel). Then, control aliquots were taken as in **Figures 6A–D** and samples were processed further, as in **B**. The designation of lanes is as in **A**. One of the three replicate experiments is shown.

use to reliably and conveniently measure the efficiency of digestion for 12 mln of starting cells as described in **Figures 2A–C**.

The first thing that we found was that fixed chromatin treated with SDS/Triton X-100 is much more readily digested than non-fixed chromatin treated with SDS/Triton X-100 in the same way (**Figure 7A**). We assume that cell fixation preserves the nuclear architecture and nuclear pores in particular and thereby contributes to a more complete release of histones and other nuclear proteins from the nuclei upon their SDS/Triton X-100 treatment, as it generally anticipated for the 3C procedure (Dekker et al., 2002) and was demonstrated in

Gavrilov (2016). A possible alternative explanation is that, without cross-linking, cells/nuclei are broken during SDS treatment (either at 37°C or 65°C), which leads to aggregation of the nuclei (DNA in aggregated nuclei is not digestible by REs). Thus, fixed chromatin appears to be more permissive to restriction digestion than non-fixed chromatin.

Second, we did not find any ligation yield after digestion and ligation of non-fixed chromatin; no regeneration of the DpnII site occurred in non-fixed chromatin against the background of uncut product (compare **Figures 7A,F,G**). The data are consistent with early observations by Dekker et al. (2002) and Gavrilov and Razin (2008), who noted that the
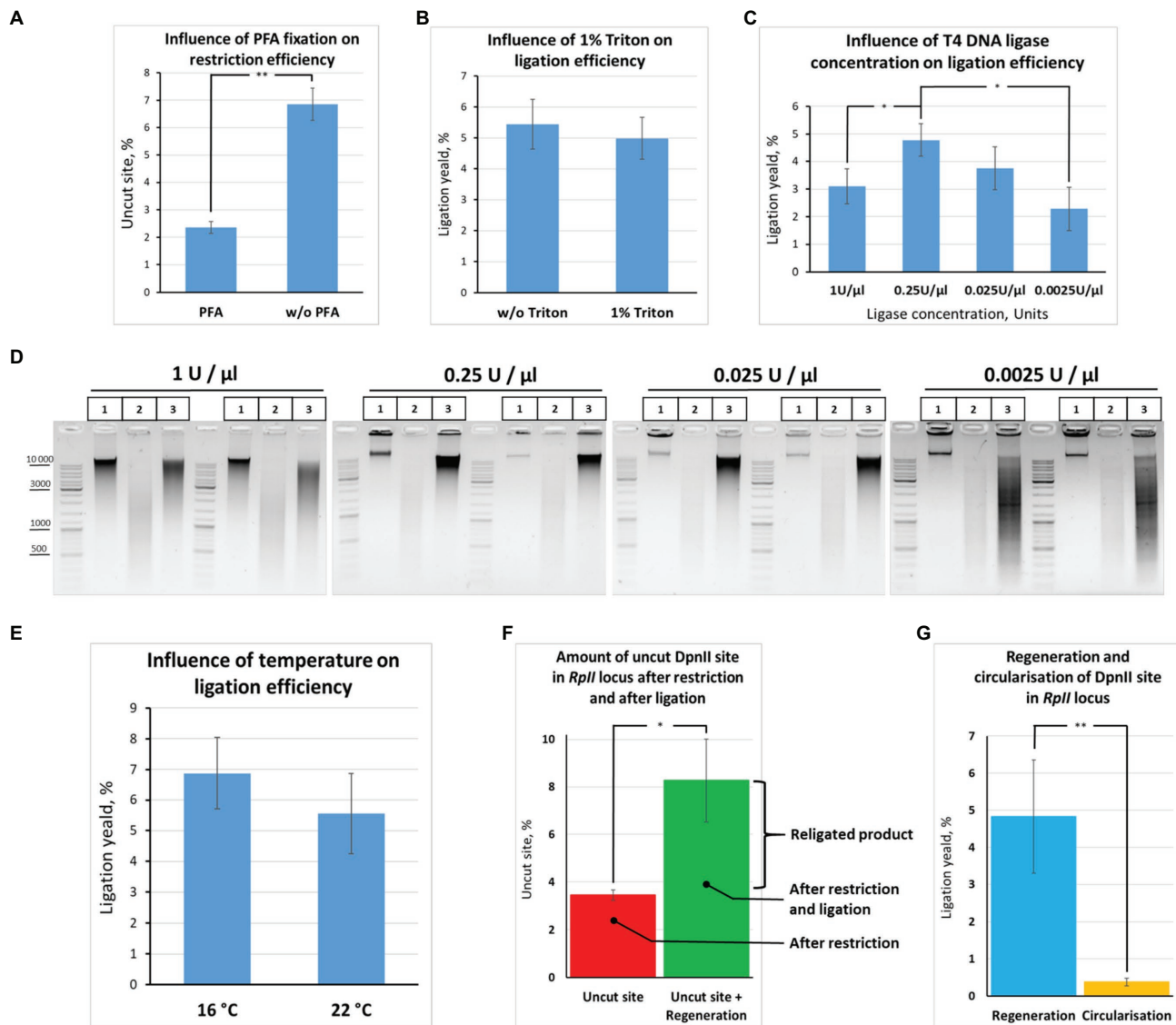
FIGURE 7 | Quantification of restriction digestion and ligation efficiency depending on reaction conditions. **(A)** Cell fixation enables efficient chromatin digestion. The experiment with unfixed cells was performed as in **Figure 6B**, except that ligation O/N was at 16°C. The experiment with fixed cells was carried out as in **Figure 6B**, except that cells were fixed with 0.5% FA, quenched, washed after quenching with 1X PBS as in **Figures 2A–C**, and ligation O/N was at 16°C. Control aliquots were taken as in **Figures 5A–D** and samples were processed further as in **Figure 6B**. The graph shows the frequency of intact (uncut) restriction site in the samples fixed with FA or non-fixed. The proportion of the uncut site was determined using PCR-stop analysis as described in **Figures 8A,B**. Error bars indicate SDs of four technical PCR measurements from at least three independent biological replicates of 3C library. Asterisks indicate significance levels: $^{**}p < 0.001$, $n = 22$. **(B)** Triton X-100 does not influence ligation efficiency. The experiment was carried out as in **(A)**. The ligation yield was calculated as described in **Supplementary Figures S2A,B**. No statistically significant difference between the groups was found ($p < 0.5$, $n = 20$). **(C,D)** High T4 DNA ligase concentration decreases ligation efficiency. The experiments were carried out as in **A**, except that ligation was performed at different T4 DNA ligase concentrations, as indicated, O/N at 16°C. The ligation yield was calculated as described in **Supplementary Figures S2A,B**. Error bars indicate SDs of four technical PCR measurements from at least three independent biological replicates of 3C library. Statistically significant differences between groups are indicated with asterisks: $^{*}p < 0.05$. The representative experiment presented under **D** was carried out in two replicates. The designation of lanes in **D** are as in **Figures 5A–D**. **(E)** Ligation efficiency does not significantly differ at 16°C and 22°C. The experiment was carried out as in **A**. The ligation yield was calculated as described in **Supplementary Figures S2A,B**. No statistically significant difference between the groups was found ($p < 0.25$, $n = 8$). **(F,G)** Quantitative values reflecting the effectiveness of the 3C procedure at the *RpII* site. The experiments were implemented as in **A**. The amount of intact (uncut or religated) DpnII site, the ligation yield, and the amount of circularized product were estimated as described in **Supplementary Figures S2A,B**. Asterisks indicate significance levels: for **F** $^{*}p < 0.05$, $n = 24$; for **G** $^{**}p < 0.001$, $n = 20$.

ligation yield is very low in non-fixed yeast and mammalian cells, respectively, not exceeding 5% of the amount of ligation in fixed cells. The data agree with later observations by

Belaghzal et al. (2017) that leaving out the cross-linking step leads to dramatic loss of detected contacts and the inability to reconstruct the chromatin conformation beyond a few kb.

TABLE 3 | Amounts of uncut site and the ligation yield in different type of protocols.

	Protocol used	Amount of uncut site, %	Ligation yield, %	Material	FA conc., %	RE	Conditions of chromatin treatment, °C
This study	In-nucleus ligation protocol	3.5	5	Drosophila cells	0.5	DpnII	65
Gavrilov, 2016	Dilution protocol	20	9	Mouse cells	2	MboI	37
van de Werken et al., 2012	*In situ* ligation protocol	20	ND*	Human cells	2	DpnII	37

*ND, Not determined

Third, we observed that the presence of 1% Triton X-100 in the ligation reaction did not exert a significant effect on the ligation efficiency (**Figure 7B**). The finding was consistent with electrophoresis data (**Figure 6D**). Thus, Triton X-100 added up to 1% does not inhibit ligation in nuclei. Moreover, we observed that the ligation efficiency may be improved in the presence of Triton X-100 in the ligation reaction since some amount of Triton X-100 prevents adhesion of nuclei to the tube walls. To prevent adhesion of nuclei, 0.1% Triton X-100 is sufficient.

Further, we quantified how the T4 DNA ligase concentration and the temperature affect the efficiency of DNA ligation in nuclei. We found that ligation was most efficient at T4 DNA ligase concentrations ranging from 0.25 to 0.025 U/μl. We did not find significant differences in ligation efficiency between the two concentrations, while lowering the ligase concentration to 0.0025 U/μl led to a significant decrease in ligation efficiency (**Figures 7C,D**). Unexpectedly, increasing the T4 DNA ligase concentration in the reaction from 0.25 to 1 U/μl decreased the ligation efficiency (**Figures 7C,D**).

We assume that an increase in T4 DNA ligase concentration leads to rapid consumption of the ATP pool and that this might be critical during the long incubation time of the reaction. Thus, the reaction of chromatin ligation in nuclei should be neither overloaded with T4 DNA ligase nor lacking it. Our results are in good agreement with the data of other authors. For example, the final concentrations of T4 DNA ligase were 0.0012, 0.001, and 0.006 Weiss U/μl in the articles (Lieberman-Aiden et al., 2009; Naumova et al., 2013; Falk et al., 2019), respectively, where the Dekker team's protocol was used. The concentration of T4 DNA ligase was also not high in other classical works and protocols. For example, the concentrations were 0.006, 0.008, and 0.014 Weiss U/μl in (Stadhouders et al., 2013; Rao et al., 2014; Nagano et al., 2015b), the concentrations were 0.006, 0.008 and 0.014 Weiss U/μl, respectively, and 0.013 Weiss U/μl in Splinter et al. (2012), van de Werken et al. (2012) and 0.02 and 0.024 in Comet et al. (2011) for 3C and H3C, respectively (Cavalli team's protocol). TThe range of the most efficient T4 DNA ligase concentrations (0.25-0.025 U/μl) that we found for S2 cells is more consistent with the T4 DNA ligase concentration described for *Drosophila* tissues (from adults, pupae, or embryos; Comet et al., 2011). As for the temperature, we found no significant differences in ligation efficiency between 16 and (**Figure 7E**).

Next, to quantify the ligation efficiency, we measured the amounts of the uncut site and ligated product (target ligation,

regeneration of the original restriction site, or ligation yield). The amounts of the uncut site and religated product were estimated at 3.4 ± 0.2 and $8.3 \pm 1.7\%$, respectively (**Figure 7F**). The average range of variation between technical replicates was only $\pm 0.4\%$ after determining the amount of the uncut site, while it was $\pm 1.5\%$ after determining the ligation yield, the two values differing significantly ($p < 0.05$, $N = 20$). Moreover, there was no significant correlation between the amount of the uncut site and the ligation yield either in the presence of 1% Triton X-100 ($r = 0.49$, $p < 0.15$, $N = 10$) or in the absence of Triton X-100 ($r = 0.333$, $p < 0.35$, $N = 10$).
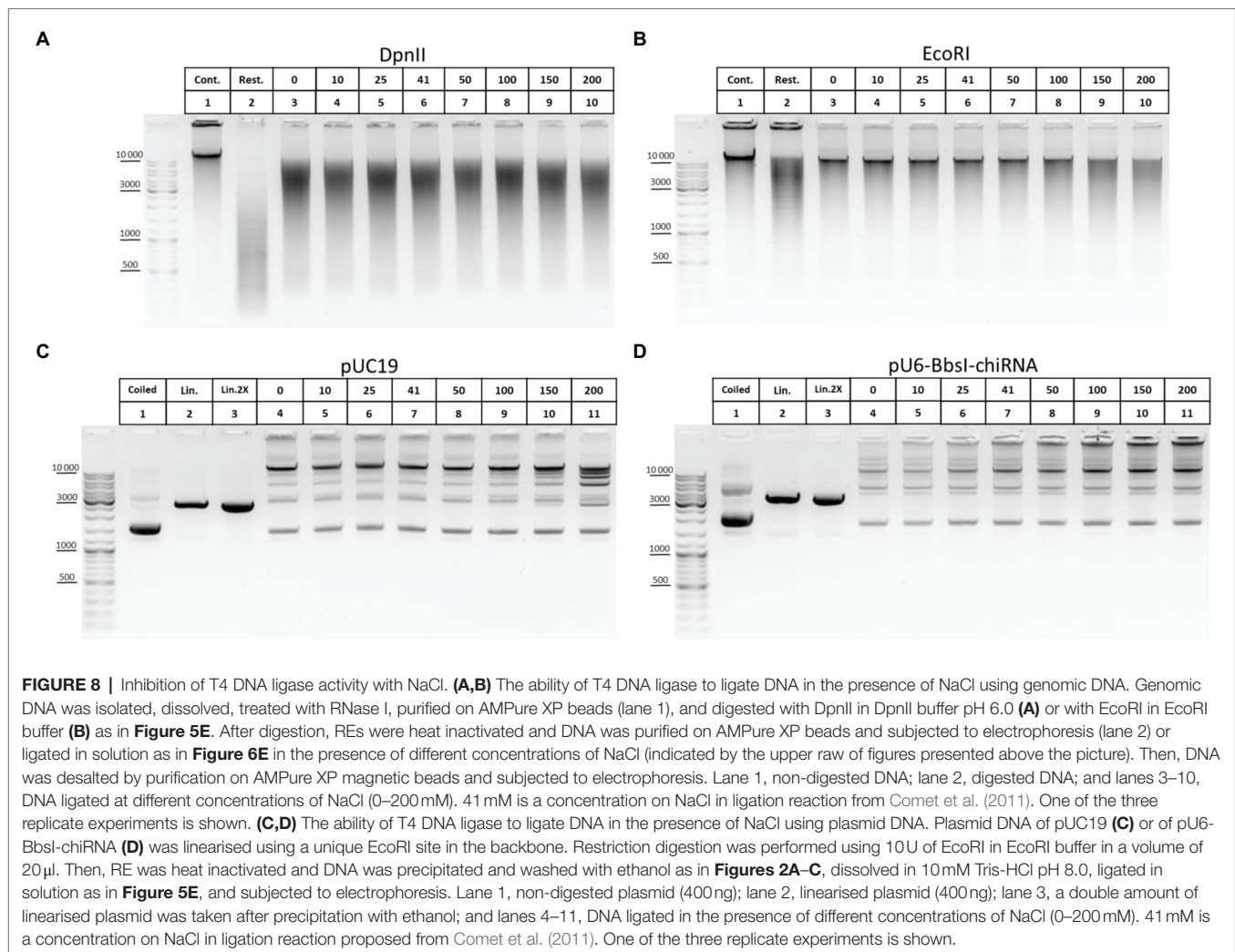
We concluded that our experimental conditions make it possible to achieve efficient chromatin digestion with DpnII and to detect the ligation products at an acceptable level, above the background of non-cleaved DNA. At the same time, a larger data variation observed after ligation than after digestion is possibly a reflection of the fact that ligation of target DNA ends is a rare event in the 3C procedure (Gavrilov et al., 2013a; Gavrilov, 2016), thus, requiring prolong incubation in the case of fixed nuclei (**Figures 5A–D,F, Figures 6B,C**) and being therefore statistically more variable according to our results.

The efficiency of DNA digestion should be as high as possible, preferably higher than 80% (Louwers et al., 2009; Naumova et al., 2012; van de Werken et al., 2012) and may vary between samples and cell types used (Splinter et al., 2012).

Several works were performed to measure the frequencies of restriction site regeneration (ligation yield) and the percentage of uncut site copies (van de Werken et al., 2012; Gavrilov, 2016). We compared our results with the results from these works (**Table 3**).

The percent of cut site in our conditions (96.6%) was much higher than achieved with the dilution (Gavrilov, 2016) and with *in situ* (van de Werken et al., 2012) protocols. This might be explained by the lower FA concentration used, the different source of experimental material, and different conditions of chromatin treatment in our study compared with the above works.

At the same time, the ligation yield in our study (~5%) was lower than with the dilution protocol (Gavrilov, 2016; **Table 3; Figure 7G**). This can be partly explained by using of different RE (DpnII instead of MboI) or is more likely be a consequence of different conditions of chromatin treatment [65°C in our study vs. 37°C in Gavrilov (2016)] since it seems that chromatin treatment at 37°C might give a higher ligation yield, as the **Figures 5A–D** imply.

FIGURE 8 | Inhibition of T4 DNA ligase activity with NaCl. **(A,B)** The ability of T4 DNA ligase to ligate DNA in the presence of NaCl using genomic DNA. Genomic DNA was isolated, dissolved, treated with RNase I, purified on AMPure XP beads (lane 1), and digested with DpnII in DpnII buffer pH 6.0 **(A)** or with EcoRI in EcoRI buffer **(B)** as in **Figure 5E**. After digestion, REs were heat inactivated and DNA was purified on AMPure XP beads and subjected to electrophoresis (lane 2) or ligated in solution as in **Figure 6E** in the presence of different concentrations of NaCl (indicated by the upper raw of figures presented above the picture). Then, DNA was desalted by purification on AMPure XP magnetic beads and subjected to electrophoresis. Lane 1, non-digested DNA; lane 2, digested DNA; and lanes 3–10, DNA ligated at different concentrations of NaCl (0–200 mM). 41 mM is a concentration on NaCl in ligation reaction from Comet et al. (2011). One of the three replicate experiments is shown. **(C,D)** The ability of T4 DNA ligase to ligate DNA in the presence of NaCl using plasmid DNA. Plasmid DNA of pUC19 **(C)** or of pU6-BbsI-chiRNA **(D)** was linearised using a unique EcoRI site in the backbone. Restriction digestion was performed using 10 U of EcoRI in EcoRI buffer in a volume of 20 μl. Then, RE was heat inactivated and DNA was precipitated and washed with ethanol as in **Figures 2A–C**, dissolved in 10 mM Tris-HCl pH 8.0, ligated in solution as in **Figure 5E**, and subjected to electrophoresis. Lane 1, non-digested plasmid (400 ng); lane 2, linearised plasmid (400 ng); lane 3, a double amount of linearised plasmid was taken after precipitation with ethanol; and lanes 4–11, DNA ligated in the presence of different concentrations of NaCl (0–200 mM). 41 mM is a concentration on NaCl in ligation reaction proposed from Comet et al. (2011). One of the three replicate experiments is shown.

In summary, the ratios between the amounts of non-cleaved and religated restriction site in our study were approximately the same as previously shown for mouse embryonic liver cells (Gavrilov, 2016), emphasizing the universality of our observations.

We additionally estimated the amount of a circular ligation product (284 bp; **Figure 7A**). The amount was found to be more than 10 times lower than the amount of the regenerated ligation product (0.37% ± 0.10% vs. 5%; **Figure 7G**). Thus, ligation of adjacent restriction fragments is predominantly observed with our in-nucleus ligation protocol, whereas circularization and ligation of adjacent restriction fragments are equally possible for solubilized chromatin in the dilution protocol (Gavrilov, 2016). Our observations coincide with the observations of other authors (Arkadiy Golov and Maxim Imakaev, personal communications).

Thus, the absence of solubilization contributes to directional ligation of adjacent chromatin fragments.

The ability of T4 DNA ligase to ligate DNA in the presence of NaCl was another important issue that we investigated concerning ligation. In some 3C protocols,

a salt-free 1X T4 DNA ligase buffer is used to dilute the restriction reaction mixture after RE inactivation at 65°C (Comet et al., 2011; Rao et al., 2014; Vermeulen et al., 2020). However, 1X restriction buffer (RB) for DpnII contains 100 mM NaCl, which has a potential inhibitory effect because T4 DNA ligase is salt-sensitive and is most active in a salt-free buffer (Raae et al., 1975; Hayashi et al., 1985). The effect of T4 DNA ligase inhibition with NaCl is manifested differently in ligation of DNA cut with different REs (Hayashi et al., 1985) and was not studied for DpnII. Unexpectedly, no significant inhibition of T4 DNA ligase was observed with up to 150–200 mM NaCl (**Figure 8A**). However, the activity of T4 DNA ligase was inhibited at a much lower NaCl concentration in the above studies.

We assumed that the effect of salt inhibition of T4 DNA ligase activity would be more clearly detected if genomic DNA is digested with a 6-bp cutter since after ligation DNA assembles in solution into fragments of higher MWs, larger than 10–12 kb (**Figure 5E**). Again, noticeable inhibition of ligation was observed sonly at NaCl concentrations of 150–200 mM (**Figure 8B**).

With plasmid DNA, such inhibition was not observed at all or was very slight at 150–200 mM NaCl (**Figures 8C,D**).

We concluded that T4 DNA ligase can function efficiently in the presence of NaCl used at up to 100 mM and that even 150–200 mM NaCl is possible to use in the ligation reaction without significantly affecting the ligase activity.

The following questions are also discussed in the supplement: SDS sequestration with Triton X-100; washings of nuclei; optimization of Taq-man PCR conditions; reversion of cross-links and isolation of the 3C library (discussion of the role of EDTA, ionic strength, and dilution in maintaining DNA integrity in the 3C procedure and the role of the temperature and composition of the extraction buffer (EB) in maintaining DNA integrity in the 3C procedure); treatment of the 3C library with RNases; and purification of the 3C library on magnetic beads.

A detailed protocol of 3C library preparation consisting of the optimized steps is given in the Supplement.

DISCUSSION

In general, the 3C procedure requires optimization for each specific cell type. We provide a useful framework for optimization of the protocol and carry it out for *Drosophila* S2 cells.

A sequence of steps and features was combined into the 3C protocol to allow keeping the nuclei as intact as possible. The features include a lysis buffer that ensures hypotonic conditions (Tolhuis et al., 2002) and has the detergent concentrations as proposed by Splinter et al. (2012), van de Werken et al. (2012), a washing of the nuclei with a hypotonic lysis buffer after lysis in hypotonic conditions (Rao et al., 2014; the washing can also be done with 1X PBS, 1X RB, or even with water), a washing of the nuclei with 1X RB to remove SDS/Triton X-100 (Flyamer et al., 2017), and a washing of the nuclei with a 1X T4 DNA ligase buffer to remove the RE instead of heating the nuclei at 65°C to inactivate RE (Flyamer et al., 2017; Golov et al., 2020). The efficiency of these steps is evidenced, for example, from the absence of DNA release from nuclei upon their treatment at 37°C and a more distinct ligation pattern (**Figures 5C–D**).

We studied in detail the basic stages of the 3C protocol in this work, paying special attention to the preservation of DNA integrity throughout the procedure. The absence of DNA degradation at all stages makes the method as reliable as possible and results in reproducible profiles at an output.

The causes of DNA degradation are the most mysterious aspect of the method. We and others noticed that DNA degradation occurs early in the 3C protocol, often at the cell lysis step, and is commonly attributed to contaminating nucleases (Louwers et al., 2009; Naumova et al., 2012). Thus, initial stages of the protocol appear to be absolutely critical for maintaining DNA integrity throughout the procedure. According to our results, cell lysis and subsequent chromatin treatment with SDS/Triton X-100 are the most crucial steps in this regard. There are a lot of enzymes with DNase activity in eukaryotic

cells, and some of them function in the nucleus, cytosol, and lysosomes (Yang, 2011; Kawane et al., 2014; Fujiwara et al., 2017). The isolated cytosol of eukaryotic cells exhibits divalent cation-dependent DNase activity (Lechardeur et al., 1999). We assumed that residual DNase activity may exert a detrimental effect on the integrity of a 3C DNA library during cell lysis and treatment of nuclei.

We found that treatment of cells with a hypotonic lysis buffer which provides for a more efficient release of nucleoplasmic proteins in comparison with an isotonic buffer (Méndez and Stillman, 2000; Golov et al., 2015) and a subsequent thorough washing of the nuclei ensure DNA integrity at subsequent stages. Nucleases may be released from fixed nuclei *via* diffusion through the disrupted nuclear envelope as fixed nuclei swell in hypotonic conditions in the 3C procedure (Gavrilov et al., 2013b).

It is conceivable that two groups of nucleases, cytoplasmic and nuclear, may be responsible for DNA degradation in the 3C procedure. Predominantly, cytoplasmic nucleases are likely to be inactivated and washed off during cell lysis, a washing of nuclei, and chromatin treatment at 65°C in isotonic conditions. The assumption is supported by the observations that a prolonged heating of nuclei at 65°C after isotonic lysis and a washing of nuclei were accompanied by DNA degradation, indicating nuclear damage and, probably, a nucleoplasm release. A transition to hypotonic conditions during reversion of cross-links in the 1X T4 DNA ligase buffer induces a release of nucleoplasmic nucleases, and high EDTA concentrations in EB are required for preventing DNA degradation. However, it is unclear how these nuclear nucleases avoid solubilization with SDS at the step of SDS/Triton X-100 treatment of nuclei at 65°C/37°C.

At the same time, some amount of genomic DNA was released into the supernatant fraction after hypotonic lysis, a washing of nuclei, and short-term heating at 65°C. The more severe the processing conditions, the greater amount of DNA was found in the supernatant. However, this release was not accompanied by DNA degradation in the supernatant and nuclear pellet fractions. The finding proved that hypotonic lysis is more efficient.

If nucleases are suspected to occur in the sample, as in the case of isotonic lysis or an improper washing of nuclei after lysis, dilution of the restriction reaction mixture with a large volume of a hypotonic 1X T4 DNA ligation buffer prior to DNA ligation may provide a means to preserve the integrity of the 3C library upon ligation and subsequent reversion of cross-links. This approach is unconsciously used in the dilution ligation protocol (Dekker et al., 2002; Tolhuis et al., 2002; Lieberman-Aiden et al., 2009; Comet et al., 2011; Stadhouders et al., 2013; Ulianov et al., 2016; Vermeulen et al., 2020). In other cases, the addition of EDTA at up to 30 mM can help to preserve DNA integrity upon reversion of cross-links.

In experiments with *Drosophila* S2 cells, we found that the addition of EDTA above a certain threshold concentration (>22 mM) prevents DNA degradation and maintains DNA integrity. The finding further confirms that DNA degradation

is directed by nucleases depending on divalent cations and, primarily, Mg^{2+} as the most abundant divalent cation in the cell. Our results from experiments on preserving EDTA concentrations correlate well with the literature data on physiological concentrations of Mg^{2+} in cells and hemolymph of *Drosophila* larvae (Begg and Cruickshank, 1962; Larrivee, 1979; van der Meer and Jaffe, 1983; Stewart et al., 1994; Echalier et al., 2018), as well as in the extracellular and perivitelline fluids (~20 mM; van der Meer and Jaffe, 1983).

Several other important findings made in our experiments are listed below.

Quality of the resulting 3C libraries may depend on how and with what RNase the library is treated. In our experiments, *E. coli* RNase I performed best. The enzyme did not degrade DNA at room temperature and 37°C even when used at high concentrations. In contrast, bovine RNase A, which is most frequently used in such experiments, often degraded the library at 37°C, but not at room temperature. An increase in bovine RNase A concentration led to complete DNA degradation at both room temperature and 37°C.

Quality of the resulting libraries may depend on the conditions of elution from magnetic beads at the step of additional purification of the library. When DNA was eluted with water at an elevated temperature of 55°C (heating facilitates elution), DNA degradation occurred, while DNA was stabilized in a Tris buffer or in the presence of NaCl.

An increase in SDS concentration during extraction of uncross-linked proteins from nuclei increases the amount of uncross-linked RNA released from the nuclei. The observation is important for studying chromatin-associated RNAs using the Hi-C method (Li et al., 2017; Sridhar et al., 2017; Bell et al., 2018; Yan et al., 2019; Bonetti et al., 2020; Gavrilov et al., 2020). In this method, SDS is routinely used to remove the cytoplasm and to extract the proteins and RNAs that have not been cross-linked to DNA before ligation of chromatin-associated RNAs to DNA. Our data suggest that, when RNA molecules are not cross-linked to DNA, but can participate in the formation of ligation products with DNA, harsh treatment with a high SDS concentration will decrease RNA-DNA cross-linking and cause a loss of signal.

We observed that non-fixed chromatin ligates much faster than fixed chromatin and that ligation in nuclei is strongly inhibited with 0.1% SDS even after its sequestration with 1% Triton X-100. Thus, it is necessary for efficient ligation to wash off SDS/Triton X-100 as suggested by Flyamer et al. (2017). We also noticed that DNA ligation in nuclei is slightly more efficient at 16°C in the presence of T4 DNA ligase at a concentration not exceeding 0.25 U/µl.

We additionally showed that the circularization of a restriction fragment is a very rare event under the conditions of the *in situ*/in-nucleus protocol compared to direct ligation of adjacent DNA restriction fragments. We measured the ligation frequencies for only one site; however, it was previously shown that values measured for several sites do not fundamentally differ (Gavrilov and Razin, 2008).

No ligation yield was observed after digestion and ligation of non-fixed chromatin (compare absolute values in **Figures 7A,F,G**). At the same time, we showed that non-fixed chromatin is ligated much faster than fixed chromatin (compare **Figure 5F** and **Figures 6B,C**). This discrepancy can be explained by the fact that mass ligation of DNA fragments still proceeds in non-fixed nuclei, although specific interactions of restriction fragments are lost in non-fixed chromatin (Dekker et al., 2002; Gavrilov and Razin, 2008; Belaghzal et al., 2017). In addition, non-fixed chromatin is less frequently cleaved (**Figure 7A**), thus producing DNA fragments of higher MWs. The fragments are ligated in a pattern that mimics efficient ligation (compare the patterns obtained with 6-bp and 4-bp cutters in **Figure 5E**). As a result, the library from non-fixed cells looks as having been efficiently ligated but contains very few specific ligation products (Dekker et al., 2002; Gavrilov and Razin, 2008; Belaghzal et al., 2017). Thus, non-fixed chromatin is cut worse and ligated faster, but this ligation turns out to be meaningless, while fixed chromatin is cut more efficiently and ligated slower, but this ligation makes sense.

The conditions that we selected for S2 cells are also suitable for *Drosophila* larvae [(Shidlovskii et al., 2021); Bylino et al., manuscript in preparation]. We suppose that the same conditions might be suitable for working with more complex models than *Drosophila* tissues, including mammalian and human cells and mammalian tissues. However, additional steps of tissue processing with collagenase, elastase, or trypsin should be introduced into the procedure to obtain a cell suspension

TABLE 4 | Conditions found to be optimal for stages of the 3C protocol.

#	Stage of the protocol	Conditions studied	Optimal
1	Inactivation of FA with glycine	1. Inactivation with 0.125 M glycine	Inactivation with equimolar or slightly excessive glycine
		2. Inactivation with equimolar or slightly excessive glycine	
	Storage of fixed nuclei	1. After using 0.125 M glycine	Move immediately to the lysis stage without storage of nuclei
		2. After using equimolar or slightly excessive glycine	
2	Lysis	Contributions of various components of the lysis buffer	Hypotonic conditions (10 mM NaCl) with high amounts of non-ionic detergents (1% Triton X-100 + 0.5% NP-40)
3	Chromatin heating in the presence of SDS (nucleoplasm release)	(1) 65°C for 10 min, (2) 65°C for 5 min, (3) 37°C for 10 min, (4) 37°C for 1 h at different SDS concentrations (0.1, 0.3, 0.5%)	65°C for 5 min or 37°C for 10 min with 0.1–0.3% SDS

(Continued)

TABLE 4 | Continued

#	Stage of the protocol	Conditions studied	Optimal
4	Sequestration of SDS with Triton X-100	The following chromatin treatment regimens were studied (SDS/Triton X-100 concentrations were 0.1 and 1.8%, respectively): (1) 65°C for 10 min with SDS in 1X RB + 37°C for 1 h with Triton X-100 in 1X RB, (2) 65°C for 10 min with SDS in 1X RB + addition of Triton X-100 without any incubation, (3) 65°C for 10 min with SDS in water + 37°C for 15 min with Triton X-100 in water, (4) 65°C for 10 min with SDS in 1X RB + 37°C for 15 min with Triton X-100 in 1X RB	65°C, 10 min, SDS in water + 37°C, 15 min, Triton X-100 in water
5	Washing of nuclei after cell lysis	(1) With hypotonic lysis buffer containing 1% Triton X-100 and 0.5% NP-40, (2) With 1X RB, (3) With 1X PBS, (4) With water	With hypotonic lysis buffer (1X RB, 1X PBS and water can also be used)
	Washing of nuclei after treatment with SDS/Triton X-100	(1) Without washing (addition of 10X RB into water containing SDS/Triton X-100), (2) Without washing (centrifugation, removing the supernatant, and resuspension of nuclei in 1X RB), (3) Washing three times with 1X RB, (4) Washing once with 1X PBS and resuspension in 1X RB	Washing three times with 1X RB
	Washing of nuclei after restriction digestion with or without prior RE inactivation at 65°C for 20 min	(1) Without washing [dilution of restriction mixture with 1X T4 DNA ligase buffer (1X T4 LB)], (2) Washing once with 1X LB, (3) Without washing (centrifugation, removing the supernatant, and resuspension of nuclei in 1X T4 LB), (4) Washing three times with 1X T4 LB	Without RE inactivation at 65°C for 20 min + washing three times with 1X T4 LB
6	Chromatin digestion in nuclei	1. The following regimens after different heat treatment were studied (treatment with SDS/Triton X-100 was done in water; SDS and Triton X-100 concentration was 0.3 and 1.8%, respectively; cell fixation was carried out at 1% FA): (1) 65°C for 10 min with SDS + 37°C for 15 min with Triton X-100, (2) 65°C for 5 min with SDS + 37°C for 15 min with Triton X-100, (3) 37°C for 10 min with SDS + 37°C for 15 min with Triton X-100, (4) 37°C for 1 h with SDS + 37°C for 1 h with Triton X-100. 2. Was studied after cell fixation with different FA concentrations (chromatin treatment was done at 65°C for 5 min with 0.3% SDS in water + 37°C for 15 min with 1.8% Triton X-100 in water): 0.5, 0.25, 0.1, 0%	1. 37°C for 1 h with SDS + 37°C for 1 h with Triton X-100 or 65°C for 5 min with SDS + 37°C for 15 min with Triton X-100 2. 0.5% FA
7	DNA ligation in nuclei	1. Duration of ligation: 1 h, overnight 2. Presence of 0.1% SDS and 1% Triton X-100 3. Effect of 1% Triton X-100 4. Effect of T4 DNA ligase concentration: 1, 0.25, 0.025, 0.0025 U/μl 5. Effect of the reaction temperature: 16°C, RT (22°C) 6. Effect of NaCl concentration: 10–200 mM	1. Overnight 2. Exclude 0.1% SDS sequestered with 1% Triton X-100 from ligation reaction 3. Does not affect the ligation process; at 0.1% prevents nuclei from sticking to tube walls upon shaking 4. Range 0.25–0.025 U/μl. 5. 16°C 6. Ligation is most effective at up to 100 mM NaCl and can be carried out without significant loss of T4 DNA ligase activity at up to 150–200 mM NaCl
8	Reversion of cross-links and isolation of 3C library	1. Effect of EDTA concentration in EB on DNA integrity 2. Effect of diluting restriction mixture with 1X T4 LB on DNA integrity 3. Influence of temperature on efficiency of cross-link reversion 4. Effect of ionic strength and EB composition on cross-link reversion and the success of DNA extraction with Ph/Chl	1. 30 mM EDTA is required to maintain DNA integrity after cell lysis in isotonic conditions, but not after cell lysis in hypotonic conditions 2. Dilution helps to maintain DNA integrity after lysis in isotonic conditions, but is not required after lysis in hypotonic conditions 3. Reversion of cross-links is efficient at 56°C for 13.5 h or 50°C for 19 h 4. Hypertonic conditions are not recommended. In isotonic conditions, adding 1% SDS or 30 mM EDTA is optional. In hypotonic conditions, adding 1% SDS or 30 mM EDTA is required
9	Treatment of 3C library with RNases	Bovine RNase A, Recombinant RNase A, RNase If, RNase T1, RNase I	RNase I
10	Purification of 3C library on magnetic beads (SPRI)	1. Elution from beads with: Water at 55°C and at RT 2. 10 or 50 mM Tris-HCl pH 8.0, 10 or 25 mM NaCl	1. Elution with water at RT 2. Elution with 10–50 mM Tris-HCl pH 8.0 or 10–25 mM NaCl at RT or at 55°C
11	DNA storage	On magnetic beads: overnight at RT and at −20°C under 75% ethanol On ice: over a weekend, for a week, for 2 weeks	Any condition is suitable

in this case. After isolation, cells can be treated using the steps shown and discussed here. We hope that the sequence of steps proposed here may be useful for single-cell methods since several steps were taken from the work by Flyamer et al. (2017), which focused on single cells. We summarized our results in **Table 4**.

CONCLUSION

We characterized the critical points of the 3C procedure and offer options to bypass these bottlenecks. Improvements introduced to the procedure make it possible to carry out the 3C method with the maximum yield, to preserve DNA integrity at all stages, and to increase the stability and reproducibility of the method.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at European Nucleotide Archive, accession no: ERR6474674.

AUTHOR CONTRIBUTIONS

YS and OB: conceptualization and methodology. OB: formal analysis and investigation. OB, AP, and YS: writing. OB and AI: visualization. YS: supervision, project administration,

and funding acquisition. All authors contributed to manuscript revision, read, and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.733937/full#supplementary-material

REFERENCES

Begg, M., and Cruickshank, W. J. (1962). XVI—A partial analysis of drosophila larval Hæmolymph. *Proc. Royal Soc. Edin. Sec. B. Biol. Sci.* 68, 215–236. doi: 10.1017/S0080455X00001053

Belaghzal, H., Dekker, J., and Gibcus, J. H. (2017). Hi-C 2.0: An optimized hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. *Methods* 123, 56–65. doi: 10.1016/j.ymeth.2017.04.004

Bell, J. C., Jukam, D., Teran, N. A., Risca, V. I., Smith, O. K., Johnson, W. L., et al. (2018). Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNA-to-DNA contacts. *Elife* 7:e27024. doi: 10.7554/eLife.27024

Belton, J.-M., McCord, R. P., Gibcus, J., Naumova, N., Zhan, Y., and Dekker, J. (2012). Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* 58, 268–276. doi: 10.1016/j.ymeth.2012.05.001

Bonetti, A., Agostini, F., Suzuki, A. M., Hashimoto, K., Pascarella, G., Gimenez, J., et al. (2020). RADICL-seq identifies general and cell type-specific principles of genome-wide RNA-chromatin interactions. *Nat. Commun.* 11:1018. doi: 10.1038/s41467-020-14337-6

Campos, P. F., and Gilbert, T. M. P. (2012). DNA extraction from formalin-fixed material. *Methods Mol. Biol.* 840, 81–85. doi: 10.1007/978-1-61779-516-9_11

Caudron-Herger, M., and Rippe, K. (2012). Nuclear architecture by RNA. *Curr. Opin. Genet. Dev.* 22, 179–187. doi: 10.1016/j.gde.2011.12.005

Chathoth, K. T., and Zabet, N. R. (2019). Chromatin architecture reorganization during neuronal cell differentiation in drosophila genome. *Genome Res.* 29, 613–625. doi: 10.1101/gr.246710.118

Comet, I., Schuettengruber, B., Sexton, T., and Cavalli, G. (2011). A chromatin insulator driving three-dimensional Polycomb response element (PRE) contacts and Polycomb association with the chromatin fiber. *PNAS* 108, 2294–2299. doi: 10.1073/pnas.1002059108

Cubeñas-Potts, C., Rowley, M. J., Lyu, X., Li, G., Lei, E. P., and Corces, V. G. (2017). Different enhancer classes in drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. *Nucleic Acids Res.* 45, 1714–1730. doi: 10.1093/nar/gkw1114

Cullen, K. E., Kladde, M. P., and Seyfred, M. A. (1993). Interaction between transcription regulatory regions of prolactin chromatin. *Science* 261, 203–206. doi: 10.1126/science.8327891

Dekker, J. (2006). The three "C" s of chromosome conformation capture: controls, controls, controls. *Nat. Methods* 3, 17–21. doi: 10.1038/nmeth823

Dekker, J. (2007). GC- and AT-rich chromatin domains differ in conformation and histone modification status and are differentially modulated by Rpd3p. *Genome Biol.* 8:R116. doi: 10.1186/gb-2007-8-6-r116

Dekker, J., Belmont, A. S., Guttman, M., Leshyk, V. O., Lis, J. T., Lomvardas, S., et al. (2017). The 4D nucleome project. *Nature* 549, 219–226. doi: 10.1038/nature23884

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* 295, 1306–1311. doi: 10.1126/science.1067799

Ding, D.-Q., Okamasa, K., Katou, Y., Oya, E., Nakayama, J.-I., Chikashige, Y., et al. (2019). Chromosome-associated RNA-protein complexes promote pairing of homologous chromosomes during meiosis in Schizosaccharomyces pombe. *Nat. Commun.* 10:5598. doi: 10.1038/s41467-019-13609-0

Downes, D. J., Beagrie, R. A., Gosden, M. E., Telenius, J., Carpenter, S. J., Nussbaum, L., et al. (2021). High-resolution targeted 3C interrogation of cis -regulatory element organization at genome-wide scale. *Nat. Commun.* 12:531. doi: 10.1038/s41467-020-20809-6

Eagen, K. P., Aiden, E. L., and Kornberg, R. D. (2017). Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map. *PNAS* 114, 8764–8769. doi: 10.1073/pnas.1701291114

Eagen, K. P., Hartl, T. A., and Kornberg, R. D. (2015). Stable chromosome condensation revealed by chromosome conformation capture. *Cell* 163, 934–946. doi: 10.1016/j.cell.2015.10.026

Echalier, G., Perrimon, N., and Mohr, S. E. (2018). *Drosophila Cells in Culture*. Netherlands: Elsevier.

El-Sharnouby, S., Fischer, B., Magbanua, J. P., Umans, B., Flower, R., Choo, S. W., et al. (2017). Regions of very low H3K27me3 partition the drosophila genome into topological domains. *PLoS One* 12:e0172725. doi: 10.1371/journal.pone.0172725

Falk, M., Feodorova, Y., Naumova, N., Imakaev, M., Lajoie, B. R., Leonhardt, H., et al. (2019). Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* 570, 395–399. doi: 10.1038/s41586-019-1275-3

Flyamer, I. M., Gassler, J., Imakaev, M., Brandão, H. B., Ulianov, S. V., Abdennur, N., et al. (2017). Single-nucleus hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* 544, 110–114. doi: 10.1038/nature21711

Fujiwara, Y., Wada, K., and Kabuta, T. (2017). Lysosomal degradation of intracellular nucleic acids—multiple autophagic pathways. *J. Biochem.* 161, 145–154. doi: 10.1093/jb/mvw085

Gabdank, I., Ramakrishnan, S., Villeneuve, A. M., and Fire, A. Z. (2016). A streamlined tethered chromosome conformation capture protocol. *BMC Genomics* 17:274. doi: 10.1186/s12864-016-2596-3

Gavrilov, A. A. (2016). Doctor of Sc. Thesis. Spatial organization of the eukaryote genome in the context of transcription regulation. Available at: http://www.genebiology.ru/dissovet/o-dissertatcionnom-sovete/. (Accessed August 27, 2021).

Gavrilov, A. A., Golov, A. K., and Razin, S. V. (2013a). Actual ligation frequencies in the chromosome conformation capture procedure. *PLoS One* 8:e60403. doi: 10.1371/journal.pone.0060403.

Gavrilov, A. A., Gushchanskaya, E. S., Strelkova, O., Zhironkina, O., Kireev, I. I., Iarovaia, O. V., et al. (2013b). Disclosure of a structural milieu for the proximity ligation reveals the elusive nature of an active chromatin hub. *Nucleic Acids Res.* 41, 3563–3575. doi: 10.1093/nar/gkt067

Gavrilov, A. A., and Razin, S. V. (2008). Spatial configuration of the chicken α-globin gene domain: immature and active chromatin hubs. *Nucleic Acids Res.* 36, 4629–4640. doi: 10.1093/nar/gkn429

Gavrilov, A., and Razin, S. V. (2009). Formaldehyde fixation of cells does not greatly reduce the ability to amplify cellular DNA. *Anal. Biochem.* 390, 94–96. doi: 10.1016/j.ab.2009.04.018

Gavrilov, A., Razin, S. V., and Cavalli, G. (2015). In vivo formaldehyde cross-linking: it is time for black box analysis. *Brief. Funct. Genomics* 14, 163–165. doi: 10.1093/bfgp/elu037

Gavrilov, A., Zharikova, A. A., Galitsyna, A. A., Luzhin, A. V., Rubanova, N. M., Golov, A. K., et al. (2020). Studying RNA-DNA interactome by red-C identifies noncoding RNAs associated with various chromatin types and reveals transcription dynamics. *Nucleic Acids Res.* 48, 6699–6714. doi: 10.1093/nar/gkaa457

Golov, A. K., Gavrilov, A. A., and Razin, S. V. (2015). The role of crowding forces in juxtaposing β-globin gene domain remote regulatory elements in mouse erythroid cells. *PLoS One* 10:e0139855. doi: 10.1371/journal.pone.0139855

Golov, A. K., Ulianov, S. V., Luzhin, A. V., Kalabusheva, E. P., Kantidze, O. L., Flyamer, I. M., et al. (2020). C-TALE, a new cost-effective method for targeted enrichment of hi-C/3C-seq libraries. *Methods* 170, 48–60. doi: 10.1016/j.ymeth.2019.06.022

Gridina, M., Mozheiko, E., Valeev, E., Nazarenko, L. P., Lopatkina, M. E., Markova, Z. G., et al. (2021). A cookbook for DNase hi-C. *Epigenetics Chromatin* 14:15. doi: 10.1186/s13072-021-00389-5

Hagège, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., et al. (2007). Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat. Protoc.* 2, 1722–1733. doi: 10.1038/nprot.2007.243

Hall, L. L., and Lawrence, J. B. (2016). RNA as a fundamental component of interphase chromosomes: could repeats prove key? *Curr. Opin. Genet. Dev.* 37, 137–147. doi: 10.1016/j.gde.2016.04.005

Hayashi, K., Nakazawa, M., Ishizaki, Y., and Obayashi, A. (1985). Influence of monovalent cations on the activity of T4 DNA ligase in the presence of polyethylene glycol. *Nucleic Acids Res.* 13, 3261–3271. doi: 10.1093/nar/13.9.3261

Hou, C., Li, L., Qin, Z. S., and Corces, V. G. (2012). Gene density, transcription, and insulators contribute to the partition of the drosophila genome into physical domains. *Mol. Cell* 48, 471–484. doi: 10.1016/j.molcel.2012.08.031

Hug, C. B., Grimaldi, A. G., Kruse, K., and Vaquerizas, J. M. (2017). Chromatin architecture emerges during zygotic genome activation independent of transcription. *Cell* 169, 216–228. doi: 10.1016/j.cell.2017.03.024

Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., and Chen, L. (2012). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.* 30, 90–98. doi: 10.1038/nbt.2057

Kawane, K., Motani, K., and Nagata, S. (2014). DNA degradation and its defects. *Cold Spring Harb. Perspect. Biol.* 6:a016394. doi: 10.1101/cshperspect.a016394

Kim, T. H., and Dekker, J. (2018a). ChIP. *Cold Spring Harb. Protoc.* 4, 314–316. doi: 10.1101/pdb.prot082610

Kim, T. H., and Dekker, J. (2018b). Formaldehyde cross-linking. *Cold Spring Harb. Protoc.* 4, 306–310. doi: 10.1101/pdb.prot082594

Larrivee, D. C. (1979). *A Biochemical Analysis of the Drosophila Rhabdomere and its Extracellular Environment. Ph.D. Thesis. Doctoral Dissertation.* Indiana: Purdue University

Lechardeur, D., Sohn, K.-J., Haardt, M., Joshi, P. B., Monck, M., Graham, R. W., et al. (1999). Metabolic instability of plasmid DNA in the cytosol: a potential barrier to gene transfer. *Gene Ther.* 6, 482–497. doi: 10.1038/sj.gt.3300867

Li, L., Lyu, X., Hou, C., Takenaka, N., Nguyen, H. Q., Ong, C.-T., et al. (2015). Widespread rearrangement of 3D chromatin organization underlies Polycomb-mediated stress-induced silencing. *Mol. Cell* 58, 216–231. doi: 10.1016/j.molcel.2015.02.023

Li, X., Zhou, B., Chen, L., Gou, L.-T., Li, H., and Fu, X.-D. (2017). GRID-seq reveals the global RNA-chromatin interactome. *Nat. Biotechnol.* 35, 940–950. doi: 10.1038/nbt.3968

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293. doi: 10.1126/science.1181369

Lo Sardo, F. (2021). "The chromosome conformation capture (3C) in Drosophila melanogaster," in *Capturing Chromosome Conformation: Methods and Protocols Methods in Molecular Biology*. eds. B. Bodega and C. Lanzuolo (New York, NY: Springer US), 9–17.

Louwers, M., Splinter, E., van Driel, R., de Laat, W., and Stam, M. (2009). Studying physical chromatin interactions in plants using chromosome conformation capture (3C). *Nat. Protoc.* 4, 1216–1229. doi: 10.1038/nprot.2009.113

Méndez, J., and Stillman, B. (2000). Chromatin Association of Human Origin Recognition Complex, Cdc6, and Minichromosome maintenance proteins during the cell cycle: assembly of Prereplication complexes in late mitosis. *Mol. Cell. Biol.* 20, 8602–8612. doi: 10.1128/MCB.20.22.8602-8612.2000

Michieletto, D., and Gilbert, N. (2019). Role of nuclear RNA in regulating chromatin structure and transcription. *Curr. Opin. Cell Biol.* 58, 120–125. doi: 10.1016/j.ceb.2019.03.007

Miele, A., Gheldof, N., Tabuchi, T. M., Dostie, J., and Dekker, J. (2006). Mapping chromatin interactions by chromosome conformation capture. *Curr. Protoc. Mol. Biol.* 74, 21.11.1–21.11.20. doi: 10.1002/0471142727.mb2111s74

Nagano, T., Lubling, Y., Stevens, T. J., Schoenfelder, S., Yaffe, E., Dean, W., et al. (2013). Single-cell hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502, 59–64. doi: 10.1038/nature12593

Nagano, T., Lubling, Y., Várnai, C., Dudley, C., Leung, W., Baran, Y., et al. (2017). Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* 547, 61–67. doi: 10.1038/nature23001

Nagano, T., Lubling, Y., Yaffe, E., Wingett, S. W., Dean, W., Tanay, A., et al. (2015a). Single-cell hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell. *Nat. Protoc.* 10, 1986–2003. doi: 10.1038/nprot.2015.127

Nagano, T., Várnai, C., Schoenfelder, S., Javierre, B.-M., Wingett, S. W., and Fraser, P. (2015b). Comparison of hi-C results using in-solution versus in-nucleus ligation. *Genome Biol.* 16:175. doi: 10.1186/s13059-015-0753-7

Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B. R., Mirny, L. A., et al. (2013). Organization of the mitotic chromosome. *Science* 342, 948–953. doi: 10.1126/science.1236083

Naumova, N., Smith, E. M., Zhan, Y., and Dekker, J. (2012). Analysis of long-range chromatin interactions using chromosome conformation capture. *Methods* 58, 192–203. doi: 10.1016/j.ymeth.2012.07.022

Oksuz, B. A., Yang, L., Abraham, S., Venev, S. V., Krietenstein, N., Parsi, K. M., et al. (2020). Systematic evaluation of chromosome conformation capture assays. BioRxiv 2020:12.26.424448. doi: 10.1101/2020.12.26.424448

Orlando, V., Strutt, H., and Paro, R. (1997). Analysis of chromatin structure by in vivo formaldehyde cross-linking. *Methods* 11, 205–214. doi: 10.1006/meth.1996.0407

Raae, A. J., Kleppe, R. K., and Kleppe, K. (1975). Kinetics and effect of salts and polyamines on T4 polynucleotide ligase. *Eur. J. Biochem.* 60, 437–443. doi: 10.1111/j.1432-1033.1975.tb21021.x

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A three-dimensional map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680. doi: 10.1016/j.cell.2014.11.021

Razin, S. V., Ulianov, S. V., and Gavrilov, A. A. (2019). 3D genomics. *Mol. Biol. (Mosk)* 53, 911–923. doi: 10.1134/S0026898419060156

Rowley, M. J., Nichols, M. H., Lyu, X., Ando-Kuri, M., Rivera, I. S. M., Hermetz, K., et al. (2017). Evolutionarily conserved principles predict 3D chromatin organization. *Mol. Cell* 67, 837–852. doi: 10.1016/j.molcel.2017.07.022

Schalbetter, S. A., Fudenberg, G., Baxter, J., Pollard, K. S., and Neale, M. J. (2019). Principles of meiotic chromosome assembly revealed in *S. cerevisiae*. *Nat. Commun.* 10:4795. doi: 10.1038/s41467-019-12629-0

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., et al. (2012). Three-dimensional folding and functional organization principles

of the drosophila genome. *Cell* 148, 458–472. doi: 10.1016/j.cell. 2012.01.010

Shidlovskii, Y. V., Bylino, O. V., Shaposhnikov, A. V., Kachaev, Z. M., Lebedeva, L. A., Kolesnik, V. V., et al. (2021). Subunits of the PBAP chromatin Remodeler are capable of mediating enhancer-driven transcription in drosophila. *Int. J. Mol. Sci.* 22:2856. doi: 10.3390/ijms22062856

Splinter, E., de Wit, E., van de Werken, H. J. G., Klous, P., and de Laat, W. (2012). Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods* 58, 221–230. doi: 10.1016/j.ymeth.2012.04.009

Splinter, E., Grosveld, F., and de Laat, W. (2004). 3C technology: analyzing the spatial organization of genomic loci in vivo. *Methods Enzymol.* 375, 493–507. doi: 10.1016/s0076-6879(03)75030-7

Sridhar, B., Rivas-Astroza, M., Nguyen, T. C., Chen, W., Yan, Z., Cao, X., et al. (2017). Systematic mapping of RNA-chromatin interactions In vivo. *Curr. Biol.* 27, 602–609. doi: 10.1016/j.cub.2017.01.011

Stadhouders, R., Kolovos, P., Brouwer, R., Zuin, J., van den Heuvel, A., Kockx, C., et al. (2013). Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat. Protoc.* 8, 509–524. doi: 10.1038/nprot.2013.018

Stadler, M. R., Haines, J. E., and Eisen, M. B. (2017). Convergence of topological domain boundaries, insulators, and polytene interbands revealed by high-resolution mapping of chromatin contacts in the early Drosophila melanogaster embryo. *elife* 6:e29550. doi: 10.7554/eLife.29550

Stewart, B. A., Atwood, H. L., Renger, J. J., Wang, J., and Wu, C.-F. (1994). Improved stability of drosophila larval neuromuscular preparations in haemolymph-like physiological solutions. *J. Comp. Physiol. A* 175, 179–191. doi: 10.1007/BF00215114

Thakur, J., and Henikoff, S. (2020). Architectural RNA in chromatin organization. *Biochem. Soc. Trans.* 48, 1967–1978. doi: 10.1042/BST20191226

Tolhuis, B., Palstra, R.-J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active β-globin locus. *Mol. Cell* 10, 1453–1465. doi: 10.1016/S1097-2765(02)00781-5

Ulianov, S. V., Doronin, S. A., Khrameeva, E. E., Kos, P. I., Luzhin, A. V., Starikov, S. S., et al. (2019). Nuclear lamina integrity is required for proper spatial organization of chromatin in drosophila. *Nat. Commun.* 10, 1–11. doi: 10.1038/s41467-019-09185-y

Ulianov, S. V., Khrameeva, E. E., Gavrilov, A. A., Flyamer, I. M., Kos, P., Mikhaleva, E. A., et al. (2016). Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res.* 26, 70–84. doi: 10.1101/gr.196006.115

Ulianov, S. V., Zakharova, V. V., Galitsyna, A. A., Kos, P. I., Polovnikov, K. E., Flyamer, I. M., et al. (2021). Order and stochasticity in the folding of individual drosophila genomes. *Nat. Commun.* 12:41. doi: 10.1038/s41467-020-20292-z

van Berkum, N. L., and Dekker, J. (2009). "determining spatial chromatin organization of large genomic regions using 5C technolosgy," in *Chromatin Immunoprecipitation Assays: Methods and Protocols Methods in Molecular Biology.* ed. P. Collas (Totowa, NJ: Humana Press), 189–213.

van de Werken, H. J. G., de Vree, P. J. P., Splinter, E., Holwerda, S. J. B., Klous, P., de Wit, E., et al. (2012). "4C technology: protocols and data analysis," in *Methods in Enzymology.* eds. C. Wu and C. David Allis (Netherlands: Elsevier), 89–112.

van der Meer, J. M., and Jaffe, L. F. (1983). Elemental composition of the perivitelline fluid in early drosophila embryos. *Dev. Biol.* 95, 249–252. doi: 10.1016/0012-1606(83)90025-8

Vermeulen, C., Allahyar, A., Bouwman, B. A. M., Krijger, P. H. L., Verstegen, M. J. A. M., Geeven, G., et al. (2020). Multi-contact 4C: long-molecule sequencing of complex proximity ligation products to uncover local cooperative and competitive chromatin topologies. *Nat. Protoc.* 15, 364–397. doi: 10.1038/s41596-019-0242-7

Yan, Z., Huang, N., Wu, W., Chen, W., Jiang, Y., Chen, J., et al. (2019). Genome-wide colocalization of RNA-DNA interactions and fusion RNA pairs. *Proc. Natl. Acad. Sci. U. S. A.* 116, 3328–3337. doi: 10.1073/pnas.1819788116

Yang, W. (2011). Nucleases: diversity of structure, function and mechanism. *Q. Rev. Biophys.* 44, 1–93. doi: 10.1017/S0033583510000181

Zhang, Y., and Li, G. (2020). Advances in technologies for 3D genomics research. *Sci. China Life Sci.* 63, 811–824. doi: 10.1007/s11427-019-1704-2

Diversity of Nuclear Lamin A/C Action as a Key to Tissue-Specific Regulation of Cellular Identity in Health and Disease

Anna Malashicheva* and Kseniya Perepelina*

Laboratory of Regenerative Biomedicine, Institute of Cytology, Russian Academy of Sciences, St. Petersburg, Russia

A-type lamins are the main structural components of the nucleus, which are mainly localized at the nucleus periphery. First of all, A-type lamins, together with B-type lamins and proteins of the inner nuclear membrane, form a stiff structure—the nuclear lamina. Besides maintaining the nucleus cell shape, A-type lamins play a critical role in many cellular events, such as gene transcription and epigenetic regulation. Nowadays it is clear that lamins play a very important role in determining cell fate decisions. Various mutations in genes encoding A-type lamins lead to damages of different types of tissues in humans, collectively known as laminopathies, and it is clear that A-type lamins are involved in the regulation of cell differentiation and stemness. However, the mechanisms of this regulation remain unclear. In this review, we discuss how A-type lamins can execute their regulatory role in determining the differentiation status of a cell. We have summarized recent data focused on lamin A/C action mechanisms in regulation of cell differentiation and identity development of stem cells of different origin. We also discuss how this knowledge can promote further research toward a deeper understanding of the role of lamin A/C mutations in laminopathies.

Keywords: lamin A/C, laminopathies, *LMNA* gene, cell differentiation, LADs, regulation of gene expression, chromatin organization

INTRODUCTION

A-type lamins are the structural components of the nucleus, which together with B-type lamins and inner nuclear membrane proteins form a scaffold, termed the nuclear lamina. Mostly A-type lamins are included in the lamina at the nucleus periphery; however, a small fraction of lamin A/C is found throughout the nucleoplasm (Naetar et al., 2017; Briand and Collas, 2020). Primarily, nuclear lamin A/C was thought to undertake solely a structural role, providing shape and stiffness to the

Abbreviations: BAF, barrier to autointegration factor; CaaX, motive required for prelamin A processing, where C is cysteine, a – an aliphatic amino acid and X can be any amino acid; CDK, cyclin-dependent kinase; cLADs, constitutive lamina-associated domains; CTCF, CCCTC binding factor; DamID, DNA adenine methyltransferase identification; ECM, extracellular matrix; EDMD, Emery-Dreifuss muscular dystrophy; FAC, focal adhesion complex; F-actin, filamentous actin; fLADs, facultative lamina-associated domains; FPLD, Dunnigan-type familial partial lipodystrophy; FTase, farnesyltransferase enzyme; ICMT, isoprenylcysteine carboxyl methyltransferase; Ifs, intermediate filaments; INM, inner nuclear membrane; iPSC, indused pluripotent stem cell; LADs, lamina-associated domains; LAP, lamina-associated polypeptide; LBR, LMNB receptor; LINC, linker of nucleoskeleton and cytoskeleton complex; MAD, Mandibuloacral dysplasia; MSC, mesenchymal stem cell; MTs, microtubules; NETs, nuclear envelope transmembrane proteins; NLS, nuclear localization signal; NPC, nuclear pore complex; ONM, outer nuclear membrane; PNS, perinuclear space (space between ONM and INM); pRb, retinoblastoma protein; Rce1, RAS converting enzyme 1; Zmpste24, zinc metalloprotease.

nucleus. Currently, A-type lamins are known as essential regulators of gene expression and key mediators of cell fate determination. Lamin A/C participation in chromatin organization, DNA replication, gene transcription regulation, cell differentiation, and tissue-specific functions has been extensively investigated (de Leeuw et al., 2018; Zhang et al., 2019; Alcorta-Sevillano et al., 2020). A-type lamins perform most of these functions by interacting with the inner nuclear membrane proteins, transcription factors, and DNA (Parnaik, 2008; Prokocimer et al., 2009; Lambert, 2019; Almendáriz-Palacios et al., 2020). A-type lamins are believed to regulate important signaling pathways' activity in cells (such as Rb/E2F, Wnt/β-catenin, TGFβ, Notch) through their direct or indirect interactions with other proteins (Maraldi et al., 2010, 2011; Gerace and Tapia, 2018).

In this review we mainly focused on recent data about the role of lamin A/C in cell differentiation. The latter has become a focus of increasing attention over the past decade, initially because of the identification of new mutations in the *LMNA* gene and associated diseases—laminopathies (Wong and Stewart, 2020). The most famous lamin modification is progerin, causing a severe developmental disorder—premature aging syndrome, or progeria. This known disease is extremely rare (Worman, 2012; Schreiber and Kennedy, 2013; Gonzalo et al., 2017). At the same time, *LMNA* point mutations leading to damage of various types of tissues occur more often. It is supposed that the development of the disease is associated with an abnormality of the stem cell differentiation process. Taking into account that A-type lamins are expressed in all differentiated cell types, it seems still unclear why only certain differentiated tissues are selectively affected for each type of laminopathy (Worman, 2012; Gruenbaum and Foisner, 2015; Robson et al., 2016; Zhang et al., 2019).

Recently, research has focused on the molecular mechanisms of laminopathies' development. The mechanisms proposed for the pathology development include disturbances in chromatin organization, intracellular signal transduction, as well as epigenetic changes. Consequently, all this leads to dysregulation of genes responsible for cell differentiation (van Steensel and Belmont, 2017; Zhang et al., 2019). Regions of lamin–chromatin interaction (lamina-associated domains—LADs) are known to be implicated in regulation of gene expression. LADs contain a diversity of differentiation-related genes, that are in an active or inactive state depending on their association with chromatin. Active gene expression is associated with releasing LADs from the nuclear lamina. In contrast, inactivation of expression is associated with the attachment of LADs to the lamina (Briand and Collas, 2020; Bitman-Lotan and Orian, 2021; Shah et al., 2021).

Apparently, cellular context is essential for further development of tissue-specific disease phenotypes. There is strong evidence of lamin A/C involvement in the processes of cell differentiation, in particular, in the adipogenic (Oldenburg et al., 2017; Perepelina et al., 2018), osteogenic (Avnet et al., 2011; Alcorta-Sevillano et al., 2020), myogenic (Steele-Stallard et al., 2018), and cardiogenic (Mounkes et al., 2005; Shah et al., 2021) directions. Recently, a theory has been proposed about the action of mechanical signals coming from the extracellular matrix to the lamin network resulting in a redistribution of chromatin and a change in the availability of DNA for transcription factors (Osmanagic-Myers et al., 2015; Martino et al., 2018; Donnaloja et al., 2020). Recent studies show that all these events affect the biophysical properties of the nucleus, which, in turn, affects the fate and differentiation of cells (Alcorta-Sevillano et al., 2020).

The difficulties in studying the role of lamin A/C in cell differentiation are in particular related to the lack of a unified experimental model. Some studies were performed using mesenchymal stem cells, fibroblast (van Tienen et al., 2019; Ikegami et al., 2020), cardiac mesenchymal cells (Perepelina et al., 2019), osteoblast precursors (Avnet et al., 2011; Y. Liu et al., 2012), etc. Some studies used mouse models (Arimura et al., 2005; Mounkes et al., 2005; Le Dour et al., 2017; Hamczyk et al., 2018). Currently, studying the mechanisms of laminopathies' development using induced pluripotent stem cell (iPSC) models is especially relevant (Crasto and Di Pasquale, 2018; Steele-Stallard et al., 2018; Giacomelli et al., 2020; Shah et al., 2021).

Here we summarized the current knowledge about the crucial role of lamin A/C in regulating differentiation of stem cells of various origin. We specifically focus on the issue of how the regulation of differentiation by A-type lamins could be involved in the pathogenesis of laminopathies for which generally accepted concept of the pathogenesis is still absent.

ORGANIZATION, MATURATION, AND ASSEMBLY OF LAMIN A/C

Nuclear lamins in metazoan cells are members of the type V intermediate filament (IF) family. There are two groups of lamins, the A type and the B type, which, in association with inner nuclear membrane proteins, form a stiff meshwork under the inner nuclear membrane termed the nuclear lamina (**Figure 1**). While B-type lamins are expressed overall in all cells, A-type lamins are only expressed in differentiated cells, which apparently determines the specific functions of this type of lamin in the cell. Moreover, the expression level of lamin A/C varies in different tissues (Gruenbaum and Foisner, 2015). As a result of alternative splicing of *LMNA* gene transcript, several isoforms such as A, C and minor isoforms AΔ10 and C2 are generated.

Along with all IFs, lamin filaments contain three structural domains: a central α-helical rod domain, a short globular amino-terminal "head" domain and a long carboxyterminal "tail" domain. The rod domain of lamins includes three helical segments (1A, 1B, and 2), connected by short linkers L1 and L12 (Ahn et al., 2019). Lamins have several differences from cytoplasmic IFs: (1) they contain 42 additional amino acids in their rod domain; (2) they have a shorter head domain; and (3) their carboxyl-terminal "tail" domain includes the nuclear localization signal (NLS)—which is required for their nuclear transport after synthesis in the cytoplasm—, an immunoglobulin-like (Ig-) fold domain, a chromatin binding site, and—with the exception of lamin C—a CaaX motif (where C is cysteine, a—an aliphatic amino acid, and X—any amino acid) (Wu et al., 2014; Gruenbaum and Foisner, 2015; **Figure 2**).
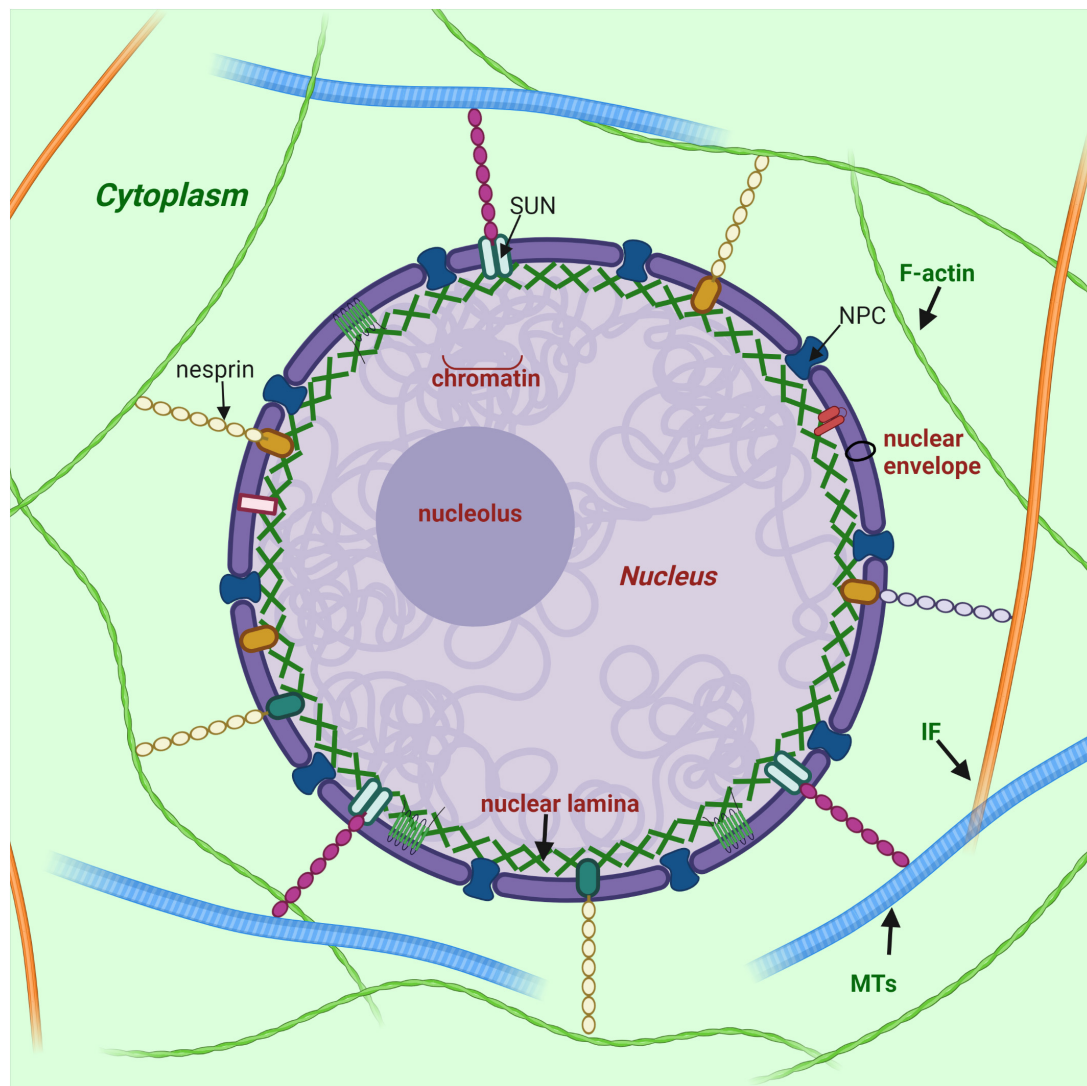
FIGURE 1 | Nuclear lamina position and its interplay with other structures of cell. Nuclear lamina is a stiff meshwork consisting of A-type lamins and B-type localized between the nuclear envelope and chromatin. Nuclear lamins interact with a wide range of nuclear envelope proteins (NEPs). Also, nuclear lamins can interact with the cytoskeleton (filamentous actin – F actin; microtubules – MTs; and intermediate filaments – IF) *via* SUN proteins and nesprins. Created with BioRender.com.
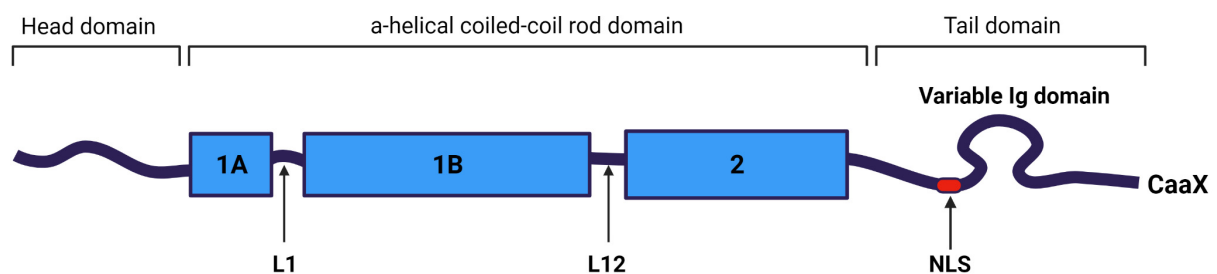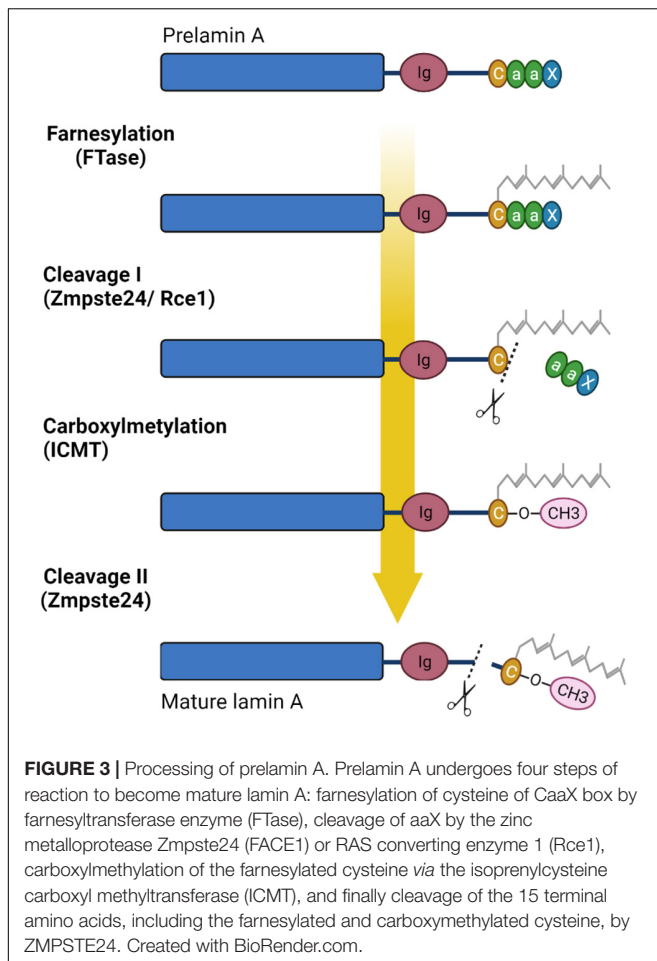

FIGURE 2 | Structural organization of lamin A/C filament. Lamin A/C is an intermediate filament, and contains a central coiled-coil rod domain divided into 1A, 1B, and 2 sub-domains connected *via* L1 and L12 linkers; a head domain; and a tail domain containing nuclear localization signal (NLS), Ig-like domain and carboxyterminal CaaX box (apart of lamin C), where C is cysteine, a – aliphatic amino acid, and X can be any amino acid. Created with BioRender.com.

FIGURE 3 | Processing of prelamin A. Prelamin A undergoes four steps of reaction to become mature lamin A: farnesylation of cysteine of CaaX box by farnesyltransferase enzyme (FTase), cleavage of aaX by the zinc metalloprotease Zmpste24 (FACE1) or RAS converting enzyme 1 (Rce1), carboxylmethylation of the farnesylated cysteine *via* the isoprenylcysteine carboxyl methyltransferase (ICMT), and finally cleavage of the 15 terminal amino acids, including the farnesylated and carboxymethylated cysteine, by ZMPSTE24. Created with BioRender.com.

Lamin C is translated as a mature protein without multiple post-translated modifications as in the case of A-type lamin, and lacking 98 amino acids with CaaX motif; A-type lamin is expressed in cells as prelamin, which undergoes multiple post-translational modifications of the carboxyterminal "tail" domain. At the first stage of prelamin A processing, farnesyltransferase enzyme (FTase) adds a farnesyl group to the C-terminal cysteine. Then the three residues (aaX) are cleavaged *via* the zinc metalloprotease Zmpste24 (FACE1) or RAS converting enzyme 1 (Rce1). At the next stage C-terminal cysteine is carboxymethylated by the isoprenylcysteine carboxyl methyltransferase (ICMT). Finally, enzyme Zmpste24 cleaves the last 15 C-terminal amino-acids of lamin A, thereby removing the carboxy farnesylated and methylated cysteine (Fisher et al., 1986; Corrigan et al., 2005; Adam et al., 2013; Wu et al., 2014). **Figure 3** presents the lamin A post-translational modifications during maturation.

Other types of post-translational modifications of lamins are known such as sumoylation, ubiquitylation, acetylation, and phosphorylation. These modifications obviously play a significant role in regulating lamin translocation during the cell cycle (Prokocimer et al., 2009; Donnaloja et al., 2020). Phosphorylation of lamins is involved in plenty of cellular process. To date,

some research has shown that phosphorylation contributes to the interaction between B-type lamins and histone H2A/H2B in *Drosophila* (Mattout et al., 2007). In mammalian cells two specific sites flanking the lamin's rod domain are phosphorylated by cyclin-dependent kinase (CDK)-1. This event is required for lamin disassembly into dimers during mitosis (Chaffee et al., 2014; Naetar et al., 2017). Moreover, phosphorylation contributes to the dynamic interaction of lamins with other proteins as well as lamin A/C solubility, and lamina meshwork formation. Remarkably, all these processes could be activated/inactivated, as phosphorylation is a reversible modification (Kochin et al., 2014; Liu and Ikegami, 2020). Apparently, lamin phosphorylation takes place in the modulation of enhancer activity. It has been reported that S22-phosphorylated lamins connect with active genomic enhancer sites and this interaction is violated in progeroid cells (Ikegami et al., 2020). Sumoylation has been shown to be important for normal lamin A/C functions and also for the regulation of lamin A/C assembly (Kim et al., 2011). It has been shown that mutant lamin A/C (E203G and E203K) leads to a decreased level of lamin sumoylation in fibroblasts and increased cell death (Zhang and Sarge, 2008).

The main structural unit of lamin A/C filaments is a coiled dimer formed as a result of the interaction of two central rod domains of lamin proteins. These dimers are connected head-to-tail and form protofilaments, which could be combined in various configurations to form 10 nm lamin filaments (Herrmann and Aebi, 2004; Prokocimer et al., 2009; Gruenbaum and Foisner, 2015). The structure of the lamin "tail" domain is similar to immunoglobulin and could mediate specific intermolecular interactions with other proteins (Donnaloja et al., 2020; **Figure 4**).

Recent evidence suggests that the structure of the nuclear lamina in somatic cells is more complex and less homogeneous than was previously thought. It has been shown by cryo-electron tomography tests that lamin filaments are assembled not into 10 nm fibers, but into 3.5 nm fibers with a protruding zone of Ig domain (Donnaloja et al., 2020). In addition, the presence of intermediate "bridges" between neighboring lamina filaments, the nature of which is unknown, has been shown (Burke and Stewart, 2013). In mammals, most lamins can interact with each other; however, some evidence suggests that the bond strength between different lamins can vary, and that A-type and B-type lamins predominantly polymerize into separate homopolymers.

LAMIN A/C-BINDING PROTEINS

Undoubtedly, A-type lamins are essential components of the nucleus which perform a multiplicity of vital cell functions, from stabilization of nucleus shape to involvement in more complex processes such as cell proliferation, migration, signaling transduction, cell differentiation, and others (Gruenbaum and Foisner, 2015; Enyedi and Niethammer, 2017; Naetar et al., 2017; Karoutas and Akhtar, 2021). The abundance of lamin A/C functions is implemented through their direct or mediated interaction with a plethora of inner nuclear membrane (INM) and nucleoplasm proteins. Over 80 nuclear envelope
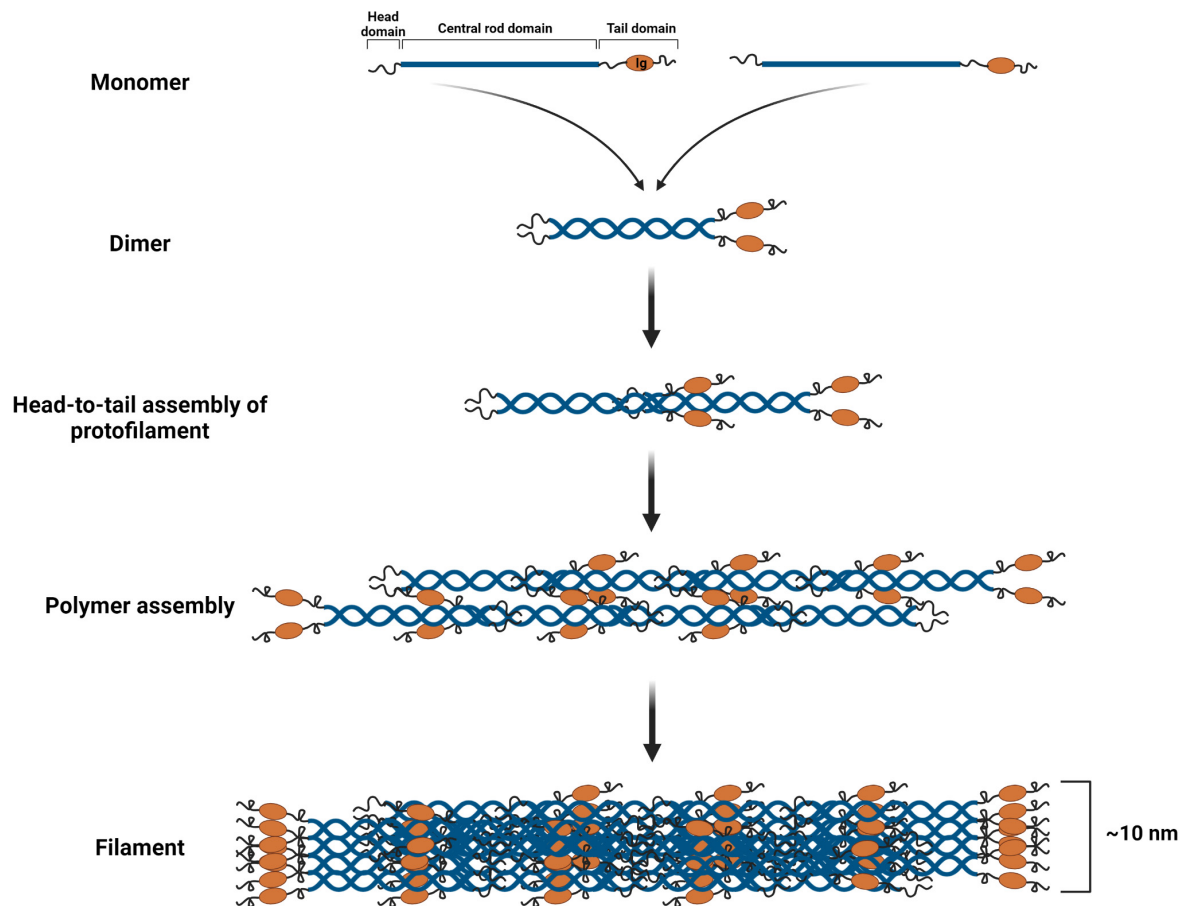
FIGURE 4 | Lamin A/C filament assembly. Lamin A/C dimers are formed from monomers that associate in a parallel, forming a coiled-coil through the central α-helical rod domain. Then lamin A/C dimers are assembled in a head-to-tail manner forming protofilament. Protofilaments through the anti-parallel association of lamin A/C polymers produce a polymer. Finally, several polymers cooperate resulting in a 10 nm lamin A/C filament. Created with BioRender.com.

transmembrane proteins (NETs) have been identified as likely interacting with lamins, by means of proteomic analyses in rat (Schirmer et al., 2003). The most important and widespread NETs include lamina-associated polypeptide (LAP) 1, LAP2a, LMNB receptor (LBR), and emerin (Almendáriz-Palacios et al., 2020). In addition, NETs were shown to vary in different tissues, and could contribute to the tissue-specific lamin A/C actions (Korfali et al., 2012).

Lamin-binding proteins are divided into three general groups: (1) proteins providing mechanical support of the nucleus by interacting with subnuclear elements, chromatin and INM; (2) signaling transmission components taking part in cell regulation of vital processes, such as cell differentiation, homeostasis, etc.; (3) proteins regulating gene expression and chromatin organization (Gruenbaum and Foisner, 2015; Martino et al., 2018; Zhang et al., 2019).

Among the proteins interacting with lamin A/C, emerin and lamina-associated polypeptides (LAPs) are to stand separately. Emerin, LAP2, and MAN1 (called LEM proteins) contain a special domain of 40 amino acid residues termed the LEM

domain, which interacts with the barrier to autointegration factor (BAF), a DNA-binding factor involved in organizing chromatin structure and assembling the nuclear envelope. There are LEM proteins that lack a transmembrane domain and therefore they are localized to the nucleoplasm or cytoplasm (Brachner and Foisner, 2011). In addition to the BAF-mediated effect on chromatin structure, lamins interact with epigenetic regulator ING, which binds to core histones, deacetylases, and histone acetyltransferases, as well as mediators of epigenetic regulation. Moreover, lamin A/C can also directly cooperate with chromatin by tethering specific chromatin regions called lamina-associated domains (LADs) at the nuclear periphery (Shevelyov and Ulianov, 2019; **Figure 5**).

SUN and KASH domain proteins are important nuclear membrane proteins localized in the INM and ONM (outer nuclear membrane), respectively. The SUN proteins interact directly with lamin A/C. KASH proteins bind to major cytoskeleton members, including actin filaments (through nesprin-1 and -2), intermediate filaments (via interaction with nesprin-3), and microtubules (via kinesin and dynein motor
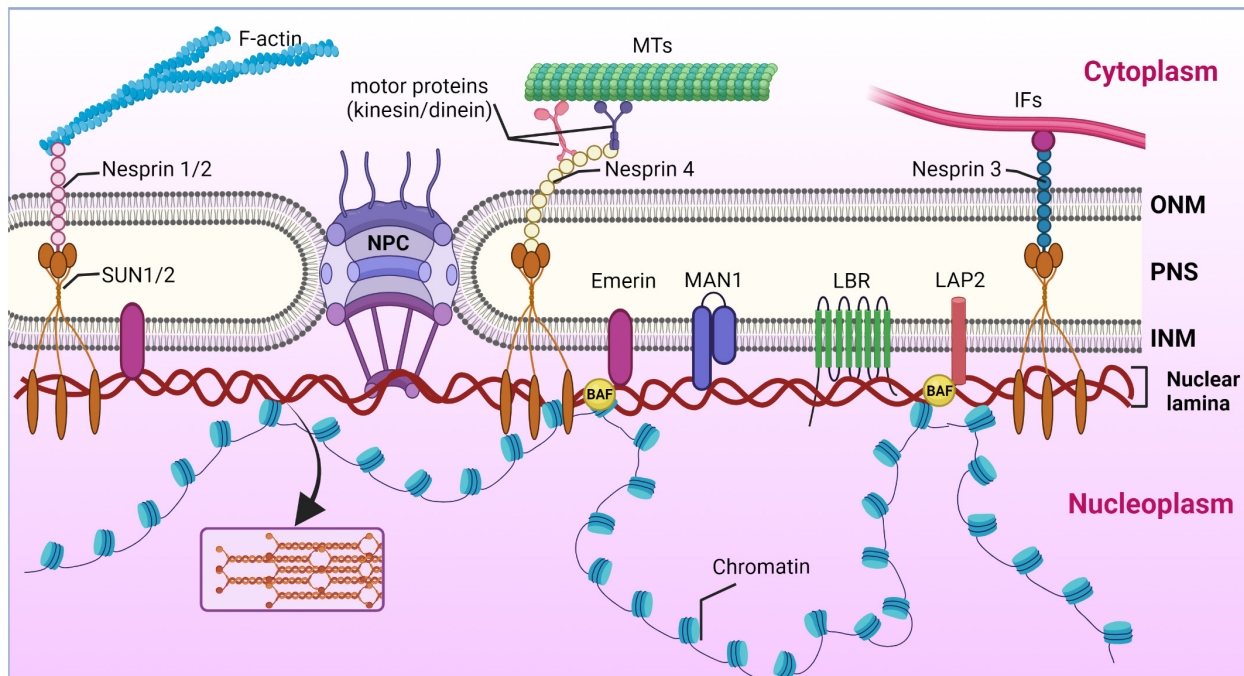
FIGURE 5 | Cooperation of nuclear lamina with nuclear envelope proteins and chromatin. Nuclear lamina is localized between the inner nuclear membrane (INM) and chromatin. Schematic representation of lamin interaction with inner nuclear membrane proteins, the most important of which are MAN1, LAP2, SUN1/2, Emerin, and LBR. The nuclear pore complex (NPC) spans both the inner nuclear membrane (INM) and the outer nuclear membrane (ONM) and mediates macromolecular transport. *Via* SUN1/2 and the nesprins interacting with them, located in the ONM, lamins cooperate with cytoskeleton components, namely filamentous actin (F-actin), microtubules (MTs), and intermediate filaments (IFs). The space between the ONM and INM is termed the perinuclear space (PNS). Created with BioRender.com.

proteins binding to nesprin-1, -2, -4, and KASH5) (Haque et al., 2006). Thus, SUN and KASH domain proteins, nesprins, together with lamin A/C, form a protein complex called the LINC (linker of nucleoskeleton and cytoskeleton) complex, which unites the nucleus and the cytoskeleton and enables force transmission across the nuclear envelope during nuclear positioning and migration (Lee and Burke, 2018).

In addition, through protein–protein interactions, A-type lamins are believed to interact with and regulate the activity and availability of important signaling pathway proteins in the cell, such as Rb/E2F, Wnt/β-catenin, TGFβ, SMAD, and MAPK (Gerbino et al., 2018; Worman, 2018). More details on the participation of lamin A/C and associated proteins will be outlined below.

LAMIN A/C AS TISSUE-SPECIFIC REGULATOR OF CELL DIFFERENTIATION

Lamin A/C Participate in Mechanosignaling Defining Cell Differentiation

Currently, the role of lamin A/C in mechanosignaling is considered to be essential for regulating vital processes in the cells including migration, homeostasis, growth, and differentiation

(Martino et al., 2018; Donnaloja et al., 2020). Mechanosignaling is the cell's ability to modulate the mechanical signals into a biological response by acting on several cell functions above. In this case, lamin A/C serve as mechanosensor, receiving external stimuli from the extracellular matrix (ECM), and then transforming them into internal biological responses. Thus, lamins are mediators, helping the cells to adapt to a changing microenvironment (Isermann and Lammerding, 2013; Guilluy et al., 2014; Osmanagic-Myers et al., 2015; Martino et al., 2018).

The first piece of knowledge about the fact that external mechanical force could lead to cell response resulting in nucleus deformation was obtained by Maniotis et al. (1997) in the 1990s. Since then, knowledge has been accumulating about the mechanisms underlying the signal transduction from the exogenous environment through the cytoplasm to the nucleus.

ECM includes many components (proteins, glycosaminoglycans, and proteoglycans) that impact the cell surface in a specific manner. The most common of them are collagen, laminin, and fibronectin (**Figure 6**). The ECM composition is unique for a given tissue and could be changed in response to alteration of the environment, especially in the case of a disease (Bonnans et al., 2014). The mechanical signal from ECM is transmitted to membrane-bound integrins that perform a sensor role. Integrins mediate the transformation of mechanical stimuli into biochemical signals. Interestingly, depending on the quantity and type of integrins, cells can react in a different way (Israeli-Rosenberg et al., 2014). Through the

accumulation of proteins termed focal adhesion complex (FAC), integrins are associated with the cytoskeleton. FAC proteins, such as talin, a-actinin, and vinculin, define the strength of interaction between integrins and filamentous actin (F-actin), which is a general cytoskeleton component (Chin et al., 2019). Then the signal is translocated *via* the LINC complex to nuclear lamins, the main sensors of mechanotransduction (**Figure 6**).

The importance of lamin A/C in mechanotransduction was confirmed in studies where cells lacking lamin A/C or expressing *LMNA* mutants were unable to directly transmit forces to the nucleus (Poh et al., 2012). In addition, signaling cascades, such as ROCK, Src, and ERK are known to be implicated in the mechanotransduction process (Osmanagic-Myers et al., 2015).

Despite the identification of a spectrum of molecular components involved in mechanotransduction, it remains completely unknown how these components act and adapt to each other to affect cellular functions and stem cell fate. The differentiation process is believed to be mechanosensitive, and cell fate could be determined by type and physical force of external stimuli. Current proposed model could be as follows. During cell differentiation A-type lamins get information about the changing microenvironment from nearby cells and ECM through the cytoskeleton. This leads to a rearranging meshwork and chromatin structures, or urges conformational changes in nuclear proteins such as transcription factors and components of signaling pathways. It is supposed that these conversions lead to chromatin segments' translocation away from or to the lamina, resulting in activation/repression of differentiation-related genes (Swift et al., 2013; Alcorta-Sevillano et al., 2020). First of all, this could be determined by the physical properties of the tissues. Some researchers have revealed correlations between substrate stiffness and gene transcription intensity of lamin A/C in a tissue-specific manner. For instance, Heo et al. (2016) have demonstrated that low external stimuli promote mesenchymal stem cell (MSC) to adipogenic differentiation associated with inhibited lamin A/C production. Other authors revealed that medium force stimuli induce MSC to differentiate into myocytes' direction, which is accompanied by elevation of lamin A/C expression (Engler et al., 2006; Swift et al., 2013). In addition, high lamin A/C expression level of hard tissues (such as bone) stabilizes the nucleus against mechanical stress. At the same time, soft tissues, such as fat, are characterized by a low expression level of lamin A/C. It has been demonstrated that lamin A/C knockdown enhances mesenchymal stem cell differentiation on a soft matrix, which contributed to fat phenotype development. In contrast, lamin A/C overexpression enhances cell differentiation on a stiff matrix toward a bone phenotype (Swift et al., 2013; Alcorta-Sevillano et al., 2020). In addition, lamin A/C overexpression leads to an inhibition of chromatin remodeling, and also to an activation of other actions such as expression of stress-related proteins implicated in cell differentiation, and transcriptional regulator YAP1 involved in cell proliferation and the suppression of apoptotic genes and Hippo pathway (Swift et al., 2013).

Thus, *via* adhesion proteins and cytoskeleton meshwork, ECM transmits information into the nucleus about the microenvironment to stabilize proper shape and stiffness of the nucleus by means of the quantity of lamins. High lamin A/C expression protects all components of the nucleus from severe forces coming from a stiff ECM, for example in a bone tissue. This mechanism reflects a mechanical theory of lamin A/C's role in the cells (Osmanagic-Myers and Foisner, 2019; **Figure 6**).

Some researchers have demonstrated the importance of the Ig-domain of lamin A/C in stress-related changes in terms of lamina rearrangement. In response to stress, electrostatic interaction between the positively charged Ig-tail domain and negatively charged regions of the rod domain of a nearby lamin's filament is disrupted, resulting in lamina reorganization (Makarov et al., 2019).

Thus, the expression level of lamin A/C determines tissue-specific differentiation of cells. In this way, mechanical signals coming from the intercellular matrix can direct lamins to proper stabilization of the genome in response to mechanical stress and tissue-specific gene expression during cell differentiation. These events are necessary to support nucleus shape and prevent the DNA from breaking.

Lamin A/C Regulates Chromatin Organization and Gene Expression

Genomic DNA in the eukaryote nucleus is known to be extensively packaged in chromosomes, each of which occupies a certain area termed the chromosome territory (Cremer and Cremer, 2010). According to transcriptional activity, chromatin is divided into euchromatin, which includes the majority of actively expressed genes, and heterochromatin, including transcriptionally inactive genes. Heterochromatin mostly occupies the nuclear periphery, whereas euchromatin is localized in the interior part of the nucleus. In addition, heterochromatin is sub-divided into constitutive heterochromatin, which is localized in the pericentromeric and subtelomeric regions of chromosomes, and facultative heterochromatin, localized in chromosome shoulders (Lieberman-aiden et al., 2009; Ou et al., 2017). It has been shown that heterochromatin is associated with lamin A/C forming the nuclear lamina, while euchromatin dominating in the nuclear interior is connected with a small number of nucleoplasmic lamin A/C. A-type lamins are considered to regulate the repressive state of genes included in facultative heterochromatin (Gruenbaum and Foisner, 2015; de Leeuw et al., 2018; Bitman-Lotan and Orian, 2021). This three-dimensional organization of chromatin contributes to the gene expression regulation and maintenance of silencing of heterochromatic genes.

Nowadays, the multiplicity of methods such as super-resolution microscopy (Cremer et al., 2017; Ricci et al., 2017), chromosome capture methods (Dekker et al., 2002), and chromatin immunoprecipitation (ChIP) allow deeper investigation of 3D nuclear architecture (Collas, 2010; Oldenburg and Collas, 2016). In this way direct interactions of chromatin with lamin A/C were identified using DNA adenine methyltransferase identification (DamID) (Van Steensel and Henikoff, 2000; Guelen et al., 2008) and chromatin immunoprecipitation methods (Lund et al., 2014, 2015). These regions now are broadly known as lamina-associated domains
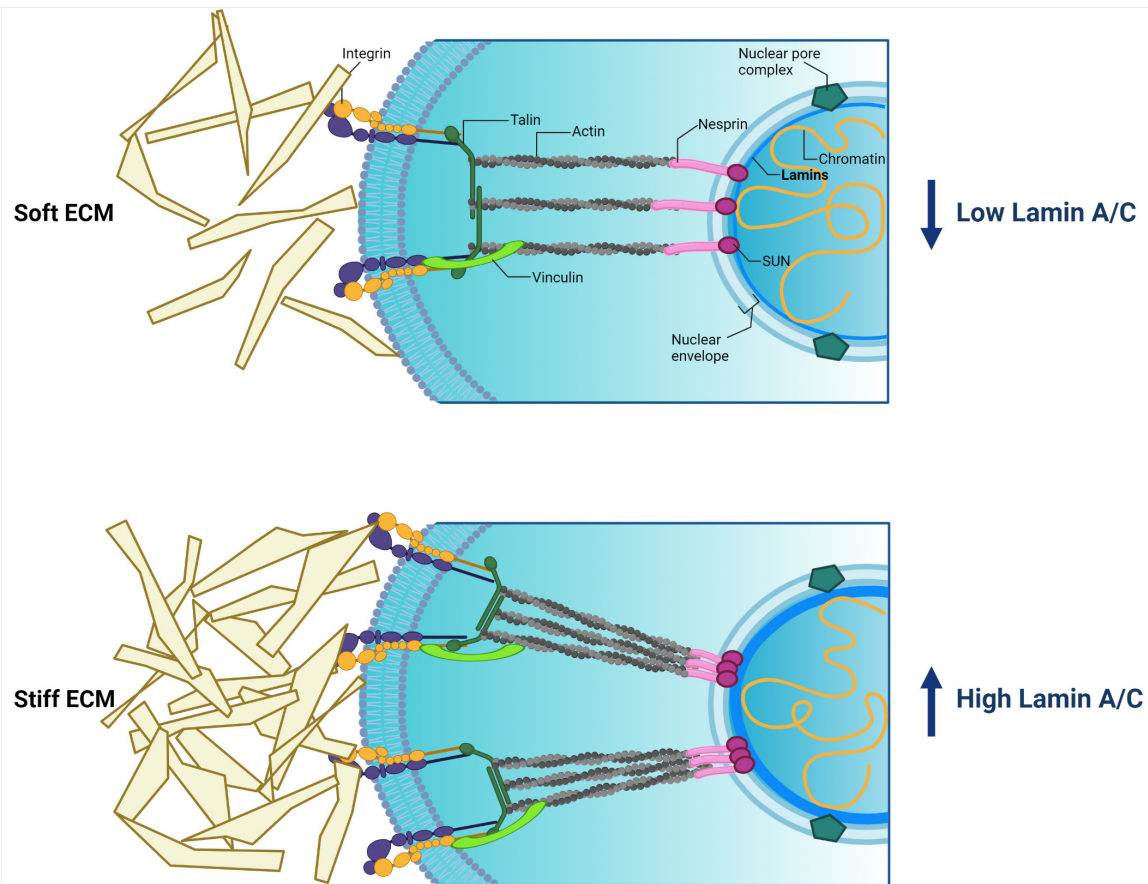
FIGURE 6 | Cytoplasmic and nuclear components involved in mechanotransduction and their relations. External stimuli coming from extracellular matrix (ECM) pass to cytoplasmic membrane integrins. Then the signal transmits though focal adhesion proteins (talin, a-actinin, vinculin) toward actin. Actin directly binds with the LINC complex (nesprin + SUN + lamins), resulting in the signal being transmitted into the nucleus. The force of external stimuli affects lamin A/C production, thereby driving nucleus stiffness and protecting the genome. Created with BioRender.com.

(LADs). Approximately 30–40% of the genome is occupied by LADs, which contain different gene sets in a silent state according to the particular type of cells. Moreover, it has been suggested that lamin A/C located in the nuclear interior as well as peripheral lamin A/C (as a part of lamina) are involved in gene repression (Naetar et al., 2017). Similar to heterochromatin, there are facultative and constitutive LADs (fLADs and cLADs, respectively). The set of cLADs is very identical in cells from several origins. Conversely, fLADs are unique for different cells types (Melcer and Meshorer, 2010). During several studies, it has been demonstrated that fLADs are spatially positioned in tissue-specific and embryo stage-dependent ways (Robson et al., 2016; Poleshko et al., 2017, 2019). Recent research conducted on induced pluripotent stem cells (iPSC) carrying a tissue-specific *LMNA* mutation has confirmed this fact and determined that disruption of lamin–chromatin bonds occurs in regions with specific characteristics. Using three cell types such as cardiomyocytes (iPS-CMs), adipocytes (iPS-adips), and hepatocytes (iPS-heps), obtained from iPSC with one of the two cardiac-specific *LMNA* mutations (T10I and R541C), it has been

determined that LADs have cell-specific organization. Moreover, cardiac-specific *LMNA* mutations have a more destructive effect on iPS-CMs compared with iPS-adips and iPS-heps (Shah et al., 2021).

During mitosis, dividing cells undergo some nuclear events, including release of transcription factors and chromatin reorganization accompanied by rearrangement of LADs. Interestingly, these cell-type– specific changes could be reconstructed after mitosis (Shevelyov and Ulianov, 2019). The molecular mechanisms of maintenance of cell-specific orientation of LADs remain unknown. Also, it is not fully clear how chromatin is attached to nuclear lamina. Several studies aimed on identification of relations of nuclear lamins with genome in different cells' types throughout the cell cycle. For instance, Kind et al. (2013) investigated the dynamics of interaction of LADs with nuclear lamina using DNA adenine methylation to visualize and track LADs in single human cells. Early after mitosis a stochastic character of the LADs-nuclear lamina contacts was identified and this was associated with gene repression and positively correlated with H3K9me2 histone

modification (Kind et al., 2013). In another report modified DamID protocol was used to map the interaction of nuclear lamins with the genome in single human cells. Gene-poor LADs contacting with the genome constitutively in case of different cells' types were identified. In addition, there are LADs with more variable lamin-genome interaction, that are cell-type specific. Furthermore, constitutive LADs are characterized by low gene activity and heterochromatic histone modification H3K9me (Kind et al., 2015). Recent research has shown significant results about the dynamics of LADs during interphase, in particular at the onset of G1 phase and during DNA replication. Antibody-based variant of the DamID technology (pA-DamID) allowed to map and visualize nuclear lamina–genome interactions with high temporal resolution. Obtained results showed that after mitosis lamins-genome contacts are widespread on distal regions of chromosomes. Small LADs appear to be gradually displaced from the nuclear lamina by larger LADs. In addition, lamins contacts are increased during DNA replication (van Schaik et al., 2020). Although all these findings are important for understanding principles of the spatial LADs dynamic throughout the cell cycle, yet it remains unclear how this spatial LADs architecture is brought about and which other players are involved.

Multiple interactions of the genome and lamina along large LAD regions are known to be dependent on histone post-translational modifications. Poleshko et al. (2019) have demonstrated that the H3K9me2 mark takes part in 3D spatial heterochromatin organization at the nuclear periphery, and re-associates with the forming nuclear lamina after mitosis. Besides, H3K27me3 marks, as well as CTCF binding sites, flank LADs, mediating their anchoring to the nuclear envelope (Harr et al., 2015).

In addition, apart from lamin A/C participating in chromatin organization, INM proteins can bind genome regions with nuclear lamina, resulting in gene silencing. So it has been shown that LBR is connected with the histone modification H3K9me3 through heterochromatin-binding protein 1 (HP1) (Hirano et al., 2012). Emerin is able to interact with HDAC3 by initiating its catalytic activity (Demmerle et al., 2012). The LAP2β protein plays a critical role in genome organization, gene expression and differentiation process via interaction with the ATP-dependent chromatin remodeling complex BAF (mammalian SWI/SNF complex) (Margalit et al., 2007). There are more examples of the involvement of INM proteins in the regulation of chromatin architecture which can be found in previous reviews (Cai et al., 2001; Zuleger et al., 2011).

The processes of maintaining stem cells in a pluripotent state, as well as their decision to differentiate in a certain direction, are under regulation via complex intracellular programs. These programs can be realized throughout changes of the activity of transcription factors, chromatin organization reconstitutions, epigenetic regulator activity, and many other events. In this regard, it is worth noting the exclusive role of lamin A/C as a part of chromatin organization and regulation of differentiation-related gene expression, resulting in the cell's choice of further fate and specification of an identity. During cell differentiation, spatial relocation of genomic regions toward or away from lamina occurs, as is shown in **Figure 7**. Thus, genes non-relevant to

differentiation interact with lamina and become silent. At the same time, differentiation-related genes unattached from lamina are available for their expression, facilitating the development of a particular cell identity (Bitman-Lotan and Orian, 2021).

The active role of lamin A/C in stem cell identity and cell differentiation has been investigated in several studies. For example, during myogenesis some genes move in and out of LADs in a specific way, leading to changes in their expression state. Some of these genes encode NETs (see above). This tissue-specific NET expression is significant for selective chromosomes docking near the nuclear periphery (Robson et al., 2016). Our group has shown an impact of various *LMNA* mutations on unique expression pattern of genes during MSC differentiation (Malashicheva et al., 2015). Another study of our group has shown that tissue-specific R482L *LMNA* mutation interferes with the differentiation of human cardiac mesenchymal cells into adipocytes when the Notch pathway is activated (Perepelina et al., 2018). Using lentiviral introduction of *LMNA* mutations we have also shown that A-type lamins participate in driving the osteogenic phenotype of four cell lines of mesenchymal origin in a different way (Perepelina et al., 2019).

Besides the functions described above, A-type lamins bind to the retinoblastoma protein pRb, one of the main cell cycle regulators, and are also involved in the regulation of apoptosis and in the processes of muscle and adipogenic differentiation (Boban et al., 2010; Kennedy and Pennypacker, 2014). Lamin A/C involvement in cell differentiation is also confirmed by the direct interaction of lamin A/C with cyclin D3 in muscle cells as well as with SREBP1, an important factor of adipogenic differentiation, in pre-adipocytes (Mariappan et al., 2007). The complex of lamin A/C and emerin could also interact with α-catenin and thereby determine the onset of adipogenesis (Boban et al., 2010). In addition, A-type lamins retain factor c-Fos at the nuclear periphery, which leads to the repression of transcriptional activity of AP-1 factor, a well-known regulator of cell proliferation, differentiation, and apoptosis (Mirza et al., 2021). Thus, A-type lamins are associated with many transcriptional regulators in the nucleus and can influence gene expression by binding to these factors or by affecting the basic transcriptional complexes assembly.

The C-terminal immunoglobulin-like domain of lamin A/C directly interacts with the PCNA replication factor, which plays an important role in DNA replication (Shumaker et al., 2008; Cobb et al., 2016). In natural conditions, lamin A/C expression leads to inhibition of PCNA and dephosphorylation of Rb, which consequently inactivates transcription factors of the E2F group. This leads to the arrest of the cell cycle, suppression of DNA replication, and initiation of the differentiation process. Impaired lamin A/C expression could lead to phosphorylation of Rb by the cyclin D - cdk4/6 complex and the release of transcription factor E2F. As a result, cells do not proceed to the process of differentiation, and the apoptotic mechanisms are activated (Chen et al., 2019).

Despite many discoveries regarding the role of lamin A/C in the regulation of gene expression and chromatin organization, there is still no clear understanding of all the molecular participants in these processes. Given the complexity and

FIGURE 7 | Chromatin reorganization resulting in change of gene expression profile of a cell during differentiation. Lamina-associated domains (LADs) attached to nuclear lamina are transcriptionally inactive, whereas LADs unattached to lamina are available for transcription. The set of LADs is unique for certain differentiated cells and different from undifferentiated cells. Here two hypothetical differentiated cell types are shown as "Differentiated cell type A" and "Differentiated cell type B" having different LADs with transcriptionally active (red) and inactive (blue) genes. Created with BioRender.com.

distinction of each specific cell type mechanism of differentiation regulation, further studies are needed on the development mechanism of severe hereditary diseases associated with impaired tissue differentiation—laminopathies.

Lamin A/C Cooperates With Signaling Pathways During Cell Differentiation

Aside from the lamin A/C functions discussed above, they are capable of modulating the activity of signaling molecules *via* their interaction with gene regulators, promoters, and the other components of signaling cascades in the cells. Intermolecular interactions of lamin A/C with plenty of molecular signaling components or their intermediates occur due to different post-translational modifications that lamin A/C may undergo (Maraldi et al., 2010; Gerace and Tapia, 2018). As a whole, post-translational modifications of lamin A/C can be subdivided into phosphorylation, sumoylation, farnesylation, and carboxymethylation. However, the influence of these modifications on lamin A/C cooperation mechanisms with other molecules and proteins remains largely unknown (Andrés and González, 2009; Gerbino et al., 2018).

Wnt/β-Catenin

The Wnt/β-catenin signaling pathway plays a decisive role in the differentiation of various cells *via* regulation of

the genes involved in mesenchymal tissue proliferation and differentiation. It has been shown that β-catenin (intracellular signal transducer in the Wnt/β-catenin signaling) is capable of interacting with lamin-binding protein emerin, thereby controlling the expression level of emerin in differentiated cells. Inhibition of GSK3 kinase, an important step in β-catenin activation, is required for adipogenic lineage differentiation. In contrast, GSK3-kinase activation leads to differentiation of stem cells toward the osteogenic lineage (Maraldi et al., 2011). Using knockout mice (*Lmna* −/−), Tong et al. (2011) have shown that the absence of lamin A/C synthesis leads to suppression of myogenic and osteogenic cell differentiation, which correlates with an increase of adipose tissue content and with expression of adipogenic markers, as well as with decreased activity of the Wnt/β-catenin signaling pathway. The implication of Wnt/β-catenin signaling in osteogenic differentiation promotion of MSCs was confirmed in several studies (Tong et al., 2011; Wang et al., 2017), whereas adipogenic and chondrogenic direction of differentiation was suppressed when Wnt/β-catenin was activated (Case and Rubin, 2010; Ullah et al., 2015).

Notch Pathway

Notch signaling is a key regulator of main cellular processes including proliferation, differentiation, and apoptosis in both

the adult organism and the developing embryo (Schwanbeck et al., 2011; Hori et al., 2013). The Notch pathway includes four Notch receptors (Notch1, Notch2, Notch3, Notch4), five ligands (Jag-1, Jag-2, DLL1, DLL3, DLL4), and gene regulators. Receptors and ligands are mainly transmembrane forms of proteins that ensure the interaction of neighboring cells with each other. Notch receptors undergo sequential proteolytic cleavages upon binding of their ligand, resulting in the release of Notch intracellular domain (NICD) from the cellular membrane. NICD is translocated into the nucleus, where it interacts with transcription factors, thereby activating expression of target genes (Andersson et al., 2011; Henrique and Schweisguth, 2019).

Notch is established to regulate the cell differentiation process (Bray, 2006). Moreover, the involvement of Notch signaling in Hutchinson-Gilford progeria syndrome (HGPS) has been shown (Pereira et al., 2008). HGPS is associated with expression of a truncated form of prelamin A called progerin, whose accumulation mainly leads to abnormal nuclear shape and chromatin structure. Thus, mostly mesenchymal tissues are thought to be damaged. Scaffidi and Misteli (2008) showed that the expression of progerin in human MSCs causes hyperactivation of the main targets of the Notch signaling pathway—HEY1 and HES1 (2008). This contributes to a change in the expression of differentiation markers: enhanced adipogenic and reduced osteogenic ones. However, changes in the chondrogenic differentiation in the cells carrying the mutation, in contrast to the wild-type LMNA, were not observed. As a possible mechanism, it has been suggested that the presence of progerin causes a disruption of a connection of lamin A/C with the transcription factor SKIP, an activator of genes of the Notch family, thereby increasing Notch-related gene expression inside the nucleus. In addition, Notch genes probably can directly interact with the nuclear lamina, and their regulation is associated with epigenetic modifications (Scaffidi and Misteli, 2008).

The impact of various LMNA mutations on the Notch pathway during differentiation of the cells of various mesenchymal origin has been reported. In our previous work, we proposed that the cooperation of lamin A/C with Notch signaling could be one of the mechanisms regulating MSC differentiation, based on the facts that tissue-specific LMNA mutations are able to influence the Notch signaling activity in MSCs (Bogdanova et al., 2014). Involvement of Notch signaling in adipogenic and osteogenic differentiation has been analyzed in another study by our group. A specific LMNA mutation (R482L), associated with Dunningan-type familial partial lipodystrophy, contributes to the impairment of adipogenic differentiation when Notch signaling is activated (Perepelina et al., 2018). One more study has revealed the opposite effect of R527C LMNA mutation associated with osteogenic phenotype of laminopathy on the expression level of RUNX2 (a master gene of osteogenic differentiation) during osteogenic differentiation of mesenchymal cells, such as human cardiac mesenchymal cells and human aortic valve interstitial cells upon Notch activation. These results confirmed the fact of interaction of lamin A/C with Notch signaling (Perepelina et al., 2019). Thus, specific mutations in the LMNA gene are implicated in functional changes of Notch signaling during cell differentiation.

TGF-β/Smad Pathway

There is considerable evidence that the TGF-β/Smad pathway is involved in bone abnormalities via contravention of the osteogenic differentiation process. Smad2 is known to interact with lamin-binding protein MAN1. Kondé et al. (2010) described in more detail this interaction via structural analysis, and revealed a UHM domain of MAN1 participating with Smad2-MAN1 link. Heterozygous loss-of-function mutation in the MAN1 gene leads to bone abnormalities in humans, such as osteopoikilosis (sclerotic bone lesions) with or without manifestations of Buschke-Ollendorff syndrome, and melorheostosis (aberrant growth of new bone tissue on the surface of existing bones). These abnormal changes lead to increasing bone density and overexpression of TGF-b (Hellemans et al., 2004). It has been shown that MAN1 could be implicated in inactivation through competition with transcription factors for binding to Smad2 and Smad3, and it contributes to their dephosphorylation by phosphatase PPM1A (Bourgeois et al., 2013). In addition, lamin A/C can impact TGF-β/Smad signaling activity via interplay with protein phosphatase 2A (Van Berlo et al., 2005). To understand how A-type lamins facilitate functional changes of TGF-β/Smad pathway, further research is obviously needed.

Mitogen-Activated Protein Kinase Pathway

The mitogen-activated protein kinase (MAPK) pathway regulates the cell cycle and differentiation process (Maraldi et al., 2011).

A-type lamins mediate retaining c-Fos (transcription factor that regulates key cellular processes, including differentiation) at the periphery of the nucleus. Cooperation of lamin A/C with c-Fos factor could be disrupted due to phosphorylation of c-Fos by MAPK Erk. This result suggests the participation of lamin A/C in MAPK pathway activity (Gonzàlez et al., 2008). In knockout mouse models of dilated cardiomyopathy with the LMNA H222P mutation in response to mechanical stress in cardiomyocytes, activation of the MAPK signaling pathway was observed, in which kinases such as ERK1/2 and JNK were involved. In addition, inhibitors of this signaling pathway were found to prevent the development of cardiomyopathies associated with a mutation in the LMNA gene, but did not affect the development of muscular dystrophy (Muchir et al., 2007).

Thus, A-type lamins are associated with many signaling pathways and transcriptional regulators in the nucleus and could influence gene expression by binding to these factors or by affecting the assembly of basic transcriptional complexes.

LAMINOPATHIES AS A CONSEQUENCE OF MUTATIONS IN THE LMNA GENE

Laminopathies are a group of hereditary diseases caused by mutations in genes encoding (a) nuclear lamins; (b) proteins associated with post-translational modifications of lamins (such as ZMPSTE24); (c) proteins that interact with lamins (emerin, LAP2, LBR, MAN1, nesprins), and (d) proteins that make up nuclear pores (Zaremba-Czogalla et al., 2011).

Over the past 20 years, it has been found that most laminopathies are caused by mutations in the *LMNA* gene, which encodes lamin A/C. To date, over 15 different diseases have been described, associated with 498 mutations in the *LMNA* gene as reported by UMD-LMNA, universal mutations database.[1] Laminopathies are characterized by a wide range of clinical phenotypes, in which one type of tissue is most often affected, mainly of mesenchymal origin, for example, lipodystrophy (damage to adipose tissue), mandibuloacral dysplasia (damage to bone tissue), cardiomyopathy and muscular dystrophy (damage of the heart and skeletal muscles) (Rankin and Ellard, 2006). There are some groups of laminopathies in which different tissues are affected, resulting in overlapping or systemic phenotypes (Bertrand et al., 2011; Zaremba-Czogalla et al., 2011; Crasto and Di Pasquale, 2018; **Figure 8**).

Premature aging syndrome, also known as progeria, is one of the best-studied human diseases with overlapping phenotypes, in which several tissues are affected. The pathology is caused by mutations in the ZMPSTE24 gene, mutations in the *LMNA* gene, as well as by mutations in genes encoding DNA repair proteins, such as in RecQ protein-like helicases (RECQLs) and nuclear excision repair (NER) proteins and others (Navarro et al., 2006). The most famous form of progeria is Hutchinson-Guildford syndrome (Rankin and Ellard, 2006; Worman et al., 2010). This is an extremely rare autosomal dominant disease, a childhood form of progeria, characterized by changes in the skin and internal organs caused by premature aging of the body. In 2003, the mechanism of this disease development was described. A mutation in the *LMNA* gene causes the substitution of cytosine with thymine amino acid, thus forming an additional splice site in exon 11, resulting in a truncated mRNA of *LMNA* transcript. In the process of translation, an altered form of prelamin A is synthesized, in which the CaaX motif is not cleaved, and instead of the mature lamin A, the progerin protein is formed, which cannot be incorporated into the nuclear lamina resulting in disruption the scaffold of the nucleus (Gonzalo et al., 2017).

Unlike HGPS, for other diseases associated with mutations in the *LMNA* gene, the molecular mechanisms of pathogenesis are still poorly understood. Most mutations in the *LMNA* gene affect the heart or skeletal muscles. Among such diseases, Emery-Dreifuss autosomal dominant and recessive forms of muscular dystrophy (EDMD) could be distinguished (Worman, 2012). The disease was found to be associated with the R453W point mutation of the *LMNA* gene mapped to locus lq 21.2–21.3 (Favreau et al., 2004). Later, missense mutations were found, for example, G232E, Q294P, and R386K, leading to the development of EDMD (Muchir and Worman, 2007). Other diseases of the heart and skeletal muscles associated with mutations in the *LMNA* gene were soon described: dilated cardiomyopathy 1A (Fatkin et al., 1999) and limb-girdle progressive muscular dystrophy 1B (Muchir et al., 2000). EDMD, isolated dilated cardiomyopathy, and limb-girdle muscular dystrophy are characterized by overlapping clinical phenotypes and dilated cardiomyopathy associated with cardiac conduction abnormalities (Cattin et al., 2013).

[1]http://www.umd.be/LMNA/

Dunnigan-type familial partial lipodystrophy, also known as FPLD, is an autosomal dominant disorder characterized by a loss of hypodermic adipose tissue in the limbs and torso after puberty and excess fat deposition in the head and neck region. A total of 90% of the *LMNA* mutations in this syndrome are missense mutations located in exon 8 (Boguslavsky et al., 2006). Several such mutations have been described, for example, R482Q, R482W, G465D in exon 8, and R582H in exon 11 of the *LMNA* gene (Garg et al., 2001).

Mandibuloacral dysplasia (MAD) is a rare autosomal recessive disorder characterized by post-natal bone anomalies. MAD occurs due to point *LMNA* mutations associated with amino acid substitutions (Garg et al., 2005). Mandibuloacral dysplasia could also be caused by mutations in the ZMPSTE24 protease, involved in the processing of prelamin A to lamin A/C (Agarwal et al., 2003).

Thus, these few main examples of laminopathies demonstrate that mutations in the same *LMNA* gene could lead to the development of severe abnormalities, characterized by a wide range of clinical tissue-specific phenotypes. However, the mechanism of development of these diseases is still not fully understood.

Several years ago, scientists proposed two hypotheses explaining the development of laminopathies: a structural hypothesis and a gene expression hypothesis. According to the structural hypothesis, mutations in the *LMNA* gene cause, first of all, weakening of the nuclear membrane, which makes it vulnerable to damage resulting in cell death and a replacement of differentiated tissue in specific cells. Another hypothesis is based on molecular mechanisms, and is related to the fact that A-type lamins are regulators of gene expression of some proteins, and mutations in the *LMNA* genes, therefore, disrupt their regulatory capacity and contribute to the disease development (Osmanagic-Myers and Foisner, 2019). Currently, there is evidence for both hypotheses. However, it is interesting that cluster analysis of *LMNA* mutations gives preference to one or another hypothesis depending on the localization of *LMNA* mutations associated with a particular type of laminopathies. Thus, it has been shown that mutations in the *LMNA* gene located upstream of the nuclear localization signal (NLS) affect the conserved core domain necessary for the formation and maintenance of the integrity of the nuclear cytoskeleton, while mutations located downstream interact more closely with chromatin and transcription factors (Hegele, 2005). Since the first group of mutations is mainly associated with a large group of muscular dystrophies and cardiomyopathies, scientists suggested that the causes of these diseases are, first of all, a violation of the formation of the lamina structure and mechanical defects. The second group of mutations belongs to other types of laminopathies—in particular, to progeroid syndromes, FPLD and MAD— and is most likely associated with disturbances in the interaction and regulation of important signaling pathways in the cell (Cattin et al., 2013; **Figure 9**).

Recently, research has mainly focused on the study of the molecular mechanisms of the development of laminopathies (Osmanagic-Myers and Foisner, 2019; Alcorta-Sevillano et al., 2020; Shah et al., 2021). The molecular mechanisms proposed

FIGURE 8 | Phenotypic spectrum of laminopathies and the different tissues affected. Created with BioRender.com.

FIGURE 9 | Scheme of lamin A and lamin C structures with mutated regions. *LMNA* mutations upstream of the nuclear localization signal (NLS) affect the conserved core domain. Resulting in striated muscular dystrophies and cardiomyopathies. *LMNA* mutations located downstream of NLS lead to progeroid syndromes, Dunnigan-type familial partial lipodystrophy (FPLD) and mandibuloacral dysplasia (MAD). Created with BioRender.com.

by scientists include disturbances in the organization of heterochromatin, intracellular signal transduction, and in the process of autophagy, which ultimately leads to the regulation of the expression of various genes (Wong and Stewart, 2020). It is clear that iPSC-derived patient-specific models could greatly contribute to the understaning of laminopathies. Particularly, we recently generated several iPSC lines carrying disease-specific mutations in the *LMNA* gene. These iPSC lines would be a useful tool to investigate disease development of laminopathies (Perepelina et al., 2019, 2020a,b; Klauzen et al., 2020).

CONCLUSION

Nuclear A-type lamins play a critical role in vital cell functions including migration, growth, homeostasis, proliferation, differentiation, and many others. Despite the fact that a broad range of studies are devoted to investigating the role of lamin A/C in the cells, all their functions are still incompletely understood and demand further investigations. A-type lamins undoubtedly play critical role in cell differentiation and determination of their identity.

To date, a large amount of data has been collected that is consistent with the hypothesis that A-type lamins define cell identity *via* the organization of chromatin architecture, epigenetic regulation and expression of differentiation related genes. Apparently, during cell differentiation, A-type lamins perform a key role in directing stem cells to a proper differentiating state. It is supposed and, for some cases experimentally shown, that for this, pluripotent or inappropriate differentiation-related genes are included in the LADs, resulting in inactivation of their expression. On the other hand, genes required for specific differentiation are activated *via* detachment of genome regions from the nuclear lamina. These genome reorganizations are cell cycle stage-dependent, and vary in the different cells from different origins (Bitman-Lotan and Orian, 2021; Shah et al., 2021).

Laminopathies are known to be associated with cell differentiation impairment. The failure of lamin A/C folding in the right way due to *LMNA* mutations leads to the disruption of integrity and function of the nuclear lamina. As a result, the interaction of chromatin with lamins can be disrupted, which leads to the essential cell and genome functions breaking, including transcription repression of the genes responsible for differentiation, and, ultimately, to the development of the disease (Scaffidi and Misteli, 2008; Melcer and Meshorer, 2010; Zhang et al., 2019). For a better understanding of the mechanisms leading to the disruption of cell differentiation, it is important to study the totality of possible changes in the cell that occur as a result of mutations in the *LMNA* gene. Currently, the most relevant studies are devoted to the epigenetic mechanisms of laminopathy development. Post-translational modifications of histones, as well as A-type lamins, are key mechanisms directing chromatin attachment to the lamina (Scaffidi and Misteli, 2008; Collas, 2010; Melcer and Meshorer, 2010; Poleshko et al., 2019; Shevelyov and Ulianov, 2019; Zhang et al., 2019). However, currently it is not completely understood what could be other mechanisms providing LADs forming.

Numerous studies have highlighted the importance of mechanical stimuli in shaping cell and tissue function including the cell differentiation process (Bonnans et al., 2014; Makarov et al., 2019; Donnaloja et al., 2020). Mechanotransduction mediates the link between ECM and components of the cell, including cytoskeleton members and A-type lamins. Their relations make it possible to transform internal stimuli into biological response (Bonnans et al., 2014; Osmanagic-Myers et al., 2015; Heo et al., 2018). Although knowledge about regulation of cytoskeleton dynamics has already been obtained, we lack a general idea about the mechanisms by which tension, propagated through the cytoskeleton, regulates mechanical signal transition. Moreover, the complete list of members involved in the mechanosignaling process is not known.

Multiple examples demonstrate that signaling pathways are implicated in relations with A-type lamins, resulting in participation in differentiating process (Parnaik, 2008; Andrés and González, 2009; Maraldi et al., 2010, 2011; Gerbino et al., 2018). Here we have discussed the main signaling pathways taking part in lamin-mediated regulation of cell fate determination, but it is not a complete list of them. An important goal for future research will be to elaborate strategies aimed at unraveling the interactions among different signaling pathways and nuclear lamins. The study of particular cell lines and cellular pathways and a detailed description of an individual molecular portrait of each cell line should explain in more details some particular features of lamin A/C action in directing cellular differentiation.

Increasing our understanding of the functional consequences of mutations in the *LMNA* gene is certainly important. For this, novel investigations using a broad range of cell types and origins are needed, since this indeed could reflect the tissue-specific features of laminopathies. Finally, a further challenge for the future will be to apply existing knowledge to the therapy of severe hereditary diseases—laminopathies.

AUTHOR CONTRIBUTIONS

AM supervised the work, analyzed literature data, discussed and wrote the manuscript, and acquired funding. KP analyzed literature data, designed the figures, discussed, and drafted the manuscript. Both authors contributed to the article and approved the submitted version.

FUNDING

REFERENCES

Adam, S. A., Butin-Israeli, V., Cleland, M. M., Shimi, T., and Goldman, R. D. (2013). Disruption of lamin B1 and lamin B2 processing and localization by farnesyltransferase inhibitors. *Nucleus* 4, 142–150. doi: 10.4161/nucl.24089

Agarwal, A. K., Fryns, J. P., Auchus, R. J., and Garg, A. (2003). Zinc metalloproteinase ZMPSTE24, is mutated in mandibuloacral dysplasia. *Hum. Mol. Genet.* 12, 1995–2001. doi: 10.1093/hmg/ddg213

Ahn, J., Jo, I., Kang, S. M., Hong, S., Kim, S., Jeong, S., et al. (2019). Structural basis for lamin assembly at the molecular level. *Nat. Commun.* 10:3757. doi: 10.1038/s41467-019-11684-x

Alcorta-Sevillano, N., Macías, I., Rodríguez, C. I., and Infante, A. (2020). Crucial Role of Lamin A/C in the Migration and Differentiation of MSCs in Bone. *Cells* 9:1330. doi: 10.3390/cells9061330

Almendáriz-Palacios, C., Gillespie, Z. E., Janzen, M., Martinez, V., Bridger, J. M., Harkness, T. A. A., et al. (2020). The nuclear lamina: protein accumulation and disease. *Biomedicines* 8:188. doi: 10.3390/BIOMEDICINES8070188

Andersson, E. R., Sandberg, R., and Lendahl, U. (2011). Notch signaling: simplicity in design, versatility in function. *Development* 138, 3593–3612. doi: 10.1242/dev.063610

Andrés, V., and González, J. M. (2009). Role of A-type lamins in signaling, transcription, and chromatin organization. *J. Cell Biol.* 187, 945–957. doi: 10.1083/jcb.200904124

Arimura, T., Helbling-Leclerc, A., Massart, C., Varnous, S., Niel, F., Lacène, E., et al. (2005). Mouse model carrying H222P-Lmna mutation develops muscular dystrophy and dilated cardiomyopathy similar to human striated muscle laminopathies. *Hum. Mol. Genet.* 14, 155–169. doi: 10.1093/hmg/ddi017

Avnet, S., Pallotta, R., Perut, F., Baldini, N., Pittis, M. G., Saponari, A., et al. (2011). Osteoblasts from a mandibuloacral dysplasia patient induce human blood precursors to differentiate into active osteoclasts. *Biochim. Biophys. Acta Mol. Basis Dis.* 1812, 711–718. doi: 10.1016/j.bbadis.2011.03.006

Bertrand, A. T., Chikhaoui, K., Yaou, R. B., and Bonne, G. (2011). Clinical and genetic heterogeneity in laminopathies. *Biochem. Soc. Trans.* 39, 1687–1692. doi: 10.1042/BST20110670

Bitman-Lotan, E., and Orian, A. (2021). Nuclear organization and regulation of the differentiated state. *Cell. Mol. Life Sci.* 78, 3141–3158. doi: 10.1007/s00018-020-03731-4

Boban, M., Braun, J., and Foisner, R. (2010). Lamins: "Structure goes cycling.". *Biochem. Soc. Trans.* 38, 301–306. doi: 10.1042/BST0380301

Bogdanova, M. A., Gudkova, A., Zabirnik, A. S., Ignat'eva, E. V., Dmitrieva, R. I., Smolina, N. A., et al. (2014). Nuclear lamins regulate osteogenic differentiation of mesenchymal stem cells. *Tsitologiia* 56, 260–267.

Boguslavsky, R. L., Stewart, C. L., and Worman, H. J. (2006). Nuclear lamin A inhibits adipocyte differentiation: implications for Dunnigan-type familial partial lipodystrophy. *Hum. Mol. Genet.* 15, 653–663. doi: 10.1093/hmg/ddi480

Bonnans, C., Chou, J., and Werb, Z. (2014). Remodelling the extracellular matrix in development and disease. *Nat. Rev. Mol. Cell Biol.* 15, 786–801. doi: 10.1038/nrm3904

Bourgeois, B., Gilquin, B., Tellier-Lebègue, C., Östlund, C., Wu, W., Pérez, J., et al. (2013). Inhibition of TGF-β signaling at the nuclear envelope: characterization of interactions between MAN1, Smad2 and Smad3, and PPM1A. *Sci. Signal.* 6:ra49. doi: 10.1126/scisignal.2003411

Brachner, A., and Foisner, R. (2011). Evolvement of LEM proteins as chromatin tethers at the nuclear periphery. *Biochem. Soc. Trans.* 39, 1735–1741. doi: 10.1042/BST20110724

Bray, S. J. (2006). Notch signalling: a simple pathway becomes complex. *Nat. Rev. Mol. Cell Biol.* 7, 678–689. doi: 10.1038/nrm2009

Briand, N., and Collas, P. (2020). Lamina-associated domains: peripheral matters and internal affairs. *Genome Biol.* 21:85. doi: 10.1186/s13059-020-02003-5

Burke, B., and Stewart, C. L. (2013). The nuclear lamins: flexibility in function. *Nat. Rev. Mol. Cell Biol.* 14, 13–24. doi: 10.1038/nrm3488

Cai, M., Huang, Y., Ghirlando, R., Wilson, K. L., Craigie, R., and Clore, G. M. (2001). Solution structure of the constant region of nuclear envelope protein LAP2 reveals two LEM-domain structures: one binds BAF and the other binds DNA. *EMBO J.* 20, 4399–4407. doi: 10.1093/emboj/20.16.4399

Case, N., and Rubin, J. (2010). β-catenin - A supporting role in the skeleton. *J. Cell. Biochem.* 110, 545–553. doi: 10.1002/jcb.22574

Cattin, M. E., Muchir, A., and Bonne, G. (2013). "State-of-the-heart" of cardiac laminopathies. *Curr. Opin. Cardiol.* 28, 297–304. doi: 10.1097/HCO.0b013e32835f0c79

Chaffee, B. R., Shang, F., Chang, M. L., Clement, T. M., Eddy, E. M., Wagner, B. D., et al. (2014). Nuclear removal during terminal lens fiber cell differentiation requires CDK1 activity: appropriating mitosis-related nuclear disassembly. *Development* 141, 3388–3398. doi: 10.1242/dev.106005

Chen, S. N., Lombardi, R., Karmouch, J., Tsai, J. Y., Czernuszewicz, G., Taylor, M. R. G., et al. (2019). DNA Damage Response/TP53 Pathway Is Activated and Contributes to the Pathogenesis of Dilated Cardiomyopathy Associated with LMNA (Lamin A/C) Mutations. *Circ. Res.* 124, 856–873. doi: 10.1161/CIRCRESAHA.118.314238

Chin, I. L., Hool, L., and Choi, Y. S. (2019). A review of in vitro platforms for understanding cardiomyocyte Mechanobiology. *Front. Bioeng. Biotechnol.* 7:133. doi: 10.3389/fbioe.2019.00133

Cobb, A. M., Murray, T. V., Warren, D. T., Liu, Y., and Shanahan, C. M. (2016). Disruption of PCNA-lamins A/C interactions by prelamin A induces DNA replication fork stalling. *Nucleus* 7, 498–511. doi: 10.1080/19491034.2016.1239685

Collas, P. (2010). The current state of chromatin immunoprecipitation. *Mol. Biotechnol.* 45, 87–100. doi: 10.1007/s12033-009-9239-8

Corrigan, D. P., Kuszczak, D., Rusinol, A. E., Thewke, D. P., Hrycyna, C. A., Michaelis, S., et al. (2005). Prelamin A endoproteolytic processing in vitro by recombinant Zmpste24. *Biochem. J.* 387, 129–138. doi: 10.1042/BJ20041359

Crasto, S., and Di Pasquale, E. (2018). Induced pluripotent stem cells to study mechanisms of laminopathies: focus on epigenetics. *Front. Cell Dev. Biol.* 6:172. doi: 10.3389/fcell.2018.00172

Cremer, C., Szczurek, A., Schock, F., Gourram, A., and Birk, U. (2017). Super-resolution microscopy approaches to nuclear nanostructure imaging. *Methods* 123, 11–32. doi: 10.1016/j.ymeth.2017.03.019

Cremer, T., and Cremer, M. (2010). Chromosome territories. *Cold Spring Harb. Perspect. Biol.* 2:a003889. doi: 10.1101/cshperspect.a003889

de Leeuw, R., Gruenbaum, Y., and Medalia, O. (2018). Nuclear Lamins: thin Filaments with Major Functions. *Trends Cell Biol.* 28, 34–45. doi: 10.1016/j.tcb.2017.08.004

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* 295, 1306–1311. doi: 10.1126/science.1067799

Demmerle, J., Koch, A. J., and Holaska, J. M. (2012). The nuclear envelope protein emerin binds directly to histone deacetylase 3 (HDAC3) and activates HDAC3 activity. *J. Biol. Chem.* 287, 22080–22088. doi: 10.1074/jbc.M111.325308

Donnaloja, F., Carnevali, F., Jacchetti, E., and Raimondi, M. T. (2020). Lamin A/C Mechanotransduction in Laminopathies. *Cells* 9:1306. doi: 10.3390/cells9051306

Engler, A. J., Sen, S., Sweeney, H. L., and Discher, D. E. (2006). Matrix Elasticity Directs Stem Cell Lineage Specification. *Cell* 126, 677–689. doi: 10.1016/j.cell.2006.06.044

Enyedi, B., and Niethammer, P. (2017). Nuclear membrane stretch and its role in mechanotransduction. *Nucleus* 8, 156–161. doi: 10.1080/19491034.2016.1263411

Fatkin, D., MacRae, C., Sasaki, T., Wolff, M. R., Porcu, M., Frenneaux, M., et al. (1999). Missense Mutations in the Rod Domain of the Lamin a/C Gene as causes of diated cardiomyopathy and conduction-system disease. *N. Engl. J. Med.* 341, 1715–1724.

Favreau, C., Higuet, D., Courvalin, J.-C., and Buendia, B. (2004). Expression of a Mutant Lamin A That Causes Emery-Dreifuss Muscular Dystrophy Inhibits In Vitro Differentiation of C2C12 Myoblasts. *Mol. Cell. Biol.* 24, 1481–1492. doi: 10.1128/mcb.24.4.1481-1492.2004

Fisher, D. Z., Chaudhary, N., and Blobel, G. (1986). cDNA sequencing of nuclear lamins A and C reveals primary and secondary structural homology to intermediate filament proteins. *Proc. Natl. Acad. Sci. U. S. A.* 83, 6450–6454. doi: 10.1073/pnas.83.17.6450

Garg, A., Cogulu, O., Ozkinay, F., Onay, H., and Agarwal, A. K. (2005). A novel homozygous Ala529Val LMNA mutation in Turkish patients with mandibuloacral dysplasia. *J. Clin. Endocrinol. Metab.* 90, 5259–5264. doi: 10.1210/jc.2004-2560

Garg, A., Vinaitheerthan, M., Weatherall, P. T., and Bowcock, A. M. (2001). Phenotypic heterogeneity in patients with familial partial lipodystrophy (Dunnigan variety) related to the site of missense mutations in lamin A/C gene. *J. Clin. Endocrinol. Metab.* 86, 59–65. doi: 10.1210/jc.86.1.59

Gerace, L., and Tapia, O. (2018). Messages from the voices within: regulation of signaling by proteins of the nuclear lamina. *Curr. Opin. Cell Biol.* 52, 14–21. doi: 10.1016/j.ceb.2017.12.009

Gerbino, A., Procino, G., Svelto, M., and Carmosino, M. (2018). Role of Lamin A/C Gene Mutations in the signaling defects leading to cardiomyopathies. *Front. Physiol.* 9:1356. doi: 10.3389/fphys.2018.01356

Giacomelli, E., Meraviglia, V., Campostrini, G., Cochrane, A., Cao, X., van Helden, R. W. J., et al. (2020). Human-iPSC-Derived Cardiac Stromal Cells Enhance Maturation in 3D Cardiac Microtissues and Reveal Non-cardiomyocyte Contributions to Heart Disease. *Cell Stem Cell* 26, 862–879.e11. doi: 10.1016/j.stem.2020.05.004

Gonzàlez, J. M., Navarro-Puche, A., Casar, B., Crespo, P., and Andrès, V. (2008). Fast regulation of AP-1 activity through interaction of lamin A/C, ERK1/2, and c-Fos at the nuclear envelope. *J. Cell Biol.* 183, 653–666. doi: 10.1083/jcb.200805049

Gonzalo, S., Kreienkamp, R., and Askjaer, P. (2017). Hutchinson-Gilford Progeria Syndrome: a premature aging disease caused by LMNA gene mutations. *Ageing Res. Rev.* 33, 18–29. doi: 10.1016/j.arr.2016.06.007

Gruenbaum, Y., and Foisner, R. (2015). Lamins: nuclear intermediate filament proteins with fundamental functions in nuclear mechanics and genome regulation. *Annu. Rev. Biochem.* 84, 131–164. doi: 10.1146/annurev-biochem-060614-034115

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M. B., Talhout, W., et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 453, 948–951. doi: 10.1038/nature06947

Guilluy, C., Osborne, L. D., Van Landeghem, L., Sharek, L., Superfine, R., Garcia-Mata, R., et al. (2014). Isolated nuclei adapt to force and reveal a mechanotransduction pathway in the nucleus. *Nat. Cell Biol.* 16, 376–381. doi: 10.1038/ncb2927

Hamczyk, M. R., Villa-Bellosta, R., Gonzalo, P., Andrés-Manzano, M. J., Nogales, P., Bentzon, J. F., et al. (2018). Vascular smooth muscle–specific progerin expression accelerates atherosclerosis and death in a mouse model of Hutchinson-Gilford progeria syndrome. *Circulation* 138, 266–282. doi: 10.1161/CIRCULATIONAHA.117.030856

Haque, F., Lloyd, D. J., Smallwood, D. T., Dent, C. L., Shanahan, C. M., Fry, A. M., et al. (2006). SUN1 Interacts with Nuclear Lamin A and Cytoplasmic Nesprins To Provide a Physical Connection between the Nuclear Lamina and the Cytoskeleton. *Mol. Cell. Biol.* 26, 3738–3751. doi: 10.1128/mcb.26.10.3738-3751.2006

Harr, J. C., Luperchio, T. R., Wong, X., Cohen, E., Wheelan, S. J., and Reddy, K. L. (2015). Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins. *J. Cell Biol.* 208, 33–52. doi: 10.1083/jcb.201405110

Hegele, R. A. (2005). LMNA mutation position predicts organ system involvement in laminopathies. *Clin. Genet.* 68, 31–34. doi: 10.1111/j.1399-0004.2005.00447.x

Hellemans, J., Preobrazhenska, O., Willaert, A., Debeer, P., Verdonk, P. C. M., Costa, T., et al. (2004). Loss-of-function mutations in LEMD3 result in osteopoikilosis, Buschke-Ollendorff syndrome and melorheostosis. *Nat. Genet.* 36, 1213–1218. doi: 10.1038/ng1453

Henrique, D., and Schweisguth, F. (2019). Mechanisms of notch signaling: a simple logic deployed in time and space. *Development* 146:dev172148. doi: 10.1242/dev.172148

Heo, S. J., Cosgrove, B. D., Daiv, E. N., and Mauck, R. L. (2018). Mechano-adaptation of the stem cell nucleus. *Nucleus* 9, 9–19. doi: 10.1080/19491034.2017.1371398

Heo, S. J., Driscoll, T. P., Thorpe, S. D., Nerurkar, N. L., Baker, B. M., Yang, M. T., et al. (2016). Differentiation alters stem cell nuclear architecture, mechanics, and mechano-sensitivity. *Elife* 5:e18207. doi: 10.7554/eLife.18207

Herrmann, H., and Aebi, U. (2004). Intermediate Filaments: molecular structure, assembly mechanism, and integration into functionally distinct intracellular scaffolds. *Annu. Rev. Biochem.* 73, 749–789. doi: 10.1146/annurev.biochem.73.011303.073823

Hirano, Y., Hizume, K., Kimura, H., Takeyasu, K., Haraguchi, T., and Hiraoka, Y. (2012). Lamin B receptor recognizes specific modifications of histone H4 in heterochromatin formation. *J. Biol. Chem.* 287, 42654–42663. doi: 10.1074/jbc.M112.397950

Hori, K., Sen, A., and Artavanis-Tsakonas, S. (2013). Notch signaling at a glance. *J. Cell Sci.* 126, 2135–2140. doi: 10.1242/jcs.127308

Ikegami, K., Secchia, S., Almakki, O., Lieb, J. D., and Moskowitz, I. P. (2020). Phosphorylated Lamin A/C in the Nuclear Interior Binds Active Enhancers Associated with Abnormal Transcription in Progeria. *Dev. Cell* 52, 699–713.e11. doi: 10.1016/j.devcel.2020.02.011

Isermann, P., and Lammerding, J. (2013). Nuclear mechanics and mechanotransduction in health and disease. *Curr. Biol.* 23, R1113–R1121. doi: 10.1016/j.cub.2013.11.009

Israeli-Rosenberg, S., Manso, A. M., Okada, H., and Ross, R. S. (2014). Integrins and integrin-associated proteins in the cardiac myocyte. *Circ. Res.* 114, 572–586. doi: 10.1161/CIRCRESAHA.114.301275

Karoutas, A., and Akhtar, A. (2021). Functional mechanisms and abnormalities of the nuclear lamina. *Nat. Cell Biol.* 23, 116–126. doi: 10.1038/s41556-020-00630-5

Kennedy, B. K., and Pennypacker, J. K. (2014). RB and lamins in cell cycle regulation and aging. *Adv. Exp. Med. Biol.* 773, 127–142. doi: 10.1007/978-1-4899-8032-8_6

Kim, W., Bennett, E. J., Huttlin, E. L., Guo, A., Li, J., Possemato, A., et al. (2011). Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol. Cell* 44, 325–340. doi: 10.1016/j.molcel.2011.08.025

Kind, J., Pagie, L., De Vries, S. S., Nahidiazar, L., Dey, S. S., Bienko, M., et al. (2015). Genome-wide Maps of Nuclear Lamina Interactions in Single Human Cells. *Cell* 163, 134–147. doi: 10.1016/j.cell.2015.08.040

Kind, J., Pagie, L., Ortabozkoyun, H., Boyle, S., De Vries, S. S., Janssen, H., et al. (2013). Single-cell dynamics of genome-nuclear lamina interactions. *Cell* 153, 178–192. doi: 10.1016/j.cell.2013.02.028

Klauzen, P., Perepelina, K., Khudiakov, A., Zlotina, A., Fomicheva, Y., Pervunina, T., et al. (2020). Generation of two induced pluripotent stem cell lines (FAMRCi005-A and FAMRCi005-B) from patient carrying genetic variant LMNA p.Asp357Val. *Stem Cell Res.* 43:101719. doi: 10.1016/j.scr.2020.101719

Kochin, V., Shimi, T., Torvaldson, E., Adam, S. A., Goldman, A., Pack, C. G., et al. (2014). Interphase phosphorylation of lamin A. *J. Cell Sci.* 127, 2683–2696. doi: 10.1242/jcs.141820

Kondé, E., Bourgeois, B., Tellier-Lebegue, C., Wu, W., Pérez, J., Caputo, S., et al. (2010). Structural analysis of the Smad2-MAN1 interaction that regulates transforming growth factor-β signaling at the inner nuclear membrane. *Biochemistry* 49, 8020–8032. doi: 10.1021/bi101153w

Korfali, N., Wilkie, G. S., Swanson, S. K., Srsen, V., de las Heras, J., Batrakou, D. G., et al. (2012). The nuclear envelope proteome differs notably between tissues. *Nucleus* 3, 37–41. doi: 10.4161/nucl.22257

Lambert, M. W. (2019). The functional importance of lamins, actin, myosin, spectrin and the LINC complex in DNA repair. *Exp. Biol. Med.* 244, 1382–1406. doi: 10.1177/1535370219876651

Le Dour, C., Macquart, C., Sera, F., Homma, S., Bonne, G., Morrow, J. P., et al. (2017). Decreased WNT/β-catenin signalling contributes to the pathogenesis of dilated cardiomyopathy caused by mutations in the lamin a/C gene. *Hum. Mol. Genet.* 26, 333–343. doi: 10.1093/hmg/ddw389

Lee, Y. L., and Burke, B. (2018). LINC complexes and nuclear positioning. *Semin. Cell Dev. Biol.* 82, 67–76. doi: 10.1016/j.semcdb.2017.11.008

Lieberman-aiden, E., Berkum, N. L., Van Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). *Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–294.

Liu, S. Y., and Ikegami, K. (2020). Nuclear lamin phosphorylation: an emerging role in gene regulation and pathogenesis of laminopathies. *Nucleus* 11, 299–314. doi: 10.1080/19491034.2020.1832734

Liu, Y., Berendsen, A. D., Jia, S., Lotinun, S., Baron, R., Ferrara, N., et al. (2012). Intracellular VEGF regulates the balance between osteoblast and adipocyte differentiation. *J. Clin. Invest.* 122, 3101–3113. doi: 10.1172/JCI61209

Lund, E., Oldenburg, A. R., and Collas, P. (2014). Enriched domain detector: a program for detection of wide genomic enrichment domains robust against local variations. *Nucleic Acids Res.* 42:e92. doi: 10.1093/nar/gku324

Lund, E. G., Duband-Goulet, I., Oldenburg, A., Buendia, B., and Collas, P. (2015). Distinct features of lamin A-interacting chromatin domains mapped by Chip-sequencing from sonicated or micrococcal nuclease-digested chromatin. *Nucleus* 6, 30–39. doi: 10.4161/19491034.2014.990855

Makarov, A. A., Zou, J., Houston, D. R., Spanos, C., Solovyova, A. S., Cardenal-Peralta, C., et al. (2019). Lamin A molecular compression and sliding as mechanisms behind nucleoskeleton elasticity. *Nat. Commun.* 10:3056 . doi: 10.1038/s41467-019-11063-6

Malashicheva, A., Bogdanova, M., Zabirnyk, A., Smolina, N., Ignatieva, E., Freilikhman, O., et al. (2015). Various lamin A/C mutations alter expression profile of mesenchymal stem cells in mutation specific manner. *Mol. Genet. Metab.* 115, 118–127. doi: 10.1016/j.ymgme.2015.04.006

Maniotis, A. J., Chen, C. S., and Ingber, D. E. (1997). Demonstration of mechanical connections between integrins, cytoskeletal filaments, and nucleoplasm that stabilize nuclear structure (cell mechanics?cell engineering?tensegrity?extracellular matrix?mechanotransduction). *Cell Biol.* 94, 849–854.

Maraldi, N. M., Capanni, C., Cenni, V., Fini, M., and Lattanzi, G. (2011). Laminopathies and lamin-associated signaling pathways. *J. Cell. Biochem.* 112, 979–992. doi: 10.1002/jcb.22992

Maraldi, N. M., Lattanzi, G., Cenni, V., Bavelloni, A., Marmiroli, S., and Manzoli, F. A. (2010). Laminopathies and A-type lamin-associated signalling pathways. *Adv. Enzyme Regul.* 50, 248–261. doi: 10.1016/j.advenzreg.2009.10.019

Margalit, A., Brachner, A., Gotzmann, J., Foisner, R., and Gruenbaum, Y. (2007). Barrier-to-autointegration factor - a BAFfling little protein. *Trends Cell Biol.* 17, 202–208. doi: 10.1016/j.tcb.2007.02.004

Mariappan, I., Gurung, R., Thanumalayan, S., and Parnaik, V. K. (2007). Identification of cyclin D3 as a new interaction partner of lamin A/C. *Biochem. Biophys. Res. Commun.* 355, 981–985. doi: 10.1016/j.bbrc.2007.02.060

Martino, F., Perestrelo, A. R., Vinarský, V., Pagliari, S., and Forte, G. (2018). Cellular mechanotransduction: from tension to function. *Front. Physiol.* 9:824. doi: 10.3389/fphys.2018.00824

Mattout, A., Goldberg, M., Tzur, Y., Margalit, A., and Gruenbaum, Y. (2007). Specific and conserved sequences in D. melanogaster and C. elegans lamins and histone H2A mediate the attachment of lamins to chromosomes. *J. Cell Sci.* 120, 77–85. doi: 10.1242/jcs.03325

Melcer, S., and Meshorer, E. (2010). The silence of the LADs: dynamic genome-lamina interactions during ESC differentiation. *Cell Stem Cell* 6, 495–497. doi: 10.1016/j.stem.2010.05.006

Mirza, A. N., Gonzalez, F., Ha, S. K., and Oro, A. E. (2021). The Sky's the LEMit: new insights into nuclear structure regulation of transcription factor activity. *Curr. Opin. Cell Biol.* 68, 173–180. doi: 10.1016/j.ceb.2020.10.006

Mounkes, L. C., Kozlov, S. V., Rottman, J. N., and Stewart, C. L. (2005). Expression of an LMNA-N195K variant of A-type lamins results in cardiac conduction defects and death in mice. *Hum. Mol. Genet.* 14, 2167–2180. doi: 10.1093/hmg/ddi221

Muchir, A., Bonne, G., Van Der Kool, A. J., Van Meegen, M., Baas, F., Bolhuis, P. A., et al. (2000). Identification of mutations in the gene encoding lamins A/C in autosomal dominant limb girdle muscular dystrophy with atrioventricular conduction disturbances (LGMD1B). *Hum. Mol. Genet.* 9, 1453–1459. doi: 10.1093/hmg/9.9.1453

Muchir, A., Pavlidis, P., Bonne, G., Hayashi, Y. K., and Worman, H. J. (2007). Activation of MAPK in hearts of EMD null mice: similarities between mouse models of X-linked and autosomal dominant Emery - Dreifuss muscular dystrophy. *Hum. Mol. Genet.* 16, 1884–1895. doi: 10.1093/hmg/ddm137

Muchir, A., and Worman, H. J. (2007). Emery-Dreifuss muscular dystrophy. *Curr. Neurol. Neurosci. Rep.* 7, 78–83. doi: 10.1007/s11910-007-0025-3

Naetar, N., Ferraioli, S., and Foisner, R. (2017). Lamins in the nuclear interior - Life outside the lamina. *J. Cell Sci.* 130, 2087–2096. doi: 10.1242/jcs.203430

Navarro, C. L., Cau, P., and Lévy, N. (2006). Molecular bases of progeroid syndromes. *Hum. Mol. Genet.* 15, 151–161. doi: 10.1093/hmg/ddl214

Oldenburg, A., Briand, N., Sørensen, A. L., Cahyani, I., Shah, A., Moskaug, J. Ø., et al. (2017). A lipodystrophy-causing lamin A mutant alters conformation and epigenetic regulation of the anti-adipogenic MIR335 locus. *J. Cell Biol.* 216, 2731–2743. doi: 10.1083/jcb.201701043

Oldenburg, A. R., and Collas, P. (2016). Mapping nuclear lamin-genome interactions by chromatin immunoprecipitation of nuclear lamins. *Methods Mol. Biol.* 1411, 315–324. doi: 10.1007/978-1-4939-3530-7_20

Osmanagic-Myers, S., Dechat, T., and Foisner, R. (2015). Lamins at the crossroads of mechanosignaling. *Genes Dev.* 29, 225–237. doi: 10.1101/gad.255968.114

Osmanagic-Myers, S., and Foisner, R. (2019). The structural and gene expression hypotheses in laminopathic diseases - Not so different after all. *Mol. Biol. Cell* 30, 1786–1790. doi: 10.1091/mbc.E18-10-0672

Ou, H. D., Phan, S., Deerinck, T. J., Thor, A., Ellisman, M. H., and O'Shea, C. C. (2017). ChromEMT: visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* 357:eaag0025. doi: 10.1126/science.aag0025

Parnaik, V. K. (2008). Role of Nuclear Lamins in Nuclear Organization, Cellular Signaling, and Inherited Diseases. *Int. Rev. Cell Mol. Biol.* 266, 157–206. doi: 10.1016/S1937-6448(07)66004-3

Pereira, S., Bourgeois, P., Navarro, C., Esteves-Vieira, V., Cau, P., De Sandre-Giovannoli, A., et al. (2008). HGPS and related premature aging disorders: from genomic identification to the first therapeutic approaches. *Mech. Ageing Dev.* 129, 449–459. doi: 10.1016/j.mad.2008.04.003

Perepelina, K., Dmitrieva, R., Ignatieva, E., Borodkina, A., Kostareva, A., and Malashicheva, A. (2018). Lamin A/C mutation associated with lipodystrophy influences adipogenic differentiation of stem cells through interaction with Notch signaling. *Biochem. Cell Biol.* 96, 342–348. doi: 10.1139/bcb-2017-0210

Perepelina, K., Klauzen, P., Khudiakov, A., Zlotina, A., Fomicheva, Y., Rudenko, D., et al. (2020a). Generation of two iPSC lines (FAMRCi006-A and FAMRCi006-B) from patient with dilated cardiomyopathy and Emery–Dreifuss muscular dystrophy associated with genetic variant LMNAp.Arg527Pro. *Stem Cell Res.* 43:101714. doi: 10.1016/j.scr.2020.101714

Perepelina, K., Klauzen, P., Kostareva, A., and Malashicheva, A. (2019). Tissue-Specific Influence of Lamin A Mutations on Notch Signaling and Osteogenic Phenotype of Primary Human Mesenchymal Cells. *Cells* 8:266. doi: 10.3390/cells8030266

Perepelina, K., Kostina, A., Klauzen, P., Khudiakov, A., Rabino, M., Crasto, S., et al. (2020b). Generation of two iPSC lines (FAMRCi007-A and FAMRCi007-B) from patient with Emery–Dreifuss muscular dystrophy and heart rhythm abnormalities carrying genetic variant LMNA p.Arg249Gln. *Stem Cell Res.* 47:101895. doi: 10.1016/j.scr.2020.101895

Poh, Y. C., Shevtsov, S. P., Chowdhury, F., Wu, D. C., Na, S., Dundr, M., et al. (2012). Dynamic force-induced direct dissociation of protein complexes in a nuclear body in living cells. *Nat. Commun.* 3:866. doi: 10.1038/ncomms1873

Poleshko, A., Shah, P. P., Gupta, M., Babu, A., Morley, M. P., Manderfield, L. J., et al. (2017). Genome-Nuclear Lamina Interactions Regulate Cardiac Stem Cell Lineage Restriction. *Cell* 171, 573–587.e14. doi: 10.1016/j.cell.2017.09.018

Poleshko, A., Smith, C. L., Nguyen, S. C., Sivaramakrishnan, P., Wong, K. G., Murray, J. I., et al. (2019). H3k9me2 orchestrates inheritance of spatial positioning of peripheral heterochromatin through mitosis. *Elife* 8:e49278. doi: 10.7554/eLife.49278

Prokocimer, M., Davidovich, M., Nissim-Rafinia, M., Wiesel-Motiuk, N., Bar, D. Z., Barkan, R., et al. (2009). Nuclear lamins: key regulators of nuclear structure and activities. *J. Cell. Mol. Med.* 13, 1059–1085. doi: 10.1111/j.1582-4934.2008.00676.x

Rankin, J., and Ellard, S. (2006). The laminopathies: a clinical review. *Clin. Genet.* 70, 261–274. doi: 10.1111/j.1399-0004.2006.00677.x

Ricci, M. A., Cosma, M. P., and Lakadamyali, M. (2017). Super resolution imaging of chromatin in pluripotency, differentiation, and reprogramming. *Curr. Opin. Genet. Dev.* 46, 186–193. doi: 10.1016/j.gde.2017.07.010

Robson, M. I., de las Heras, J. I., Czapiewski, R., Lê Thành, P., Booth, D. G., Kelly, D. A., et al. (2016). Tissue-Specific Gene Repositioning by Muscle Nuclear Membrane Proteins Enhances Repression of Critical Developmental Genes during Myogenesis. *Mol. Cell* 62, 834–847. doi: 10.1016/j.molcel.2016.04.035

Scaffidi, P., and Misteli, T. (2008). Lamin A-dependent misregulation of adult stem cells associated with accelerated ageing. *Nat. Cell Biol.* 10, 452–459. doi: 10.1038/ncb1708

Schirmer, E. C., Florens, L., Guan, T., Yates, J. R., and Gerace, L. (2003). Nuclear membrane proteins with potential disease links found by subtractive proteomics. *Science* 301, 1380–1382. doi: 10.1126/science.1088176

Schreiber, K. H., and Kennedy, B. K. (2013). When lamins go bad: nuclear structure and disease. *Cell* 152, 1365–1375. doi: 10.1016/j.cell.2013.02.015

Schwanbeck, R., Martini, S., Bernoth, K., and Just, U. (2011). The Notch signaling pathway: molecular basis of cell context dependency. *Eur. J. Cell Biol.* 90, 572–581. doi: 10.1016/j.ejcb.2010.10.004

Shah, P. P., Lv, W., Rhoades, J. H., Poleshko, A., Abbey, D., Caporizzo, M. A., et al. (2021). Pathogenic LMNA variants disrupt cardiac lamina-chromatin interactions and de-repress alternative fate genes. *Cell Stem Cell* 28, 938–954.e9. doi: 10.1016/j.stem.2020.12.016

Shevelyov, Y. Y., and Ulianov, S. V. (2019). The Nuclear Lamina as an Organizer of Chromosome Architecture. *Cells* 8:136. doi: 10.3390/cells8020136

Shumaker, D. K., Solimando, L., Sengupta, K., Shimi, T., Adam, S. A., Grunwald, A., et al. (2008). The highly conserved nuclear lamin Ig-fold binds to PCNA: its role in DNA replication. *J. Cell Biol.* 181, 269–280. doi: 10.1083/jcb.200708155

Steele-Stallard, H. B., Pinton, L., Sarcar, S., Ozdemir, T., Maffioletti, S. M., Zammit, P. S., et al. (2018). Modeling skeletal muscle laminopathies using human induced pluripotent stem cells carrying pathogenic LMNA mutations. *Front. Physiol.* 9:1332. doi: 10.3389/fphys.2018.01332

Swift, J., Ivanovska, I. L., Buxboim, A., Harada, T., Dingal, P. C. D. P., Pinter, J., et al. (2013). Nuclear lamin-A scales with tissue stiffness and enhances matrix-directed differentiation. *Science* 341:1240104. doi: 10.1126/science.1240104

Tong, J., Li, W., Vidal, C., Yeo, L. S., Fatkin, D., and Duque, G. (2011). Lamin A/C deficiency is associated with fat infiltration of muscle and bone. *Mech. Ageing Dev.* 132, 552–559. doi: 10.1016/j.mad.2011.09.004

Ullah, I., Subbarao, R. B., and Rho, G. J. (2015). Human mesenchymal stem cells - Current trends and future prospective. *Biosci. Rep.* 35:e00191. doi: 10.1042/BSR20150025

Van Berlo, J. H., Voncken, J. W., Kubben, N., Broers, J. L. V., Duisters, R., van Leeuwen, R. E. W., et al. (2005). A-type lamins are essential for TGF-β1 induced PP2A to dephosphorylate transcription factors. *Hum. Mol. Genet.* 14, 2839–2849. doi: 10.1093/hmg/ddi316

van Schaik, T., Vos, M., Peric-Hupkes, D., Hn Celie, P., and van Steensel, B. (2020). Cell cycle dynamics of lamina-associated DNA. *EMBO Rep.* 21:e50636. doi: 10.15252/embr.202050636

van Steensel, B., and Belmont, A. S. (2017). Lamina-Associated Domains: links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* 169, 780–791. doi: 10.1016/j.cell.2017.04.022

Van Steensel, B., and Henikoff, S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase. *Nat. Biotechnol.* 18, 424–428. doi: 10.1038/74487

van Tienen, F. H. J., Lindsey, P. J., Kamps, M. A. F., Krapels, I. P., Ramaekers, F. C. S., Brunner, H. G., et al. (2019). Assessment of fibroblast nuclear morphology aids interpretation of LMNA variants. *Eur. J. Hum. Genet.* 27, 389–399. doi: 10.1038/s41431-018-0294-0

Wang, T., Zhang, X., and Bikle, D. D. (2017). Osteogenic Differentiation of Periosteal Cells During Fracture Healing. *J. Cell. Physiol.* 232, 913–921. doi: 10.1002/jcp.25641

Wong, X., and Stewart, C. L. (2020). The Laminopathies and the Insights They Provide into the Structural and Functional Organization of the Nucleus. *Annu. Rev. Genomics Hum. Genet.* 21, 263–288. doi: 10.1146/annurev-genom-121219-083616

Worman, H. J. (2012). Nuclear lamins and laminopathies. *J. Pathol.* 226, 316–325. doi: 10.1002/path.2999

Worman, H. J. (2018). Cell signaling abnormalities in cardiomyopathy caused by lamin A/C gene mutations. *Biochem. Soc. Trans.* 46, 37–42. doi: 10.1042/BST20170236

Worman, H. J., Ostlund, C., and Wang, Y. (2010). Diseases of the nuclear envelope. *Cold Spring Harb. Perspect. Biol.* 2:a000760. doi: 10.1101/cshperspect.a000760

Wu, D., Flannery, A. R., Cai, H., Ko, E., and Cao, K. (2014). Nuclear localization signal deletion mutants of lamin A and progerin reveal insights into lamin A processing and emerin targeting. *Nucleus* 5, 37–41. doi: 10.4161/nucl.28068

Zaremba-Czogalla, M., Dubińska-Magiera, M., and Rzepecki, R. (2011). Laminopathies: the molecular background of the disease and the prospects for its treatment. *Cell. Mol. Biol. Lett.* 16, 114–148. doi: 10.2478/s11658-010-0038-9

Zhang, B., Yang, Y., Keyimu, R., Hao, J., Zhao, Z., and Ye, R. (2019). The role of lamin A/C in mesenchymal stem cell differentiation. *J. Physiol. Biochem.* 75, 11–18. doi: 10.1007/s13105-019-00661-z

Zhang, Y. Q., and Sarge, K. D. (2008). Sumoylation regulates lamin A function and is lost in lamin A mutants associated with familial cardiomyopathies. *J. Cell Biol.* 182, 35–39. doi: 10.1083/jcb.200712124

Zuleger, N., Robson, M. I., and Schirmer, E. C. (2011). The nuclear envelope as a chromatin organizer. *Nucleus* 2, 339–349. doi: 10.4161/nucl.2.5.17846

Building a Mammalian Retina: An Eye on Chromatin Structure

Marwa Daghsni[1] and Issam Aldiri[1,2,3]*

[1]Department of Ophthalmology, School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States, [2]Department of Developmental Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States, [3]Louis J. Fox Center for Vision Restoration, University of Pittsburgh, Pittsburgh, PA, United States

Regulation of gene expression by chromatin structure has been under intensive investigation, establishing nuclear organization and genome architecture as a potent and effective means of regulating developmental processes. The substantial growth in our knowledge of the molecular mechanisms underlying retinogenesis has been powered by several genome-wide based tools that mapped chromatin organization at multiple cellular and biochemical levels. Studies profiling the retinal epigenome and transcriptome have allowed the systematic annotation of putative cis-regulatory elements associated with transcriptional programs that drive retinal neural differentiation, laying the groundwork to understand spatiotemporal retinal gene regulation at a mechanistic level. In this review, we outline recent advances in our understanding of the chromatin architecture in the mammalian retina during development and disease. We focus on the emerging roles of non-coding regulatory elements in controlling retinal cell-type specific transcriptional programs, and discuss potential implications in untangling the etiology of eye-related disorders.

Keywords: progenitors, retina, epigenetics, enhancers, histones, genome organization, neurogenesis, cell fate

RETINAL DEVELOPMENT

The retina has been an excellent system to study neurogenesis, due to its simplified anatomical structure, accessibility and well-defined cell types (Agathocleous and Harris, 2009; Demb and Singer, 2015). The vertebrate mature retina contains seven morphologically and functionally distinct cell types, including six types of neurons (ganglion cells, amacrines, bipolars, horizontal cells, and rod and cone photoreceptors) and one type of glia, the Müller glia (**Figure 1**) (Cepko et al., 1996). Retinal cells are organized into three layers (outer nuclear layer, inner nuclear layer and ganglion cell layer) interconnected by two synaptic layers that facilitate processing of visual signals (**Figure 1**) (Fisher, 1979). The visual pathway initiates by the response of the photoreceptors to a light stimulus, transducing it into action potentials that propagate to the retinal interneurons (horizontal, bipolar and amacrine cells) and ganglion cells. Eventually the visual input is relayed to the brain through retinal ganglion cell axons that collectively form the optic nerve.

Retinal differentiation initiates when multipotent retinal progenitor cells (RPCs) exit the cell cycle and differentiate into neurons or glia in a temporally conserved order under the control of gene regulatory networks and signaling pathways (**Figure 1**) (Wetts and Fraser, 1988; Turner et al., 1990; Agathocleous and Harris, 2009). Early retinal development is coordinated by a group of transcription factors (Rax, Otx2, Pax6, Six3, Lhx2, Vsx2 and other) that specifies the eye field within the developing forebrain, promotes retinal proliferation and primes RPCs for subsequent neural differentiation (Zuber et al., 2003). Mutations in many of these genes underlie severe retinal developmental

FIGURE 1 | Retinal neurogenesis and organization of the mammalian retina. **(A)** Schematic diagram illustrating waves of retinal neurogenesis and approximate timing of retinal cell type birth. Note that rod photoreceptors, bipolar cells and Müller glia are mainly formed postnatally. **(B)** DAPI staining of the adult mouse retina showing its exquisite laminar structure. **(C)** Retinal laminar position of different cell types.

disorders, as observed in microphthalmia (small eye), anophthalmia (absence of the eye), and coloboma (failure in optic fissure closure) cases (Slavotinek, 2011; Reis and Semina, 2015).

Unlike development in the mammalian cortex, retinal cell types are born in waves during which the periods of neuron generation overlap considerably (**Figure 1**) (Marquardt and Gruss, 2002). Hence, retinal cell types are often classified into early born cell types (ganglion cells, cones, amacrine and horizontal cells) and late born cell types (rods, bipolar cells and Müller glia) (Ohsawa and Kageyama, 2008). Experimental evidence suggests that the ability of retinal progenitors to produce different cell types (competence) changes as development progresses: early progenitors generate early born cell types while late progenitors produce late born cell types (Livesey and Cepko, 2001; Hafler et al., 2012).

The mechanisms that determine RPC competence are rooted in the ability of progenitor cells to integrate signaling pathways and the activities of complex networks of transcription factors (TFs) that drive cell fate decisions at the genomic level (Livesey and Cepko, 2001; Agathocleous and Harris, 2009). Chromatin regulation allows interpretation of identical genomes in a variety of ways, leading to cell type specific transcriptional outputs (Soshnev et al., 2016). Hence, chromatin architecture of the developing retina has been intensively studied, resulting in a wealth of information on transcriptional programs influenced by chromatin regulation during retinogenesis.

Chromatin regulators and retinal lineage-specific programs.

Nuclear DNA is wrapped around a disc of highly conserved proteins (histones) to form the nucleosome, the basic unit of chromatin. Histones are classified into core histones (H2A, H2B, H3 and H4), the principal components of the nucleosome, and linker histones (H1), which bind the nucleosome at the cross point of DNA entry/exit sites (Luger et al., 1997; Vignali and

Workman, 1998). Accessibility to DNA requires nucleosome mobilization, which is mediated by large complexes that utilize ATP hydrolysis in the process (Wilson and Roberts, 2011; Kadoch and Crabtree, 2015; Centore et al., 2020). The structural changes in chromatin are often associated with deposition and/or removal of chemical modifications on histone tails, facilitated by distinct multimeric complexes with enzymatic activity (Soshnev et al., 2016; Villasenor and Baubec, 2021).

Given the association between chromatin pathways and regulation of gene expression, genetic studies have focused on investigating the roles of chromatin remodelers and histone modifying complexes during retinal development, a topic that has been reviewed recently (Corso-Diaz et al., 2018; Raeisossadati et al., 2021). Briefly, these studies revealed that chromatin regulators influence retinal progenitor proliferation and cell fate determination in a context-dependent manner. For instance, multiple studies investigated the effect of loss of the polycomb repressive complex 2 (PRC2), which catalyzes the addition of the repressive mark H3K27me3, on retinal development (Aldiri and Vetter, 2009; Aldiri et al., 2013; Iida et al., 2015; Zhang et al., 2015; Yan et al., 2016; Cheng et al., 2018; Fujimura et al., 2018). Mutations in the PRC2 core subunits Ezh2 or Eed lead to reduced retinal proliferation and alteration in neuronal cell fate, particularly amacrine cells, and glia formation (Iida et al., 2015; Zhang et al., 2015; Fujimura et al., 2018). In the postnatal retina, loss of PRC2 function caused photoreceptor degeneration, mediated by a de-repression of the PRC2 targets Six1 and Eya2 (Yan et al., 2016). Meanwhile, perturbation of H3K27me3 removal by knocking down the H3K27me3 demethylase Jmjd3 impacts retinal bipolar cell formation (Iida et al., 2014). Cell-type specific alterations were also observed when MLL, the core subunit of a complex required for mono- and di-methylation of H3K4, was mutated during retinal

development. Here, a conditional knockout of MLL impacts retinal proliferation and leads to a progressive loss of horizontal cells in the differentiating retina (Brightman et al., 2018). These examples highlight how chromatin modifying enzymes control multiple aspects of retinal development.

The function of chromatin remodelers that govern nucleosome mobilization has been investigated as well (Das et al., 2007; Lamba et al., 2008; Aldiri et al., 2015). For instance, evidence indicates that Brg1, a core subunit of the SWI/SNF complex, is required for retinal proliferation and photoreceptor differentiation (Aldiri et al., 2015). The effect of Brg1 is likely mediated by its ability to influence the chromatin landscape near actively transcribed cell-type specific genes, as Brg1 predominantly occupies active cis-regulatory elements in the retina, and previous work demonstrated that Brg1 binds transcription factors that drive neurogenesis such as Pax6 and NeuroD1 (Seo et al., 2005; Ninkovic et al., 2013). Additionally, work on cell lines suggests that activities of the enhancer landscape of lineage specification genes is sensitive to the loss of SWI/SNF chromatin remodeling complexes (Aldiri et al., 2015; Alver et al., 2017).

Chromatin-associated complexes can change subunit composition during development, indicative of cell-type-specific roles (Lessard et al., 2007). Indeed, several auxiliary subunits of chromatin regulator complexes are expressed in a stage-specific manner during retinal development but the exact molecular and cellular phenotypes resulting from mutating these proteins during retinogenesis remains to be explored (Lamba et al., 2008; Aldiri et al., 2015).

More recently, chromosome confirmation capture (3C) techniques revealed that manipulation of chromatin regulators such as SWI/SNF and the polycomb repressive complexes can lead to changes in compartment-level chromatin organization (Schoenfelder et al., 2015; Barutcu et al., 2016; Cruz-Molina et al., 2017). These intriguing findings link regulation of gene expression with 3D chromatin architecture via activities of chromatin regulators, a function yet to be explored in the retina.

EPIGENETIC LANDSCAPE DYNAMICS DURING RETINOGENESIS

Genome-wide profiling of histone marks and chromatin associated proteins greatly facilitated the in depth probing of chromatin signature dynamics during developmental stages of mouse and human retina, revealing non-random genomic localization of histone marks and association with gene expression (Popova et al., 2012; Mo et al., 2016; Ueno et al., 2016; Aldiri et al., 2017). In progenitor cells, differentiation genes are poised (H3K27me3-occupied) toward activation and as retinal development proceeds, H3K27me3 is lost and cell type specific genes are expressed (Ueno et al., 2016; Aldiri et al., 2017). Interestingly, the accumulation of H3K27me3 on progenitor genes in differentiated neurons is not as common (Aldiri et al., 2017).

The retinal enhancer landscape exhibits exquisite reconfiguration concomitant with changes in gene expression

during retinal developmental transitions: whereas cis-regulatory elements of progenitor genes lose their activities, enhancers targeting differentiation genes are gradually activated (Aldiri et al., 2017). Mechanisms of enhancer potentiation have been the focus of many studies. Current models suggest that priming enhancers for activation during embryonic development can be achieved by a cooperative binding of lineage-specific TFs or by the deployment of a unique set of TFs, termed pioneer factors, that have the ability to bind closed chromatin and facilitate the recruitment of chromatin regulators and lineage-specific TFs and co-factors (Zaret and Carroll, 2011; Drouin, 2014). Retinal pioneer TFs remain poorly characterized but recent genomic data begins to shed light on their roles. For instance, a study examining the genomic profiling of the RPC gene LHX2 reveals global and local reduction of LHX2-bound chromatin accessible sites upon loss of Lhx2, including regulatory regions nearby TFs with potential pioneer function, suggesting that LHX2 functions as a pioneer factor in the developing retina (Zibetti et al., 2019). In another work, analysis of the regulatory elements bound by the photoreceptor differentiation transcription factor Crx in wild type and Crx-mutant retina in mice indicates a limited ability of CRX to remodel chromatin and points toward a cooperative TF binding module in promoting photoreceptor cell fate (Ruzycki et al., 2018). Thus, priming the retinal enhancer landscape during developmental transitions and cell fate choices likely involves multiple mechanisms and is highly context specific.

Mechanisms of enhancer-mediated transcriptional control of genes with multiphasic expression during retinogenesis are particularly interesting, and underscore the complexity of gene regulation. For instance, the transcription factor Sox2 is expressed in RPCs and is confined to amacrine cells and Müller glia in adult retina (Taranova et al., 2006). In principle, such a complex temporal and spatial expression pattern can occur via recruitment of stage- and cell-type specific TFs and/or by the utilization of cell-type exclusive enhancers. Retina-specific enhancer elements with temporally restricted activities have been identified as the case with those nearby Otx2, a transcription factor expressed in a subset of progenitor cells and marks bipolar cells and photoreceptors (Emerson and Cepko, 2011; Kaufman et al., 2021). Notably, Sox2 chromatin architecture has been studied given its essential roles in maintaining stem cell pluripotency, revealing a complex regulatory landscape with multiple putative enhancer elements, including stem cell-specific regulatory constituents that are essential for Sox2 expression (Li et al., 2014; Zhou et al., 2014; Bonev et al., 2017).

Interestingly, downregulation of Sox2 in rod photoreceptors is accompanied by site-specific deposition of the repressive histone mark H3K27me3 (Norrie et al., 2019). Whereas Sox2 coding region and nearby enhancers are occupied by H3K27me3, Sox2-regulatory elements that are hundreds of base pairs away holds limited levels. This implies that not all regulatory elements are created equally and underscores a locus-specific utilization of repressive mechanisms on enhancer elements. Florescence in situ hybridization (FISH) performed on rod nuclei indicates that while Sox2 coding region is located in euchromatin, its long-range putative enhancers reside in heterochromatin, thus likely

inaccessible to the action of repressive complexes (Norrie et al., 2019). These data are in agreement with the finding that rod photoreceptors render a substantial fraction of vestigial regulatory elements (enhancers that used to be active in earlier stages of retinogenesis) inaccessible to repression mediated by DNA methyltransferase (Mo et al., 2016).

Diverse histone marks tend to co-exist, leading to an excessive number of possible combinatorial readouts and renders interpretation of epigenomic maps challenging. To facilitate a better understanding to the biological roles of combinations of histone marks and chromatin associated proteins, a computational modeling that utilizes machine learning algorithms (ChromHMM) was developed to distinguish groups (states) of co-occurring chromatin marks across the genome (Ernst and Kellis, 2010; 2012). Applying this method to ChIP-Seq data generated from mouse and human developing retina led to the identification of several chromatin states that capture known genomic elements such as active promoters and enhancers, insulators and repressed regions (Aldiri et al., 2017). ChromHMM analysis was also informative in exploring prevailing chromatin states in retinoblastoma and retinal organoids (Hiler et al., 2015; Aldiri et al., 2017). Later, a computational work that integrates retinal chromatin states and 3D FISH imaging successfully predicted genome-wide euchromatin and heterochromatin compartmentalization in the mouse retina (Norrie et al., 2019).

Mapping of the epigenomic marks and regions of chromatin accessibility has emerged as a powerful tool to annotate retinal putative regulatory elements, particularly enhancers (Wilken et al., 2015; Aldiri et al., 2017; Hughes et al., 2017; Wang et al., 2018; Cherry et al., 2020; Xie et al., 2020). Given the essential roles of enhancer elements in controlling cell type specific differentiation programs during retinogenesis, we will discuss recent progress in the field and highlight examples related to the gene regulatory networks controlling retinal cell fate choices.

DISCOVERY OF RETINAL ENHANCERS

Enhancers are stretches of non-coding DNA elements that spatially and temporally regulate transcription by acting as platforms to recruit transcription factors and transcriptional machinery, irrespective of sequence orientation (Gasperini et al., 2020). Enhancers are the main source for communication between chromatin and the environment as they contain motifs that can bind transcription factors and recruit effectors of signaling pathways (Long et al., 2016). Biochemically, enhancers are characterized by occupancy of active histone marks (i.e., H3K27ac and/or H3K4me1) and chromatin-associated proteins (i.e., p300), overlaying areas of open chromatin (Visel et al., 2009; Creyghton et al., 2010; Rada-Iglesias et al., 2011; Thurman et al., 2012). Interestingly, while H3K27Ac has been widely validated as a hallmark for active enhancers in the animal kingdom, association of H3K27Ac deposition with active regulatory elements in plants appears species-specific (Yan et al., 2019).

Recent advances in techniques that map 3D genome organization demonstrated that enhancers may act over long genomic distances, via looping, to contact their cognate gene promoters in 3D space (Li et al., 2012; Fang et al., 2016; Mumbach et al., 2016). The prevailing model is that a promoter-enhancer interaction mediates activation of gene expression by bringing transcription factors and transcription machinery into promoter proximity (Robson et al., 2019). However, whether promoter-enhancer physical contact is a universal prerequisite mechanism for gene activation is not firmly established (Chen et al., 2018; Benabdallah et al., 2019; Crump et al., 2021). There are hundreds of thousands of putative regulatory elements in the human genome, far in excess of number of genes, underscoring the complexity of enhancer function in organ development and homeostasis.

Classically, strategies to pinpoint cell-type specific cis-regulatory elements in the developing retina have exploited DNA conservation and enrichment of lineage-specific transcription factor motifs coupled with in vivo screening for enhancer activities. This method was successful in the identification of numerous distal regulatory elements near genes essential for retinal development and cell-type specification such as Vsx2 and Grm6 (bipolar cells), Nrl, Otx2 and Prdm1 (photoreceptors), Atoh7 (ganglion cells), Onecut1 and Thrb (cones/horizontal cells), and Pax6 (RPCs, amacrine cells), among others (Kleinjan et al., 2004; Rowan and Cepko, 2005; Riesenberg et al., 2009; Willardsen et al., 2009; Emerson and Cepko, 2011; Kautzmann et al., 2011; Emerson et al., 2013; Mills et al., 2017; Goodson et al., 2020; Patoori et al., 2020).

Comparative genomics employing convergent evolution were also useful in identifying and characterizing putative retinal enhancer elements (Kvon et al., 2016; Partha et al., 2017; Roscito et al., 2018). The logic behind this interesting method is that regulatory elements that are essential for vision are under evolutionarily constraints to preserve visual structures and functions. In animals where vision is regressed, such as subterranean mammals, vision-related regulatory regions and/or their target genes undergo accelerated mutation rate and suffer sequence divergence due to relaxed evolutionarily constraints, thus revealing DNA sequences potentially essential for development of optical structures. Such a strategy was employed to investigate enhancer elements in the ground-dwelling moles, leading to the identification of several retina-specific regulatory regions associated with vision deterioration, including those nearby Pax6 (Partha et al., 2017). Still, not all regulatory elements are conserved at the DNA level, and many highly conserved enhancers lack in vivo activities in transgenic assays (Pennacchio et al., 2013). Thus, complementary approaches to profile the cis-regulome remain essential to elucidate enhancer structure and function.

With the broad availability of next generation sequencing platforms, profiling chromatin structure in the developing retina has taken a momentum, facilitating the discovery of genome wide putative distal enhancers with a relative ease. Taking advantage of transcription factors occupancy as a proxy to the identification of distal enhancer regions, numerous transcription factors involved in retinal cell fate

choices have been surveyed using ChIP-Seq and, more recently, CUT and RUN, including OTX2, ATOH7, NRL, CRX, MEF2D, RORB and LHX2 (Corbo et al., 2010; Swaroop et al., 2010; Samuel et al., 2014; Andzelm et al., 2015; Zibetti et al., 2019; Cherry et al., 2020; Brodie-Kommit et al., 2021). Likewise, histone modifications associated with active promoters and enhancers have been extensively charted in the developing retina, and hundreds of cis-regulatory elements have been catalogued (Popova et al., 2012; Wilken et al., 2015; Aldiri et al., 2017; Hughes et al., 2017; Xie et al., 2020). More recently, a transcriptional profiling of non-coding RNAs, often transcribed from active enhancer regions, was performed to delineate cone and rod regulatory elements in wild type and Nrl mutant mice (Perez-Cervantes et al., 2020).

Chromatin accessibility has become a popular method to identify cis-regulatory elements. Studies on bulk tissues from human and murine developing retina revealed temporal dynamics in chromatin accessibility associated with changes in gene expression during retinogenesis. Earlier work utilized DNaseI hypersensitivity (DHS) to profile the mouse developing retina, leading to the identification of developmentally regulated enhancer elements near the transcription factors Neurog2, Otx2 and Olig2, (Wilken et al., 2015). By far, assay for transposase-accessible chromatin with sequencing (ATAC-Seq) has become the most common technique used to profile regulatory elements, revealing enhancer landscape dynamics in the mouse and human developing retina (Mo et al., 2016; Aldiri et al., 2017; Hughes et al., 2017; Cherry et al., 2020; Xie et al., 2020). As retinal organoids become a powerful method to investigate and model retinal development and disease (Eiraku et al., 2011; Volkner et al., 2016), ATAC-Seq was used to demonstrate a high temporal correlation of regulatory landscape dynamics in retinal organoids and human fetal retina, further validating retinal organoids as a robust model to study human retina (Xie et al., 2020).

To date, most of the studies surveying retinal open chromatin regions used bulk tissues as input, which renders the determination of cell type specific deployment of regulatory elements in rare retinal cell types challenging. To overcome this limitation, ATAC-Seq, and sometimes ChIP-Seq, experiments have been performed on purified cells from transgenic mice carrying cell type-specific reporter genes, and as a result, epigenomic data from enriched rods, cones, bipolar cells and Müller glia are now available (Mo et al., 2016; Ueno et al., 2016; Hughes et al., 2017; Ueno et al., 2017; Murphy et al., 2019; VandenBosch et al., 2020).

The advent of single cell technologies, methods that circumvent heterogeneity and allow the investigation of rare cell populations at high resolution, has revolutionized the field, and a large cohort of studies focusing on surveying the adult and developing retinal single cell transcriptome has been performed (Clark et al., 2019; Kim et al., 2019; Liang et al., 2019; Menon et al., 2019; Cherry et al., 2020; Lu et al., 2020; Sridhar et al., 2020; Brodie-Kommit et al., 2021; Wu et al., 2021). Still, matching studies that investigate retinal chromatin accessibility dynamics at the single cell resolution remain limited (Xie et al., 2020). With recent technical advances that enable the simultaneous profiling of transcriptome and epigenome from the same cells, it is almost certain that work is underway to accurately outlining the epigenome dynamics in relation to gene expression in retinal cell populations (Kashima et al., 2020; Weir et al., 2021).

FUNCTIONAL VALIDATION OF RETINAL ENHANCERS

Genome-wide analysis to delineate putative regulatory elements is a robust method to infer enhancer activities but not without limitations (Halfon, 2019). With the wealth of information available on the genomic location of predicted retinal enhancers, derived primarily by biochemical annotations and computational methods, *in vivo* experimental characterization of those elements remains necessary to validate their functions. In theory, a regulatory element should recapitulate its cognate gene's spatial and temporal expression pattern, and when mutated should lead to alteration in gene expression. A large body of work has been directed toward investigating enhancer activities in the retina using reporter assays, which test the ability of a candidate enhancer sequence to activate a reporter gene (i.e., GFP, LacZ and luciferase). Electroporation of the mouse developing retina has been the main method for construct introduction into the retina, testing enhancer activity one element at a time (Montana et al., 2011; Wang et al., 2014; Goodson et al., 2020; Kaufman et al., 2021). *In vivo* transgenesis using mouse, zebrafish and the frog xenopus was also used (Hutcheson et al., 2005; Ghiasvand et al., 2011; Fang et al., 2017; Bhansali et al., 2020; Kaufman et al., 2021). High throughput strategies to interrogate the activities of retinal enhancers has been explored as well. In one such a study, massively parallel reporter assay (MPRA) was used to investigate photoreceptor cis-regulatory elements bound by CRX (White et al., 2013).

CRISPR-based genome editing technology has tremendously facilitated testing the function of retinal enhancers *in vivo* by providing a venue to efficiently delete non-coding regions with precision (Osterwalder et al., 2018). Emerging studies on enhancer elements nearby Vsx2, Otx2 and Prdm1 in retinal explants and mouse knockouts uncovered lineage- and stage-specific regulatory elements important for photoreceptor and bipolar cell fates (Wang et al., 2014; Norrie et al., 2019; Chan et al., 2020; Goodson et al., 2020; Kaufman et al., 2021). Still, whether an enhancer is required for the expression of its target gene remains challenging to address given the complex nature of the chromatin landscape. Enhancers may regulate the expression of a single target (i.e., a single or many enhancers, one target gene) or acting promiscuously on multiple genes (i.e., one enhancer, multiple target genes). As such, *in vivo* perturbations of regulatory elements, especially those nearby functionally important genes, may result in no molecular or cellular consequences, likely due to enhancer redundancy (Kurokawa et al., 2004; Osterwalder et al., 2018). Additionally, an enhancer may govern the expression of gene(s) broadly expressed in multiple tissues during embryogenesis, leading to pleiotropic effects and/or embryonic lethality upon loss of enhancer function. Still, enhancer deletion assays remain an important tool to reveal molecular mechanism underlying biological functions of enhancer landscape.

Super enhancers and retinal cell type specific programs

Developmentally critical transcription factors are often marked with strong enhancers to drive and/or maintain robust expression. Work on embryonic stem cells defined a subclass of regulatory elements, termed super-enhancers (SEs), that are selectively enriched near genes important for stem cell identity (Whyte et al., 2013). Super-enhancers tend to span large genomic regions and are strongly enriched in mediator complex and transcription factors, particularly those driving lineage-specific programs (Parker et al., 2013; Adam et al., 2015). The importance of SE size is not clear, but it was proposed that strong H3K27Ac occupancy that demarcates these regulatory clusters weakens DNA-histone interactions, thus exposing DNA to transcription factors (Parker et al., 2013). Evidence suggests that SEs drive high levels of transcriptional activity and are particularly sensitive to perturbations (Hnisz et al., 2013; Loven et al., 2013; Whyte et al., 2013; Bahr et al., 2018). Recent studies propose that transcription factors, activators and co-activators occupying SEs form condensates with liquid-phase separation properties (Hnisz et al., 2017; Boija et al., 2018; Sabari et al., 2018). However, the functional significance of SEs and whether a super-enhancer constitutes a single functional unit of cooperating regulatory clusters or a mere stretch of aggregated enhancers remains unclear (Pott and Lieb, 2015; Moorthy et al., 2017).

Given the emerging interest in SEs roles in regulating tissue-specific gene expression, dynamically regulated SEs in mouse and human developing retina have been annotated (Aldiri et al., 2017). Studies that functionally investigate SEs in the retina remain limited but available data suggest central roles in driving retinal cell fate choices. For example, a large deletion (35 kb) in an area that overlaps a super-enhancer nearby Vsx2 caused a complete loss of retinal bipolar cells, while proliferation appears to proceed normally (Norrie et al., 2019). This regulatory region contains conserved elements that can drive reporter expression in RPCs, Müller glia and bipolar cells (Rowan and Cepko, 2004; Kim et al., 2008), and a recent study demonstrated that knocking down of a smaller portion of the Vsx2 SE also impacts bipolar cell differentiation (Goodson et al., 2020). Thus, while the concept of super-enhancers is appealing, detailed functional studies are needed to elucidate the exact biological roles of SEs and their constituents in promoting retinal cell fate acquisitions.

RETINAL ENHANCEROPATHIES

Defining cis-regulatory elements is crucial to understand disease mechanisms, as variations in DNA sequences linked to inherited human disorders often lie in non-coding regions (Chatterjee and Ahituv, 2017). Retinal diseases associated with alterations in regulatory landscape have been reported but only in a handful of cases has a causative link been suggested. A clear example illustrating a direct role of regulatory elements in inherited retinal disorders comes from studies on patients with nonsyndromic congenital retinal nonattachment (NCRNA), an autosomal recessive retinal disease characterized by congenital blindness due to loss of RGCs and optic nerve atrophy (Keser et al., 2017). A deletion in a non-coding DNA region 20 kb upstream of the proneural bHLH transcription factor ATOH7 has been linked to the disease (Ghiasvand et al., 2011). Transgenic reporter assays in mouse and zebrafish demonstrated that this non-coding element has developmental activities that matched the spatiotemporal expression of Atoh7, suggesting that it acts as an enhancer element for Atoh7 (Ghiasvand et al., 2011). Subsequent studies identified pathogenic mutations in the ATOH7 coding region itself, further linking NCRNA to misregulation of Atoh7 (Keser et al., 2017; Kondo et al., 2018). Surprisingly, deleting the orthologous murine enhancer region does not recapitulate the disease phenotype, suggesting a differential biological significance of mouse and human Atoh7 enhancer landscape (Miesfeld et al., 2020). Other examples that identified variations in enhancer elements with links to ocular disorders include those nearby Pax6 (aniridia) and Samd7 (retinitis pigmentosa) (Bhatia et al., 2013; Van Schil et al., 2016).

Global alterations in retinal enhancer landscape have been observed in patients with retinal degenerative diseases. A recent study profiled the genome-wide chromatin accessibility in patients with dry age-related macular degeneration (AMD), a disease characterized by a progressive loss of photoreceptors, and revealed a genome-wide quantitative reduction in chromatin accessibility associated with advanced stages of the disease, particularly in the macular region (Wang et al., 2018). Of note, the genomic regions that recruit gene regulatory networks controlling photoreceptor gene expression seems to be most impacted in those patients (Wang et al., 2018).

Retinal diseases can be associated with genomic rearrangements that lead to the formation of a *de novo* regulatory landscape, causing gene deregulation. In one such instance, a cohort of patients with autosomal-dominant retinitis pigmentosa suffered a structural rearrangement that led to a repositioning of retina-specific regulatory landscape nearby GDPD1, a gene involved in lipid metabolism. The ectopic activation of GDPD1 driven by the newly created enhancer region likely leads to de-regulation of lipid metabolism, an essential process for phototransduction (de Bruijn et al., 2020; Fu et al., 2021). This work demonstrates how recent advances in surveying 3D genome organization can facilitate the discovery of molecular mechanisms underlying retinal diseases.

Retinal 3D Nuclear Organization and High Order Chromatin

Thanks to the rapid development of 3C techniques, the mammalian 3D genome conformation has been profiled at high resolution, illuminating that chromatin is organized into compartments in which multiple levels of DNA-DNA preferential interactions exist (Dixon et al., 2012; Lieberman-Aiden et al., 2009). At the chromosome level, transcriptionally active and inactive regions are spatially segregated into very large genomic regions, called compartments A and B, respectively (Lieberman-Aiden et al., 2009). Within each compartment distinct territories, the topologically associated domains

FIGURE 2 | Retinal 3D nuclear organization. **(A)** Model for enhancer-mediated activation of gene expression. The process involves formation of a DNA loop, facilitated by the recruitment of CTCF and cohesin, that brings distal enhancers into proximity of the promoters. Enhancers enable the recruitment of transcription factors, co-factors and transcriptional machinery to the promoter. **(B,C)** Nuclear structure of murine rod **(B)** and bipolar **(C)** cells as revealed by DAPI staining.

(TADs), exist in which promoter-enhancer contacts are heavily constrained (Dixon et al., 2012; Rao et al., 2014). How hierarchical genomic folding is formed and maintained is under intensive investigation, but evidence points toward a major role of the transcription factor CTCF (Merkenschlager and Nora, 2016; Nora et al., 2012). Current models propose that genomic contacts are established *via* loop formation that involves homo-dimerization of CTCF at the loop anchors (**Figure 2**) (Rao et al., 2014; Hnisz et al., 2016; Weintraub et al., 2017). These interactions are further stabilized by a cohesin complex that forms a ring around the loop anchor region (Kagey et al., 2011; Hnisz et al., 2016; Merkenschlager and Nora, 2016; Rao et al., 2017; Weintraub et al., 2017). The orientation of CTCF binding seems to be important for the proper formation of the loop (de Wit et al., 2015; Guo et al., 2015).

Research investigating high order chromatin of the developing retina remains limited, and available data is primarily collected from bulk mouse retina and purified rod photoreceptors (Falk et al., 2019; Norrie et al., 2019; Tan et al., 2019). Overall, retina hierarchical genomic organization is similar to what has been reported in other tissues, including in the cortex, and developmental transitions of compartments A and B correlated well with retinal chromatin signature (Dixon et al., 2012; Norrie et al., 2019). However, while the number of TADs remain relatively constant as neural progenitors differentiate into cortical neurons, rod photoreceptors have significantly more TADs than in RPCs or cortical neurons, presumably due to the compact nature of rod nuclei (Bonev et al., 2017; Norrie et al., 2019).

Genomic technologies have enabled the identification and cataloging of putative regulatory elements yet defining their cognate genes remains challenging (Buecker and Wysocka, 2012; de Laat and Duboule, 2013). Promoter-enhancer contacts are generally difficult to identify using Hi-C due to resolution limitations but work on developing neural tissues captured the dynamics of several prominent interactions associated with genes important for neurogenesis (Bonev et al.,

2017; Norrie et al., 2019). This is illustrated by Sox2 locus, where changes in the Sox2 expression during cortical and retinal differentiation is associated with re-wiring of longs-range contacts between Sox2 promoter and regulatory elements hundreds of kilobases away (Bonev et al., 2017; Norrie et al., 2019).

It is now broadly accepted that enhancers can act over great genomic distances, *via* CTCF-mediated looping, to regulate promoter activities, bypassing proximally located genes (Schoenfelder and Fraser, 2019). The specific roles of CTCF in retinal differentiation remain unclear but early studies on chick retina suggest regulatory functions associated with Pax6 (Li et al., 2006). CTCF is essential for proper retinal formation as loss of CTCF expression in the murine developing retina leads to massive cell death (Watson et al., 2014). The genome wide occupancy of CTCF in the developing retina has been profiled, revealing constitutive and dynamic CTCF occupancy across retinal genome during retinogenesis (Aldiri et al., 2017). Interestingly, work on retinal organoids suggest that maintaining a robust CTCF binding memory in stem cells reprogrammed from rod photoreceptors is important for efficient differentiation of retinal organoids (Hiler et al., 2015). Still, evidence from stem cells indicates that global loss of chromatin loops has a minimal effect on gene expression (Zuin et al., 2014; Rao et al., 2017). Thus the retina-specific roles of CTCF likely reflect gene-specific regulatory functions independent of 3D genome structure, although more work is needed to examine this idea.

Inverted Nuclear Architecture in Mouse Rod Photoreceptors

The chromatin spatial architecture is commonly shared among animal nuclei, where inactive heterochromatin is preferentially sequestered to the nuclear periphery while active euchromatin occupies the nuclear interior (Holla et al., 2020; Solovei et al., 2016). The structure of rod

photoreceptor nuclei in nocturnal animals has deviated from this organization: heterochromatin is densely concentrated in the nuclear center while euchromatin occupies the outer edges (**Figure 2**) (Solovei et al., 2009). Data suggest that the inverted nuclear arrangement in rods reduces light scattering, effectively converting the nuclei into micro-lenses that enhance vision in dim light conditions (Solovei et al., 2009; Solovei et al., 2013; Subramanian et al., 2019). As such, this inverted nuclear structure in rods represents a clear example of how 3D nuclear architecture may directly influence a physiological function. Still, inverted nuclei structure is also observed in other cell types such as olfactory sensory neurons and neutrophils but the exact biological purpose of this organization in these cells is not clear (Clowney et al., 2012; Solovei et al., 2013).

Despite the stark structural differences between inverted and conventional nuclei, Hi-C data indicate that the hierarchical chromatin compartmentalization is qualitatively similar (Falk et al., 2019; Tan et al., 2019). Additionally, studies integrating Hi-C experiments with computational modeling suggest that the spatial partitioning of heterochromatin and euchromatin in both conventional and inverted nuclei is mediated by liquid-phase separation dynamics, driven primarily by heterochromatin interactions (Falk et al., 2019; Tan et al., 2019).

The establishment of inverted nuclei occurs during rod photoreceptors terminal differentiation and is completed by postnatal day 28 in mice (Solovei et al., 2009). During this process, rod precursor nuclei experience morphological reorganization where chromocenters gradually dissociate from the nuclear periphery and coalesce centrally (Solovei et al., 2009). At the molecular level, nuclear inversion is correlated with loss of LBR and Lamin A/C, proteins essential for tethering heterochromatin to the nuclear periphery (Clowney et al., 2012; Solovei et al., 2013). The molecular mechanism involving downregulation of lamina-associated proteins during rod differentiation has not been fully explored but preliminary evidence suggests a role for the transcription factor Casz1 in association with polycomb proteins in repressing Lamin A (Mattar et al., 2018). Casz1 is also expressed in cone photoreceptors and does not seem to regulate LBR expression (Mattar et al., 2018). Thus, it is likely that repression of lamina-associated proteins in differentiating rods involves other rod-specific transcription factors (Hughes et al., 2017; Mattar et al., 2018). Interestingly, while loss of LBR can alter the nuclear structure, it does not affect global gene expression (Solovei et al., 2013; Norrie et al., 2019).

CONCLUDING MARKS

Genomic studies thus far have provided insights into modulation of retinal development by chromatin structure, yet the field is still in its infancy and a tremendous amount of work is needed to gain a comprehensive understanding on how epigenetics shape retinal development and are associated with retinal diseases. As sequencing technologies and computational analyses continue to rapidly evolve, it is likely that more high resolution data from retinal cell types will be available in the near future.

What are the long-range interactions that occur among cis-regulatory elements during retinal development and how essential are they to retinal development and homeostasis? Are these interactions disrupted in ocular diseases? If so in what way? What are the factors that govern nuclear organization in retinal neurons? How does nuclear architecture influence gene expression during retinal cell type specification? Do liquid-phase separation properties of nuclear compartments influence retinal transcriptional programs? These are some of the outstanding questions that are likely to help elucidating how chromatin influence transcriptional regulation in the retina.

Animal models have been immensely valuable in understanding molecular mechanisms underlying human biology and diseases but more studies investigating chromatin structure in human native and diseased retina are needed. This is particularly important to advance therapeutic strategies aiming at stimulating regeneration and/or preventing degeneration in the mammalian retina.

AUTHOR CONTRIBUTIONS

Both authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Adam, R. C., Yang, H., Rockowitz, S., Larsen, S. B., Nikolova, M., Oristian, D. S., et al. (2015). Pioneer Factors Govern Super-enhancer Dynamics in Stem Cell Plasticity and Lineage Choice. *Nature* 521, 366–370. doi:10.1038/nature14289

Agathocleous, M., and Harris, W. A. (2009). From Progenitors to Differentiated Cells in the Vertebrate Retina. *Annu. Rev. Cel Dev. Biol.* 25, 45–69. doi:10.1146/annurev.cellbio.042308.113259

Aldiri, I., Ajioka, I., Xu, B., Zhang, J., Chen, X., Benavente, C., et al. (2015). Brg1 Coordinates Multiple Processes During Retinogenesis and Is a Tumor Suppressor in Retinoblastoma. *Development* 142, 4092–4106. doi:10.1242/dev.124800

Aldiri, I., Moore, K. B., Hutcheson, D. A., Zhang, J., and Vetter, M. L. (2013). Polycomb Repressive Complex PRC2 Regulates Xenopus Retina Development Downstream of Wnt/β-Catenin Signaling. *Development* 140, 2867–2878. doi:10.1242/dev.088096

Aldiri, I., and Vetter, M. L. (2009). Characterization of the Expression Pattern of the PRC2 Core subunitSuz12during Embryonic Development ofXenopus Laevis. *Dev. Dyn.* 238, 3185–3192. doi:10.1002/dvdy.22120

Aldiri, I., Xu, B., Wang, L., Chen, X., Hiler, D., Griffiths, L., et al. (2017). The Dynamic Epigenetic Landscape of the Retina During Development, Reprogramming, and Tumorigenesis. *Neuron* 94, 550–568. doi:10.1016/j.neuron.2017.04.022

Alver, B. H., Kim, K. H., Lu, P., Wang, X., Manchester, H. E., Wang, W., et al. (2017). The SWI/SNF Chromatin Remodelling Complex Is Required for Maintenance of Lineage Specific Enhancers. *Nat. Commun.* 8, 14648. doi:10.1038/ncomms14648

Andzelm, M. M., Cherry, T. J., Harmin, D. A., Boeke, A. C., Lee, C., Hemberg, M., et al. (2015). MEF2D Drives Photoreceptor Development Through a Genome-wide Competition for Tissue-specific Enhancers. *Neuron* 86, 247–263. doi:10.1016/j.neuron.2015.02.038

Bahr, C., von Paleske, L., Uslu, V. V., Remeseiro, S., Takayama, N., Ng, S. W., et al. (2018). A Myc Enhancer Cluster Regulates Normal and Leukaemic Haematopoietic Stem Cell Hierarchies. *Nature* 553, 515–520. doi:10.1038/nature25193

Barutcu, A. R., Lajoie, B. R., Fritz, A. J., McCord, R. P., Nickerson, J. A., van Wijnen, A. J., et al. (2016). SMARCA4 Regulates Gene Expression and Higher-Order Chromatin Structure in Proliferating Mammary Epithelial Cells. *Genome Res.* 26, 1188–1201. doi:10.1101/gr.201624.115

Benabdallah, N. S., Williamson, I., Illingworth, R. S., Kane, L., Boyle, S., Sengupta, D., et al. (2019). Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Mol. Cel* 76, 473–484. doi:10.1016/j.molcel.2019.07.038

Bhansali, P., Cvekl, A., and Liu, W. (2020). A Distal Enhancer That Directs Otx2 Expression in the Retinal Pigment Epithelium and Neuroretina. *Dev. Dyn.* 249, 209–221. doi:10.1002/dvdy.127

Bhatia, S., Bengani, H., Fish, M., Brown, A., Divizia, M. T., de Marco, R., et al. (2013). Disruption of Autoregulatory Feedback by a Mutation in a Remote, Ultraconserved PAX6 Enhancer Causes Aniridia. *Am. J. Hum. Genet.* 93, 1126–1134. doi:10.1016/j.ajhg.2013.10.028

Boija, A., Klein, I. A., Sabari, B. R., Dall'Agnese, A., Coffey, E. L., Zamudio, A. V., et al. (2018). Transcription Factors Activate Genes Through the Phase-Separation Capacity of Their Activation Domains. *Cell* 175, 1842–1855. doi:10.1016/j.cell.2018.10.042

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G. L., Lubling, Y., et al. (2017). Multiscale 3D Genome Rewiring During Mouse Neural Development. *Cell* 171, 557–572. doi:10.1016/j.cell.2017.09.043

Brightman, D. S., Grant, R. L., Ruzycki, P. A., Suzuki, R., Hennig, A. K., and Chen, S. (2018). MLL1 Is Essential for Retinal Neurogenesis and Horizontal Inner Neuron Integrity. *Sci. Rep.* 8, 11902. doi:10.1038/s41598-018-30355-3

Brodie-Kommit, J., Clark, B. S., Shi, Q., Shiau, F., Kim, D. W., Langel, J., et al. (2021). Atoh7-independent Specification of Retinal Ganglion Cell Identity. *Sci. Adv.* 7, eabe4983. doi:10.1126/sciadv.abe4983

Buecker, C., and Wysocka, J. (2012). Enhancers as Information Integration Hubs in Development: Lessons from Genomics. *Trends Genet.* 28, 276–284. doi:10.1016/j.tig.2012.02.008

Centore, R. C., Sandoval, G. J., Soares, L. M. M., Kadoch, C., and Chan, H. M. (2020). Mammalian SWI/SNF Chromatin Remodeling Complexes: Emerging Mechanisms and Therapeutic Strategies. *Trends Genet.* 36, 936–950. doi:10.1016/j.tig.2020.07.011

Cepko, C. L., Austin, C. P., Yang, X., Alexiades, M., and Ezzeddine, D. (1996). Cell Fate Determination in the Vertebrate Retina. *Proc. Natl. Acad. Sci.* 93, 589–595. doi:10.1073/pnas.93.2.589

Chan, C. S. K., Lonfat, N., Zhao, R., Davis, A. E., Li, L., Wu, M.-R., et al. (2020). Cell Type- and Stage-specific Expression of Otx2 Is Regulated by Multiple Transcription Factors and Cis-Regulatory Modules in the Retina. *Development* 147, dev187922. doi:10.1242/dev.187922

Chatterjee, S., and Ahituv, N. (2017). Gene Regulatory Elements, Major Drivers of Human Disease. *Annu. Rev. Genom. Hum. Genet.* 18, 45–63. doi:10.1146/annurev-genom-091416-035537

Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J. B., and Gregor, T. (2018). Dynamic Interplay Between Enhancer-Promoter Topology and Gene Activity. *Nat. Genet.* 50, 1296–1303. doi:10.1038/s41588-018-0175-z

Cheng, L., Wong, L. J., Yan, N., Han, R. C., Yu, H., Guo, C., et al. (2018). Ezh2 Does Not Mediate Retinal Ganglion Cell Homeostasis or Their Susceptibility to Injury. *Plos One* 13, e0191853. doi:10.1371/journal.pone.0191853

Cherry, T. J., Yang, M. G., Harmin, D. A., Tao, P., Timms, A. E., Bauwens, M., et al. (2020). Mapping the Cis-Regulatory Architecture of the Human Retina Reveals Noncoding Genetic Variation in Disease. *Proc. Natl. Acad. Sci. USA* 117, 9001–9012. doi:10.1073/pnas.1922501117

Clark, B. S., Stein-O'Brien, G. L., Shiau, F., Cannon, G. H., Davis-Marcisak, E., Sherman, T., et al. (2019). Single-Cell RNA-Seq Analysis of Retinal Development Identifies NFI Factors as Regulating Mitotic Exit and Late-Born Cell Specification. *Neuron* 102, 1111–1126. doi:10.1016/j.neuron.2019.04.010

Clowney, E. J., LeGros, M. A., Mosley, C. P., Clowney, F. G., Markenskoff-Papadimitriou, E. C., Myllys, M., et al. (2012). Nuclear Aggregation of Olfactory Receptor Genes Governs Their Monogenic Expression. *Cell* 151, 724–737. doi:10.1016/j.cell.2012.09.043

Corbo, J. C., Lawrence, K. A., Karlstetter, M., Myers, C. A., Abdelaziz, M., Dirkes, W., et al. (2010). CRX ChIP-Seq Reveals the Cis-Regulatory Architecture of Mouse Photoreceptors. *Genome Res.* 20, 1512–1525. doi:10.1101/gr.109405.110

Corso-Díaz, X., Jaeger, C., Chaitankar, V., and Swaroop, A. (2018). Epigenetic Control of Gene Regulation During Development and Disease: A View from the Retina. *Prog. Retin. Eye Res.* 65, 1–27. doi:10.1016/j.preteyeres.2018.03.002

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., et al. (2010). Histone H3K27ac Separates Active From Poised Enhancers and Predicts Developmental State. *Proc. Natl. Acad. Sci.* 107, 21931–21936. doi:10.1073/pnas.1016071107

Crump, N. T., Ballabio, E., Godfrey, L., Thorne, R., Repapi, E., Kerry, J., et al. (2021). BET Inhibition Disrupts Transcription but Retains Enhancer-Promoter Contact. *Nat. Commun.* 12, 223. doi:10.1038/s41467-020-20400-z

Cruz-Molina, S., Respuela, P., Tebartz, C., Kolovos, P., Nikolic, M., Fueyo, R., et al. (2017). PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function During Pluripotent Stem Cell Differentiation. *Cell Stem Cell* 20, 689–705. doi:10.1016/j.stem.2017.02.004

Das, A. V., James, J., Bhattacharya, S., Imbalzano, A. N., Antony, M. L., Hegde, G., et al. (2007). SWI/SNF Chromatin Remodeling ATPase Brm Regulates the Differentiation of Early Retinal Stem Cells/progenitors by Influencing Brn3b Expression and Notch Signaling. *J. Biol. Chem.* 282, 35187–35201. doi:10.1074/jbc.M706742200

de Bruijn, S. E., Fiorentino, A., Ottaviani, D., Fanucchi, S., Melo, U. S., Corral-Serrano, J. C., et al. (2020). Structural Variants Create New Topological-Associated Domains and Ectopic Retinal Enhancer-Gene Contact in Dominant Retinitis Pigmentosa. *Am. J. Hum. Genet.* 107, 802–814. doi:10.1016/j.ajhg.2020.09.002

de Laat, W., and Duboule, D. (2013). Topology of Mammalian Developmental Enhancers and Their Regulatory Landscapes. *Nature* 502, 499–506. doi:10.1038/nature12753

Demb, J. B., and Singer, J. H. (2015). Functional Circuitry of the Retina. *Annu. Rev. Vis. Sci.* 1, 263–289. doi:10.1146/annurev-vision-082114-035334

de Wit, E., Vos, E. S. M., Holwerda, S. J. B., Valdes-Quezada, C., Verstegen, M. J. A. M., Teunissen, H., et al. (2015). CTCF Binding Polarity Determines Chromatin Looping. *Mol. Cel* 60, 676–684. doi:10.1016/j.molcel.2015.09.023

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012). Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. *Nature* 485, 376–380. doi:10.1038/nature11082

Drouin, J. (2014). Minireview: Pioneer Transcription Factors in Cell Fate Specification. *Mol. Endocrinol.* 28, 989–998. doi:10.1210/me.2014-1084

Eiraku, M., Takata, N., Ishibashi, H., Kawada, M., Sakakura, E., Okuda, S., et al. (2011). Self-organizing Optic-Cup Morphogenesis in Three-Dimensional Culture. *Nature* 472, 51–56. doi:10.1038/nature09941

Emerson, M. M., and Cepko, C. L. (2011). Identification of a Retina-specific Otx2 Enhancer Element Active in Immature Developing Photoreceptors. *Dev. Biol.* 360, 241–255. doi:10.1016/j.ydbio.2011.09.012

Emerson, M. M., Surzenko, N., Goetz, J. J., Trimarchi, J., and Cepko, C. L. (2013). Otx2 and Onecut1 Promote the Fates of Cone Photoreceptors and Horizontal Cells and Repress Rod Photoreceptors. *Dev. Cel* 26, 59–72. doi:10.1016/j.devcel.2013.06.005

Ernst, J., and Kellis, M. (2012). ChromHMM: Automating Chromatin-State Discovery and Characterization. *Nat. Methods* 9, 215–216. doi:10.1038/nmeth.1906

Ernst, J., and Kellis, M. (2010). Discovery and Characterization of Chromatin States for Systematic Annotation of the Human Genome. *Nat. Biotechnol.* 28, 817–825. doi:10.1038/nbt.1662

Falk, M., Feodorova, Y., Naumova, N., Imakaev, M., Lajoie, B. R., Leonhardt, H., et al. (2019). Heterochromatin Drives Compartmentalization of Inverted and Conventional Nuclei. *Nature* 570, 395–399. doi:10.1038/s41586-019-1275-3

Fang, R., Yu, M., Li, G., Chee, S., Liu, T., Schmitt, A. D., et al. (2016). Mapping of Long-Range Chromatin Interactions by Proximity Ligation-Assisted ChIP-Seq. *Cell Res* 26, 1345–1348. doi:10.1038/cr.2016.137

Fang, W., Guo, C., and Wei, X. (2017). Rainbow Enhancers Regulate Restrictive Transcription in Teleost Green, Red, and Blue Cones. *J. Neurosci.* 37, 2834–2848. doi:10.1523/JNEUROSCI.3421-16.2017

Fisher, L. J. (1979). Development of Synaptic Arrays in the Inner Plexiform Layer of Neonatal Mouse Retina. *J. Comp. Neurol.* 187, 359–372. doi:10.1002/cne.901870207

Fu, Z., Kern, T. S., Hellström, A., and Smith, L. E. H. (2021). Fatty Acid Oxidation and Photoreceptor Metabolic Needs. *J. Lipid Res.* 62, 100035. doi:10.1194/jlr.TR120000618

Fujimura, N., Kuzelova, A., Ebert, A., Strnad, H., Lachova, J., Machon, O., et al. (2018). Polycomb Repression Complex 2 Is Required for the Maintenance of Retinal Progenitor Cells and Balanced Retinal Differentiation. *Dev. Biol.* 433, 47–60. doi:10.1016/j.ydbio.2017.11.004

Gasperini, M., Tome, J. M., and Shendure, J. (2020). Towards a Comprehensive Catalogue of Validated and Target-Linked Human Enhancers. *Nat. Rev. Genet.* 21, 292–310. doi:10.1038/s41576-019-0209-0

Ghiasvand, N. M., Rudolph, D. D., Mashayekhi, M., Brzezinski, J. A., Goldman, D., and Glaser, T. (2011). Deletion of a Remote Enhancer Near ATOH7 Disrupts Retinal Neurogenesis, Causing NCRNA Disease. *Nat. Neurosci.* 14, 578–586. doi:10.1038/nn.2798

Goodson, N. B., Kaufman, M. A., Park, K. U., and Brzezinski, J. A. (2020). Simultaneous Deletion of Prdm1 and Vsx2 Enhancers in the Retina Alters Photoreceptor and Bipolar Cell Fate Specification, Yet Differs from Deleting Both Genes. *Development* 147. doi:10.1242/dev.190272

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D. U., et al. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* 162, 900–910. doi:10.1016/j.cell.2015.07.038

Hafler, B. P., Surzenko, N., Beier, K. T., Punzo, C., Trimarchi, J. M., Kong, J. H., et al. (2012). Transcription Factor Olig2 Defines Subpopulations of Retinal Progenitor Cells Biased Toward Specific Cell Fates. *Proc. Natl. Acad. Sci.* 109, 7882–7887. doi:10.1073/pnas.1203138109

Halfon, M. S. (2019). Studying Transcriptional Enhancers: The Founder Fallacy, Validation Creep, and Other Biases. *Trends Genet.* 35, 93–103. doi:10.1016/j.tig.2018.11.004

Hiler, D., Chen, X., Hazen, J., Kupriyanov, S., Carroll, P. A., Qu, C., et al. (2015). Quantification of Retinogenesis in 3D Cultures Reveals Epigenetic Memory and Higher Efficiency in iPSCs Derived from Rod Photoreceptors. *Cell Stem Cell* 17, 101–115. doi:10.1016/j.stem.2015.05.015

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., et al. (2013). Super-enhancers in the Control of Cell Identity and Disease. *Cell* 155, 934–947. doi:10.1016/j.cell.2013.09.053

Hnisz, D., Day, D. S., and Young, R. A. (2016). Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. *Cell* 167, 1188–1200. doi:10.1016/j.cell.2016.10.024

Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K., and Sharp, P. A. (2017). A Phase Separation Model for Transcriptional Control. *Cell* 169, 13–23. doi:10.1016/j.cell.2017.02.007

Holla, S., Dhakshnamoorthy, J., Folco, H. D., Balachandran, V., Xiao, H., Sun, L.-l., et al. (2020). Positioning Heterochromatin at the Nuclear Periphery Suppresses Histone Turnover to Promote Epigenetic Inheritance. *Cell* 180, 150–164. doi:10.1016/j.cell.2019.12.004

Hughes, A. E., Enright, J. M., Myers, C. A., Shen, S. Q., and Corbo, J. C. (2017). Cell Type-specific Epigenomic Analysis Reveals a Uniquely Closed Chromatin Architecture in Mouse Rod Photoreceptors. *Sci. Rep.* 7, 43184. doi:10.1038/srep43184

Hutcheson, D. A., Hanson, M. I., Moore, K. B., Le, T. T., Brown, N. L., and Vetter, M. L. (2005). bHLH-Dependent and -independent Modes of Ath5 Gene Regulation during Retinal Development. *Development* 132, 829–839. doi:10.1242/dev.01653

Iida, A., Iwagawa, T., Baba, Y., Satoh, S., Mochizuki, Y., Nakauchi, H., et al. (2015). Roles of Histone H3K27 Trimethylase Ezh2 in Retinal Proliferation and Differentiation. *Devel Neurobio* 75, 947–960. doi:10.1002/dneu.22261

Iida, A., Iwagawa, T., Kuribayashi, H., Satoh, S., Mochizuki, Y., Baba, Y., et al. (2014). Histone Demethylase Jmjd3 Is Required for the Development of Subsets of Retinal Bipolar Cells. *Proc. Natl. Acad. Sci. USA* 111, 3751–3756. doi:10.1073/pnas.1311480111

Kadoch, C., and Crabtree, G. R. (2015). Mammalian SWI/SNF Chromatin Remodeling Complexes and Cancer: Mechanistic Insights Gained from Human Genomics. *Sci. Adv.* 1, e1500447. doi:10.1126/sciadv.1500447

Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., van Berkum, N. L., et al. (2011). Erratum: Mediator and Cohesin Connect Gene Expression and Chromatin Architecture. *Nature* 472, 247. doi:10.1038/nature09930

Kashima, Y., Sakamoto, Y., Kaneko, K., Seki, M., Suzuki, Y., and Suzuki, A. (2020). Single-cell Sequencing Techniques from Individual to Multiomics Analyses. *Exp. Mol. Med.* 52, 1419–1427. doi:10.1038/s12276-020-00499-2

Kaufman, M. L., Goodson, N. B., Park, K. U., Schwanke, M., Office, E., Schneider, S. R., et al. (2021). Initiation of Otx2 Expression in the Developing Mouse Retina Requires a Unique Enhancer and Either Ascl1 or Neurog2 Activity. *Development* 148. doi:10.1242/dev.199399

Kautzmann, M.-A. I., Kim, D. S., Felder-Schmittbuhl, M.-P., and Swaroop, A. (2011). Combinatorial Regulation of Photoreceptor Differentiation Factor, Neural Retina Leucine Zipper Gene NRL, Revealed by In Vivo Promoter Analysis. *J. Biol. Chem.* 286, 28247–28255. doi:10.1074/jbc.M111.257246

Keser, V., Khan, A., Siddiqui, S., Lopez, I., Ren, H., Qamar, R., et al. (2017). The Genetic Causes of Nonsyndromic Congenital Retinal Detachment: A Genetic and Phenotypic Study of Pakistani Families. *Invest. Ophthalmol. Vis. Sci.* 58, 1028. doi:10.1167/iovs.16-20281

Kim, D. S., Matsuda, T., and Cepko, C. L. (2008). A Core Paired-type and POU Homeodomain-Containing Transcription Factor Program Drives Retinal Bipolar Cell Gene Expression. *J. Neurosci.* 28, 7748–7764. doi:10.1523/JNEUROSCI.0397-08.2008

Kim, S., Lowe, A., Dharmat, R., Lee, S., Owen, L. A., Wang, J., et al. (2019). Generation, Transcriptome Profiling, and Functional Validation of Cone-Rich Human Retinal Organoids. *Proc. Natl. Acad. Sci. USA* 116, 10824–10833. doi:10.1073/pnas.1901572116

Kleinjan, D. A., Seawright, A., Childs, A. J., and van Heyningen, V. (2004). Conserved Elements in Pax6 Intron 7 Involved in (Auto)regulation and Alternative Transcription. *Dev. Biol.* 265, 462–477. doi:10.1016/j.ydbio.2003.09.011

Kondo, H., Matsushita, I., Tahira, T., Uchio, E., and Kusaka, S. (2018). Mutations in ATOH7 Gene in Patients with Nonsyndromic Congenital Retinal Nonattachment and Familial Exudative Vitreoretinopathy (Vol 37, Pg 462, 2016). *Ophthalmic Genet.* 39, 553. doi:10.1080/13816810.2017.1401090

Kurokawa, D., Kiyonari, H., Nakayama, R., Kimura-Yoshida, C., Matsuo, I., and Aizawa, S. (2004). Regulation of Otx2 Expression and its Functions in Mouse Forebrain and Midbrain. *Development* 131, 3319–3331. doi:10.1242/dev.01220

Kvon, E. Z., Kamneva, O. K., Melo, U. S., Barozzi, I., Osterwalder, M., Mannion, B. J., et al. (2016). Progressive Loss of Function in a Limb Enhancer During Snake Evolution. *Cell* 167, 633–642. doi:10.1016/j.cell.2016.09.028

Lamba, D. A., Hayes, S., Karl, M. O., and Reh, T. (2008). Baf60c Is a Component of the Neural Progenitor-specific BAF Complex in Developing Retina. *Dev. Dyn.* 237, 3016–3023. doi:10.1002/dvdy.21697

Lessard, J., Wu, J. I., Ranish, J. A., Wan, M., Winslow, M. M., Staahl, B. T., et al. (2007). An Essential Switch in Subunit Composition of a Chromatin Remodeling Complex during Neural Development. *Neuron* 55, 201–215. doi:10.1016/j.neuron.2007.06.019

Li, G., Ruan, X., Auerbach, R. K., Sandhu, K. S., Zheng, M., Wang, P., et al. (2012). Extensive Promoter-Centered Chromatin Interactions Provide a Topological Basis for Transcription Regulation. *Cell* 148, 84–98. doi:10.1016/j.cell.2011.12.014

Li, T., Lu, Z., and Lu, L. (2006). Pax6 Regulation in Retinal Cells by CCCTC Binding Factor. *Invest. Ophthalmol. Vis. Sci.* 47, 5218–5226. doi:10.1167/iovs.06-0254

Li, Y., Rivera, C. M., Ishii, H., Jin, F., Selvaraj, S., Lee, A. Y., et al. (2014). CRISPR Reveals a Distal Super-enhancer Required for Sox2 Expression in Mouse Embryonic Stem Cells. *Plos One* 9, e114485. doi:10.1371/journal.pone.0114485

Liang, Q., Dharmat, R., Owen, L., Shakoor, A., Li, Y., Kim, S., et al. (2019). Single-nuclei RNA-Seq on Human Retinal Tissue Provides Improved Transcriptome Profiling. *Nat. Commun.* 10, 5743. doi:10.1038/s41467-019-12917-9

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* 326, 289–293. doi:10.1126/science.1181369

Livesey, F. J., and Cepko, C. L. (2001). Vertebrate Neural Cell-Fate Determination: Lessons from the Retina. *Nat. Rev. Neurosci.* 2, 109–118. doi:10.1038/35053522

Long, H. K., Prescott, S. L., and Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* 167, 1170–1187. doi:10.1016/j.cell.2016.09.018

Lovén, J., Hoke, H. A., Lin, C. Y., Lau, A., Orlando, D. A., Vakoc, C. R., et al. (2013). Selective Inhibition of Tumor Oncogenes by Disruption of Super-enhancers. *Cell* 153, 320–334. doi:10.1016/j.cell.2013.03.036

Lu, Y., Shiau, F., Yi, W., Lu, S., Wu, Q., Pearson, J. D., et al. (2020). Single-Cell Analysis of Human Retina Identifies Evolutionarily Conserved and Species-specific Mechanisms Controlling Development. *Dev. Cel* 53, 473–491. doi:10.1016/j.devcel.2020.04.009

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal Structure of the Nucleosome Core Particle at 2.8 Å Resolution. *Nature* 389, 251–260. doi:10.1038/38444

Marquardt, T., and Gruss, P. (2002). Generating Neuronal Diversity in the Retina: One for Nearly All. *Trends Neurosciences* 25, 32–38. doi:10.1016/s0166-2236(00)02028-2

Mattar, P., Stevanovic, M., Nad, I., and Cayouette, M. (2018). Casz1 Controls Higher-Order Nuclear Organization in Rod Photoreceptors. *Proc. Natl. Acad. Sci. USA* 115, E7987–E7996. doi:10.1073/pnas.1803069115

Menon, M., Mohammadi, S., Davila-Velderrain, J., Goods, B. A., Cadwell, T. D., Xing, Y., et al. (2019). Single-cell Transcriptomic Atlas of the Human Retina Identifies Cell Types Associated with Age-Related Macular Degeneration. *Nat. Commun.* 10, 4902. doi:10.1038/s41467-019-12780-8

Merkenschlager, M., and Nora, E. P. (2016). CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annu. Rev. Genom. Hum. Genet.* 17, 17–43. doi:10.1146/annurev-genom-083115-022339

Miesfeld, J. B., Ghiasvand, N. M., Marsh-Armstrong, B., Marsh-Armstrong, N., Miller, E. B., Zhang, P., et al. (2020). TheAtoh7remote Enhancer Provides Transcriptional Robustness During Retinal Ganglion Cell Development. *Proc. Natl. Acad. Sci. USA* 117, 21690–21700. doi:10.1073/pnas.2006888117

Mills, T. S., Eliseeva, T., Bersie, S. M., Randazzo, G., Nahreini, J., Park, K. U., et al. (2017). Combinatorial Regulation of a Blimp1 (Prdm1) Enhancer in the Mouse Retina. *Plos One* 12, e0176905. doi:10.1371/journal.pone.0176905

Mo, A., Luo, C., Davis, F. P., Mukamel, E. A., Henry, G. L., Nery, J. R., et al. (2016). Epigenomic Landscapes of Retinal Rods and Cones. *Elife* 5, e11613. doi:10.7554/eLife.11613

Montana, C. L., Myers, C. A., and Corbo, J. C. (2011). Quantifying the Activity of Cis-Regulatory Elements in the Mouse Retina by Explant Electroporation. *J. Vis. Exp.* 10, 3791. doi:10.3791/2821

Moorthy, S. D., Davidson, S., Shchuka, V. M., Singh, G., Malek-Gilani, N., Langroudi, L., et al. (2017). Enhancers and Super-enhancers Have an Equivalent Regulatory Role in Embryonic Stem Cells through Regulation of Single or Multiple Genes. *Genome Res.* 27, 246–258. doi:10.1101/gr.210930.116

Mumbach, M. R., Rubin, A. J., Flynn, R. A., Dai, C., Khavari, P. A., Greenleaf, W. J., et al. (2016). HiChIP: Efficient and Sensitive Analysis of Protein-Directed Genome Architecture. *Nat. Methods* 13, 919–922. doi:10.1038/nmeth.3999

Murphy, D. P., Hughes, A. E., Lawrence, K. A., Myers, C. A., and Corbo, J. C. (2019). Cis-regulatory Basis of Sister Cell Type Divergence in the Vertebrate Retina. *Elife* 8, e48216. doi:10.7554/eLife.48216

Ninkovic, J., Steiner-Mezzadri, A., Jawerka, M., Akinci, U., Masserdotti, G., Petricca, S., et al. (2013). The BAF Complex Interacts with Pax6 in Adult Neural Progenitors to Establish a Neurogenic Cross-Regulatory Transcriptional Network. *Cell Stem Cell* 13, 403–418. doi:10.1016/j.stem.2013.07.002

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., et al. (2012). Spatial Partitioning of the Regulatory Landscape of the X-Inactivation Centre. *Nature* 485, 381–385. doi:10.1038/nature11049

Norrie, J. L., Lupo, M. S., Xu, B., Al Diri, I., Valentine, M., Putnam, D., et al. (2019). Nucleome Dynamics During Retinal Development. *Neuron* 104, 512–528. doi:10.1016/j.neuron.2019.08.002

Ohsawa, R., and Kageyama, R. (2008). Regulation of Retinal Cell Fate Specification by Multiple Transcription Factors. *Brain Res.* 1192, 90–98. doi:10.1016/j.brainres.2007.04.014

Osterwalder, M., Barozzi, I., Tissières, V., Fukuda-Yuzawa, Y., Mannion, B. J., Afzal, S. Y., et al. (2018). Enhancer Redundancy Provides Phenotypic Robustness in Mammalian Development. *Nature* 554, 239–243. doi:10.1038/nature25461

Parker, S. C. J., Stitzel, M. L., Taylor, D. L., Orozco, J. M., Erdos, M. R., Akiyama, J. A., et al. (2013). Chromatin Stretch Enhancer States Drive Cell-specific Gene Regulation and Harbor Human Disease Risk Variants. *Proc. Natl. Acad. Sci.* 110, 17921–17926. doi:10.1073/pnas.1317023110

Partha, R., Chauhan, B. K., Ferreira, Z., Robinson, J. D., Lathrop, K., Nischal, K. K., et al. (2017). Subterranean Mammals Show Convergent Regression in Ocular Genes and Enhancers, Along with Adaptation to Tunneling. *Elife* 6, e25884. doi:10.7554/eLife.25884

Patoori, S., Jean-Charles, N., Gopal, A., Sulaiman, S., Gopal, S., Wang, B., et al. (2020). Cis-regulatory Analysis of Onecut1 Expression in Fate-Restricted Retinal Progenitor Cells. *Neural Dev.* 15, 5. doi:10.1186/s13064-020-00142-w

Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A., and Bejerano, G. (2013). Enhancers: Five Essential Questions. *Nat. Rev. Genet.* 14, 288–295. doi:10.1038/nrg3458

Perez-Cervantes, C., Smith, L. A., Nadadur, R. D., Hughes, A. E. O., Wang, S., Corbo, J. C., et al. (2020). Enhancer Transcription Identifies Cis-Regulatory Elements for Photoreceptor Cell Types. *Development* 147. doi:10.1242/dev.184432

Popova, E. Y., Xu, X., DeWan, A. T., Salzberg, A. C., Berg, A., Hoh, J., et al. (2012). Stage and Gene Specific Signatures Defined by Histones H3K4me2 and H3K27me3 Accompany Mammalian Retina Maturation In Vivo. *Plos One* 7, e46867. doi:10.1371/journal.pone.0046867

Pott, S., and Lieb, J. D. (2015). What Are Super-enhancers. *Nat. Genet.* 47, 8–12. doi:10.1038/ng.3167

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., and Wysocka, J. (2011). A Unique Chromatin Signature Uncovers Early Developmental Enhancers in Humans. *Nature* 470, 279–283. doi:10.1038/nature09692

Raeisossadati, R., Ferrari, M. F. R., Kihara, A. H., AlDiri, I., and Gross, J. M. (2021). Epigenetic Regulation of Retinal Development. *Epigenetics & Chromatin* 14, 11. doi:10.1186/s13072-021-00384-w

Rao, S. S. P., Huang, S.-C., Glenn St Hilaire, B., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K.-R., et al. (2017). Cohesin Loss Eliminates All Loop Domains. *Cell* 171, 305, 320.e24. doi:10.1016/j.cell.2017.09.026

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* 159, 1665–1680. doi:10.1016/j.cell.2014.11.021

Reis, L. M., and Semina, E. V. (2015). Conserved Genetic Pathways Associated with Microphthalmia, Anophthalmia, and Coloboma. *Birth Defect Res. C* 105, 96–113. doi:10.1002/bdrc.21097

Riesenberg, A. N., Le, T. T., Willardsen, M. I., Blackburn, D. C., Vetter, M. L., and Brown, N. L. (2009). Pax6regulation ofMath5during Mouse Retinal Neurogenesis. *Genesis* 47, 175–187. doi:10.1002/dvg.20479

Robson, M. I., Ringel, A. R., and Mundlos, S. (2019). Regulatory Landscaping: How Enhancer-Promoter Communication Is Sculpted in 3D. *Mol. Cel* 74, 1110–1122. doi:10.1016/j.molcel.2019.05.032

Roscito, J. G., Sameith, K., Parra, G., Langer, B. E., Petzold, A., Moebius, C., et al. (2018). Phenotype Loss Is Associated with Widespread Divergence of the Gene Regulatory Landscape in Evolution. *Nat. Commun.* 9, 4737. doi:10.1038/s41467-018-07122-z

Rowan, S., and Cepko, C. L. (2005). A POU Factor Binding Site Upstream of the Chx10 Homeobox Gene Is Required for Chx10 Expression in Subsets of Retinal Progenitor Cells and Bipolar Cells. *Dev. Biol.* 281, 240–255. doi:10.1016/j.ydbio.2005.02.023

Rowan, S., and Cepko, C. L. (2004). Genetic Analysis of the Homeodomain Transcription Factor Chx10 in the Retina Using a Novel Multifunctional BAC Transgenic Mouse Reporter. *Dev. Biol.* 271, 388–402. doi:10.1016/j.ydbio.2004.03.039

Ruzycki, P. A., Zhang, X., and Chen, S. (2018). CRX Directs Photoreceptor Differentiation by Accelerating Chromatin Remodeling at Specific Target Sites. *Epigenetics & Chromatin* 11, 42. doi:10.1186/s13072-018-0212-2

Sabari, B. R., Dall'Agnese, A., Boija, A., Klein, I. A., Coffey, E. L., Shrinivas, K., et al. (2018). Coactivator Condensation at Super-enhancers Links Phase Separation and Gene Control. *Science* 361, eaar3958. doi:10.1126/science.aar3958

Samuel, A., Housset, M., Fant, B., and Lamonerie, T. (2014). Otx2 ChIP-Seq Reveals Unique and Redundant Functions in the Mature Mouse Retina. *Plos One* 9, e89110. doi:10.1371/journal.pone.0089110

Schoenfelder, S., and Fraser, P. (2019). Long-range Enhancer-Promoter Contacts in Gene Expression Control. *Nat. Rev. Genet.* 20, 437–455. doi:10.1038/s41576-019-0128-0

Schoenfelder, S., Sugar, R., Dimond, A., Javierre, B.-M., Armstrong, H., Mifsud, B., et al. (2015). Polycomb Repressive Complex PRC1 Spatially Constrains the Mouse Embryonic Stem Cell Genome. *Nat. Genet.* 47, 1179–1186. doi:10.1038/ng.3393

Seo, S., Richardson, G. A., and Kroll, K. L. (2005). The SWI/SNF Chromatin Remodeling Protein Brg1 Is Required for Vertebrate Neurogenesis and Mediates Transactivation of Ngn and NeuroD. *Development* 132, 105–115. doi:10.1242/dev.01548

Slavotinek, A. M. (2011). Eye Development Genes and Known Syndromes. *Mol. Genet. Metab.* 104, 448–456. doi:10.1016/j.ymgme.2011.09.029

Solovei, I., Kreysing, M., Lanctôt, C., Kösem, S., Peichl, L., Cremer, T., et al. (2009). Nuclear Architecture of Rod Photoreceptor Cells Adapts to Vision in Mammalian Evolution. *Cell* 137, 356–368. doi:10.1016/j.cell.2009.01.052

Solovei, I., Thanisch, K., and Feodorova, Y. (2016). How to rule the nucleus: divide et impera. *Curr. Opin. Cell Biol.* 40, 47–59. doi:10.1016/j.ceb.2016.02.014

Solovei, I., Wang, A. S., Thanisch, K., Schmidt, C. S., Krebs, S., Zwerger, M., et al. (2013). LBR and Lamin A/C Sequentially Tether Peripheral Heterochromatin and Inversely Regulate Differentiation. *Cell* 152, 584–598. doi:10.1016/j.cell.2013.01.009

Soshnev, A. A., Josefowicz, S. Z., and Allis, C. D. (2016). Greater Than the Sum of Parts: Complexity of the Dynamic Epigenome. *Mol. Cel* 62, 681–694. doi:10.1016/j.molcel.2016.05.004

Sridhar, A., Hoshino, A., Finkbeiner, C. R., Chitsazan, A., Dai, L., Haugan, A. K., et al. (2020). Single-Cell Transcriptomic Comparison of Human Fetal Retina, hPSC-Derived Retinal Organoids, and Long-Term Retinal Cultures. *Cel Rep.* 30, 1644–1659. doi:10.1016/j.celrep.2020.01.007

Subramanian, K., Weigert, M., Borsch, O., Petzold, H., Garcia-Ulloa, A., Myers, E. W., et al. (2019). Rod Nuclear Architecture Determines Contrast Transmission of the Retina and Behavioral Sensitivity in Mice. *Elife* 8, e49542. doi:10.7554/eLife.49542

Swaroop, A., Kim, D., and Forrest, D. (2010). Transcriptional Regulation of Photoreceptor Development and Homeostasis in the Mammalian Retina. *Nat. Rev. Neurosci.* 11, 563–576. doi:10.1038/nrn2880

Tan, L., Xing, D., Daley, N., and Xie, X. S. (2019). Three-dimensional Genome Structures of Single Sensory Neurons in Mouse Visual and Olfactory Systems. *Nat. Struct. Mol. Biol.* 26, 297–307. doi:10.1038/s41594-019-0205-2

Taranova, O. V., Magness, S. T., Fagan, B. M., Wu, Y. Q., Surzenko, N., Hutton, S. R., et al. (2006). SOX2 Is a Dose-dependent Regulator of Retinal Neural Progenitor Competence. *Genes Dev.* 20, 1187–1202. doi:10.1101/gad.1407906

Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., et al. (2012). The Accessible Chromatin Landscape of the Human Genome. *Nature* 489, 75–82. doi:10.1038/nature11232

Turner, D. L., Snyder, E. Y., and Cepko, C. L. (1990). Lineage-independent Determination of Cell Type in the Embryonic Mouse Retina. *Neuron* 4, 833–845. doi:10.1016/0896-6273(90)90134-4

Ueno, K., Iwagawa, T., Kuribayashi, H., Baba, Y., Nakauchi, H., Murakami, A., et al. (2016). Transition of Differential Histone H3 Methylation in Photoreceptors and Other Retinal Cells during Retinal Differentiation. *Sci. Rep.* 6, 29264. doi:10.1038/srep29264

Ueno, K., Iwagawa, T., Ochiai, G., Koso, H., Nakauchi, H., Nagasaki, M., et al. (2017). Analysis of Müller Glia Specific Genes and Their Histone Modification Using Hes1-Promoter Driven EGFP Expressing Mouse. *Sci. Rep.* 7, 3578. doi:10.1038/s41598-017-03874-8

Van Schil, K., Karlstetter, M., Aslanidis, A., Dannhausen, K., Azam, M., Qamar, R., et al. (2016). Autosomal Recessive Retinitis Pigmentosa with Homozygous Rhodopsin Mutation E150K and Non-coding Cis-Regulatory Variants of CRX-Binding Regions of SAMD7. *Sci. Rep.* 6, 21307. doi:10.1038/srep21307

VandenBosch, L. S., Wohl, S. G., Wilken, M. S., Hooper, M., Finkbeiner, C., Cox, K., et al. (2020). Developmental Changes in the Accessible Chromatin, Transcriptome and Ascl1-Binding Correlate with the Loss in Müller Glial Regenerative Potential. *Sci. Rep.* 10, 13615. doi:10.1038/s41598-020-70334-1

Vignali, M., and Workman, J. L. (1998). Location and Function of Linker Histones. *Nat. Struct. Mol. Biol.* 5, 1025–1028. doi:10.1038/4133

Villaseñor, R., and Baubec, T. (2021). Regulatory Mechanisms Governing Chromatin Organization and Function. *Curr. Opin. Cel Biol.* 70, 10–17. doi:10.1016/j.ceb.2020.10.015

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., et al. (2009). ChIP-seq Accurately Predicts Tissue-specific Activity of Enhancers. *Nature* 457, 854–858. doi:10.1038/nature07730

Völkner, M., Zschätzsch, M., Rostovskaya, M., Overall, R. W., Busskamp, V., Anastassiadis, K., et al. (2016). Retinal Organoids from Pluripotent Stem Cells Efficiently Recapitulate Retinogenesis. *Stem Cel Rep.* 6, 525–538. doi:10.1016/j.stemcr.2016.03.001

Wang, J., Zibetti, C., Shang, P., Sripathi, S. R., Zhang, P., Cano, M., et al. (2018). ATAC-seq Analysis Reveals a Widespread Decrease of Chromatin Accessibility in Age-Related Macular Degeneration. *Nat. Commun.* 9, 1364. doi:10.1038/s41467-018-03856-y

Wang, S., Sengel, C., Emerson, M. M., and Cepko, C. L. (2014). A Gene Regulatory Network Controls the Binary Fate Decision of Rod and Bipolar Cells in the Vertebrate Retina. *Dev. Cel* 30, 513–527. doi:10.1016/j.devcel.2014.07.018

Watson, L. A., Wang, X., Elbert, A., Kernohan, K. D., Galjart, N., and Berube, N. G. (2014). Dual Effect of CTCF Loss on Neuroprogenitor Differentiation and Survival. *J. Neurosci.* 34, 2860–2870. doi:10.1523/JNEUROSCI.3769-13.2014

Weintraub, A. S., Li, C. H., Zamudio, A. V., Sigova, A. A., Hannett, N. M., Day, D. S., et al. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* 171, 1573–1588. doi:10.1016/j.cell.2017.11.008

Weir, K., Leavey, P., Santiago, C., and Blackshaw, S. (2021). Multiplexed Analysis of Retinal Gene Expression and Chromatin Accessibility Using scRNA-Seq and scATAC-Seq. *J. Vis. Exp.* (169). doi:10.3791/62239

Wetts, R., and Fraser, S. E. (1988). Multipotent Precursors Can Give Rise to All Major Cell Types of the Frog Retina. *Science* 239, 1142–1145. doi:10.1126/science.2449732

White, M. A., Myers, C. A., Corbo, J. C., and Cohen, B. A. (2013). Massively Parallel In Vivo Enhancer Assay Reveals that Highly Local Features Determine the Cis-Regulatory Function of ChIP-Seq Peaks. *Proc. Natl. Acad. Sci.* 110, 11952–11957. doi:10.1073/pnas.1307449110

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., et al. (2013). Master Transcription Factors and Mediator Establish Super-enhancers at Key Cell Identity Genes. *Cell* 153, 307–319. doi:10.1016/j.cell.2013.03.035

Wilken, M. S., Brzezinski, J. A., La Torre, A., Siebenthall, K., Thurman, R., Sabo, P., et al. (2015). DNase I Hypersensitivity Analysis of the Mouse Brain and Retina Identifies Region-specific Regulatory Elements. *Epigenetics & Chromatin* 8, 8. doi:10.1186/1756-8935-8-8

Willardsen, M. I., Suli, A., Pan, Y., Marsh-Armstrong, N., Chien, C.-B., El-Hodiri, H., et al. (2009). Temporal Regulation of Ath5 Gene Expression During Eye Development. *Dev. Biol.* 326, 471–481. doi:10.1016/j.ydbio.2008.10.046

Wilson, B. G., and Roberts, C. W. M. (2011). SWI/SNF Nucleosome Remodellers and Cancer. *Nat. Rev. Cancer* 11, 481–492. doi:10.1038/nrc3068

Wu, F., Bard, J. E., Kann, J., Yergeau, D., Sapkota, D., Ge, Y., et al. (2021). Single Cell Transcriptomics Reveals Lineage Trajectory of Retinal Ganglion Cells in Wild-type and Atoh7-Null Retinas. *Nat. Commun.* 12, 1465. doi:10.1038/s41467-021-21704-4

Xie, H., Zhang, W., Zhang, M., Akhtar, T., Li, Y., Yi, W., et al. (2020). Chromatin Accessibility Analysis Reveals Regulatory Dynamics of Developing Human Retina and hiPSC-Derived Retinal Organoids. *Sci. Adv.* 6, eaay5247. doi:10.1126/sciadv.aay5247

Yan, N., Cheng, L., Cho, K., Malik, M. T. A., Xiao, L., Guo, C., et al. (2016). Postnatal Onset of Retinal Degeneration by Loss of Embryonic Ezh2 Repression of Six1. *Sci. Rep.* 6, 33887. doi:10.1038/srep33887

Yan, W., Chen, D., Schumacher, J., Durantini, D., Engelhorn, J., Chen, M., et al. (2019). Dynamic Control of Enhancer Activity Drives Stage-specific Gene Expression During Flower Morphogenesis. *Nat. Commun.* 10, 1705. doi:10.1038/s41467-019-09513-2

Zaret, K. S., and Carroll, J. S. (2011). Pioneer Transcription Factors: Establishing Competence for Gene Expression. *Genes Dev.* 25, 2227–2241. doi:10.1101/gad.176826.111

Zhang, J., Taylor, R. J., La Torre, A., Wilken, M. S., Cox, K. E., Reh, T. A., et al. (2015). Ezh2 Maintains Retinal Progenitor Proliferation, Transcriptional

Integrity, and the Timing of Late Differentiation. *Dev. Biol.* 403, 128–138. doi:10.1016/j.ydbio.2015.05.010

Zhou, H. Y., Katsman, Y., Dhaliwal, N. K., Davidson, S., Macpherson, N. N., Sakthidevi, M., et al. (2014). ASox2distal Enhancer Cluster Regulates Embryonic Stem Cell Differentiation Potential. *Genes Dev.* 28, 2699–2711. doi:10.1101/gad.248526.114

Zibetti, C., Liu, S., Wan, J., Qian, J., and Blackshaw, S. (2019). Epigenomic Profiling of Retinal Progenitors Reveals LHX2 Is Required for Developmental Regulation of Open Chromatin. *Commun. Biol.* 2, 142. doi:10.1038/s42003-019-0375-9

Zuber, M. E., Gestri, G., Viczian, A. S., Barsacchi, G., and Harris, W. A. (2003). Specification of the Vertebrate Eye by a Network of Eye Field Transcription Factors. *Development* 130, 5155–5167. doi:10.1242/dev.00723

Zuin, J., Dixon, J. R., van der Reijden, M. I. J. A., Ye, Z., Kolovos, P., Brouwer, R. W. W., et al. (2014). Cohesin and CTCF Differentially Affect Chromatin Architecture and Gene Expression in Human Cells. *Proc. Natl. Acad. Sci.* 111, 996–1001. doi:10.1073/pnas.1317788111

Prioritisation of Candidate Genes Underpinning COVID-19 Host Genetic Traits Based on High-Resolution 3D Chromosomal Topology

Michiel J. Thiecke[1†], Emma J. Yang[2,3†], Oliver S. Burren[4], Helen Ray-Jones[2,3*‡] and Mikhail Spivakov[2,3*‡]

[1]Enhanc3D Genomics Ltd, Cambridge, United Kingdom, [2]Functional Gene Control Group, MRC London Institute of Medical Sciences, London, United Kingdom, [3]Institute of Clinical Sciences, Faculty of Medicine, Imperial College, London, United Kingdom, [4]Cambridge Institute of Therapeutic Immunology and Infectious Disease, Department of Medicine, University of Cambridge, Cambridge, United Kingdom

Genetic variants showing associations with specific biological traits and diseases detected by genome-wide association studies (GWAS) commonly map to non-coding DNA regulatory regions. Many of these regions are located considerable distances away from the genes they regulate and come into their proximity through 3D chromosomal interactions. We previously developed COGS, a statistical pipeline for linking GWAS variants with their putative target genes based on 3D chromosomal interaction data arising from high-resolution assays such as Promoter Capture Hi-C (PCHi-C). Here, we applied COGS to COVID-19 Host Genetic Consortium (HGI) GWAS meta-analysis data on COVID-19 susceptibility and severity using our previously generated PCHi-C results in 17 human primary cell types and SARS-CoV-2-infected lung carcinoma cells. We prioritise 251 genes putatively associated with these traits, including 16 out of 47 genes highlighted by the GWAS meta-analysis authors. The prioritised genes are expressed in a broad array of tissues, including, but not limited to, blood and brain cells, and are enriched for genes involved in the inflammatory response to viral infection. Our prioritised genes and pathways, in conjunction with results from other prioritisation approaches and targeted validation experiments, will aid in the understanding of COVID-19 pathology, paving the way for novel treatments.

Keywords: COVID-19, GWAS (genome-wide association studies), enhancers and promoters, regulatory genome, 3D chromosomal architecture

INTRODUCTION

Patients with COVID-19 disease, caused by SARS-CoV-2 infection, show a broad range of symptoms and severity, from asymptomatic disease to fatal progressive respiratory failure (Hu et al., 2021). Several known epidemiological factors increase the risk of COVID-19 severity and mortality: old age, male gender and pre-existing medical conditions such as diabetes (Docherty et al., 2020; Huang et al., 2020). These factors, however, do not fully explain the variability and clinical outcome of COVID-19. Following the outbreak of the disease caused by a related virus, SARS, in 2002–2004, it was suggested that host genetic factors influence the clinical course and outcome of coronavirus infections (de Wilde et al., 2018). These findings have provided a motivation for a systematic identification of host

genetic factors linked with COVID-19 susceptibility and severity using genome-wide association studies (GWAS). Most recently, the COVID-19 host genetic initiative (COVID-19 HGI) has joined up these efforts to produce GWAS meta-analyses in four case-control settings in ~50 K patients and ~2 M controls from 47 studies in total (at Release 5), thereby increasing the power and robustness of individual GWAS (COVID-19 Host Genetics Initiative, 2021).

Whilst GWAS have revealed the underpinning genetic components of many phenotypes (Buniello et al., 2019), translating the identified genotype-disease associations into actionable therapeutic targets has presented a major challenge. To a large extent, this is due to the fact that the absolute majority of GWAS variants map outside of the protein-coding and promoter regions of the genome and are instead enriched at distal DNA regulatory elements such as gene enhancers (Cano-Gamez and Trynka, 2020). Enhancers may localise long distances (hundreds of kilobasepairs) away from their target gene promoters and come into their physical proximity via 3D chromosomal contacts (Schoenfelder and Fraser, 2019; Ray-Jones and Spivakov, 2021).

Currently, 3D chromosomal contacts are typically measured by Hi-C, a chromatin proximity ligation technique using next-generation sequencing of the ligation junctions for detection (van Berkum et al., 2010). Theoretically, Hi-C makes it possible to map all pairwise genomic contacts in the genome at a restriction-fragment resolution. However, the high complexity of Hi-C sequencing libraries limits the practically achievable genomic coverage, leading to a reduced sensitivity and resolution of this method. This limitation can be effectively mitigated using Capture Hi-C (PCHi-C), which enriches Hi-C libraries prior to sequencing for fragment pairs that include, at least on one end, regions of interest, such as annotated gene promoters (Schoenfelder et al., 2018).

We previously developed COGS (Capture Hi-C Omnibus Gene Score), a formal statistical framework to capitalise on high-resolution chromosome conformation data such as PCHi-C to link GWAS variants with their putative target genes (Javierre et al., 2016; Burren et al., 2017). The COGS pipeline generates a Bayesian prioritisation score for each gene being causal for a given GWAS trait, with causal genes defined as those containing at least one causal variant in the coding region, promoter and/or promoter-interacting regions detected by PCHi-C.

Here we used COGS with our previously generated PCHi-C data in 17 primary blood cell types (Javierre et al., 2016) and in a SARS-CoV-2-infected lung carcinoma cell line (Ho et al., 2021) to prioritise candidate genes underpinning COVID-19 host genetic associations from COVID-19 HGI Host GWAS meta-analysis (COVID-19 Host Genetics Initiative, 2021). We prioritise 251 putative genes associated with SARS-CoV-2 infection and COVID-19 susceptibility and severity, the majority of which were not previously implicated in these traits, and characterise their expression patterns and functional annotations.

METHODS

The COGS Prioritisation Pipeline

The COGS pipeline (Javierre et al., 2016; Burren et al., 2017) takes GWAS summary data as input, fine-maps it using Wakefield synthesis (Wakefield, 2009) and aggregates the resulting posterior probabilities of a variant being casual across all promoter-interacting regions detected using PCHi-C data. It then uses LD block data to compute the probability that there is at least one causal variant in at least one gene-associated region, including promoter-connected fragments, promoter-proximal regions (the baited restriction fragment and its immediate flanking fragments) and/or the gene's coding regions, under the assumption that there is at most one causal variant per LD block. COGS scores correspond to the estimated Bayesian probabilities of having at least one causal GWAS variant associated with a gene. Since COGS is primarily a ranking algorithm, the choice of the score threshold for gene prioritisation remains subjective in the absence of a gold standard. We used a COGS score threshold of 0.3 in reporting the numbers of prioritised genes and, where required, for downstream analyses, with data presented in the last section of Results confirming that our choice of threshold was appropriate for these purposes.

We ran the COGS pipeline using each of the four COVID-19 HGI GWAS datasets (release 5 excluding 23andMe data) using *HindIII*-based PCHi-C data in 17 human primary blood cell types (Javierre et al., 2016; Burren et al., 2017) and *DpnII*-based PCHi-C data (in 5 kb bins, with the baited fragments left unbinned) in A549-ACE2 cells at 0, 8 and 24 h after SARS-CoV-2 infection (Ho et al., 2021). The cell-type specificity of COGS scores may not be consistent with the expression patterns of the prioritised genes, while using COGS in a pooled setting across multiple samples increases the sensitivity of the analysis (Javierre et al., 2016). At the same time, the coverage and design of different PCHi-C datasets may have systematic effects on detected interaction signals (Freire-Pritchett et al., 2021). Therefore, COGS was run separately for data from each GWAS meta-analysis using a pool of promoter interactions with CHiCAGO scores (Cairns et al., 2016) above 5 in at least one cell type in either dataset (Javierre: 707,583 interactions involving 21,102 baited promoter fragments; Ho: 43,265 interactions involving 9,955 baited promoter fragments). A minority of gene promoters were not baited in either PCHi-C capture system due to challenges in probe design and therefore not assayed in the respective systems. Therefore, their promoter-interacting regions could not be included in the analysis. To facilitate the analysis of their promoter-proximal variants, we generated "virtual baited fragments" for all annotated gene promoters. In addition, we included the coding variants of all annotated genes.

Data Sources

COVID-19 HGI GWAS meta-analysis release 5 data were downloaded from https://www.covid19hg.org/results/r5/. This release jointly analysed nearly 50,000 COVID-19 cases and

FIGURE 1 | The COGS prioritisation scores of genes associated with A2 COVID-19 host GWAS trait. Gene-level Manhattan plot showing COGS scores generated based on A2 COVID-19 host GWAS data and PCHi-C data from COVID Javierrere et al. (2016). The top scoring genes (COGS scores >0.3) are labelled in each locus. Multiple genes are labelled when there are several top-scoring genes with a very similar score, or lower-scoring genes with compelling biological functions. For simplicity, non-coding genes are not labelled, unless there are no prioritised protein-coding genes in the same locus. See **Supplementary Figures S1–S3** for the COGS gene-level Manhattan plots produced with the other three COVID-19 host GWAS traits based on PCHi-C data from Javierre et al. and **Supplementary Figures S4–S7** for the prioritisation results based on PCHi-C data from Ho et al. (2021).

over two million controls by combining data from 47 studies across 19 countries. Details for each study are provided on the HGI website and in the consortium paper (COVID-19 Host Genetics Initiative, 2021). The CHiCAGO-processed PCHi-C data from Javierre and Ho were downloaded from OSF (https://osf.io/u8tzp) and GEO (accession GSE164533), respectively. LD block data were generated with LD-detect (Berisa and Pickrell, 2016) and downloaded from the software author's website (http://bitbucket.org/nygcresearch/ldetect-data). Note that the LD block dataset did not include sex chromosomes, which were therefore excluded from COGS analysis. However, no strong association signals were detected on sex chromosomes in COVID-19 HGI GWAS, and therefore this limitation is unlikely to have missed strongly implicated genes.

The Javierre PCHi-C data are on GRCh37 assembly, and we used the GRCh37 versions of the COVID-19 HGI GWAS datasets, the original LD block data from Berisa and Pickrell and gene models from Ensembl GRCh37 Release 103 (https://grch37.ensembl.org) in the analysis. The Ho PCHi-C data are on GRCh38 assembly, and we used the GRCh38 versions of the COVID-19 HGI GWAS datasets, the lifted-over (GRCh37-to-38) LD block data and gene models from Ensembl GRCh38 Release 103. The results for each gene were linked between these analyses using Ensembl gene IDs as primary identifiers.

TPM-level gene expression data from GTEx and FPKM-level gene expression data from BLUEPRINT consortia were downloaded from GTEx portal (accession: phs000424. v8. p2) and EBI Gene expression atlas (accession: E-MTAB-3827), respectively. Gene sets of COVID-19 differentially expressed genes in multiple human cell types and tissues (106 conditions) were obtained from The COVID-19 Drug and Gene Set Library (https://maayanlab.cloud/covid19/) (Kuleshov et al., 2020). Differentially expressed genes in COVID-19 were obtained from Supplementary Table S2 in (Daamen et al., 2021) and the union of genes reported for peripheral blood mononuclear cells (PBMCs), lung tissue and bronchoalveolar lavage was taken. Hallmark gene sets were obtained from the Molecular Signature Database (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp) (Liberzon et al., 2015).

Gene-Level Manhattan Plots

Gene-level Manhattan plots were generated separately for each COGS run on a given GWAS and PCHi-C dataset using the R package ggplot2 (Wickham, 2016). Genes with COGS scores >0.3 were labelled in each locus. Multiple genes were labelled when there were several top-scoring genes with a very similar score, or lower-scoring genes with compelling biological functions. For simplicity, we did not label non-

coding genes unless there were no prioritised protein-coding genes in the same locus.

Comparison of COGS With Other Gene-Prioritisation Approaches

In the naive GWAS prioritisation approach, variants with nominal p-values below 10^{-8} were assigned to the nearest exon. The list of HGI-prioritised genes was taken from Figure 1 in the HGI consortium paper (COVID-19 Host Genetics Initiative, 2021). Genes outside of the regions highlighted in the Figure were defined as those whose TSSs mapped more than 1 Mb away from the lead variant. The list of TWAS- and SMR-prioritised genes was taken from Tables 2 and 3 in Baranova et al., 2021.

Gene Expression Analysis

K-means clustering was performed on the scaled expression values of COGS-prioritised genes (score >0.3) in GTEx (TPM) and BLUEPRINT (RPKM) datasets using R package pheatmap with the number of clusters determined using the Silhouette and Elbow methods. GTEx analysis included 218 genes with detectable expression, and BLUEPRINT analysis focused on the 55 genes in the top 25% of expression in blood cells.

The GSEAPreranked analysis (Mootha et al., 2003; Subramanian et al., 2005) against COVID-19 differential expression signature gene sets used all genes returned by COGS, ranked by COGS score. Analysis was performed using the GSEA software (downloaded from www.gsea-msigdb.org) with default parameters and 1,000 permutations. Results were collated into a bubble plot using the R package ggplot2 (Wickham, 2016). Precision and recall analysis of COGS-prioritised genes versus COVID-19 differentially expressed genes (Daamen et al., 2021) was performed in R.

Annotation of the Prioritised Genes

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis (Kanehisa and Goto, 2000) used COGS-prioritised genes (score >0.3). The analysis was performed using the enrichKEGG function in the ClusterProfiler package (Yu et al., 2012) with an adjusted p value of 0.05. Significantly enriched pathways were visualised with KEGG mapper (Kanehisa and Sato, 2020). The GSEAPreranked analysis on Hallmark gene sets was run as for COVID-19-associated gene sets above.

RESULTS AND DISCUSSION

Prioritisation of COVID-19 Host GWAS Genes Using PCHi-C Data

To prioritise candidate genes associated with COVID-19 susceptibility and severity, we integrated the worldwide meta-analysis data from the COVID-19 Host Genetics Initiative (COVID-19 HGI Release 5) (COVID-19 Host Genetics Initiative, 2021) with PCHi-C data using the COGS pipeline (Javierre et al., 2016; Burren et al., 2017). COVID-19 HGI divided the patients into three categories: A- very

severe cases characterised by respiratory failure, B- all hospitalised cases, and C - all cases that tested positive for SARS-CoV-2 infection, generating a GWAS meta-analysis for the following four traits: A2 (very severe cases vs population), B1 (hospitalised vs non-hospitalised Covid-19 patients), B2 (hospitalised patients vs population) and C2 (confirmed Covid-19 vs population) (COVID-19 Host Genetics Initiative, 2021).

We first used high-coverage PCHi-C data in 17 human primary blood cell types (Javierre et al., 2016), including endothelial progenitors, as the source of 3D chromosomal contacts for COGS. We prioritised 234 genes with COGS scores above 0.3 across the four GWAS, of which 37 had scores above 0.75. More than half of the prioritised genes (122/234) were detected from A2 GWAS, consistent with the number of significant variant-trait associations in this study. A total of 78 genes were uniquely prioritised from A2 and not the other three GWAS. Including B2 in the analysis contributed an additional 71 genes, followed by B1 and C2 (26 and 15 additional genes, respectively).

We expressed the prioritisation analysis results in the form of gene-level Manhattan plots (**Figure 1** and **Supplementary Figure S1–S3**), which showed that clusters of adjacent genes were often prioritised jointly. In some cases, this was due to two promoters sharing the same PCHi-C baited fragment (e.g., *VRK3* and *ZNF473*). However, multiple genes may genuinely share GWAS variant-containing enhancers (Ray-Jones and Spivakov, 2021). Therefore, we have avoided further "fine-mapping" of COGS associations to the top-scoring gene in each peak.

We next used PCHi-C data from our recent analysis of a SARS-CoV-2 infected lung cell line (ACE2-expressing A549 cells) and uninfected controls (Ho et al., 2021). This experiment used a different PCHi-C design, based on *DpnII* and analysed in 5 kb bins (outside of the baited promoter regions that were left unbinned), as opposed to *HindIII* in the Javierre et al. blood cell analysis. This analysis returned 60 prioritised genes with COGS scores above 0.3, of which 13 had scores above 0.75. The gene-level Manhattan plots for this analysis are shown in **Supplementary Figure S4–S7**. The lower number of genes compared with the blood cell data is expected given the lower sequencing coverage and the smaller number of cell types profiled in this experiment. Over 70% (43/60) of the genes prioritised using this dataset (COGS score >0.3) also had scores above 0.3 in the blood cell-based analysis, indicating that the results of COGS prioritisation show a significant degree of consistency across different cell types and PCHi-C array designs.

Overall, 251 unique genes were prioritised based on four GWAS and two PCHi-C datasets at COGS score >0.3. The full results for all genes with their associated COGS scores are presented in **Supplementary Table S1**.

Comparison of COGS With Other Gene-Prioritisation Approaches
Comparison With Nearest-Exon Variant-To-Gene Assignment

To compare the results of COGS prioritisation with a naive approach, we selected GWAS variants with nominal p-values

FIGURE 2 | Expression patterns of the prioritised genes. Heatmaps showing the results of k-means clustering of COGS-prioritised genes (scores >0.3) based on their relative expression levels across the tissues profiled by the GTEx consortium **(A)** and across primary blood cell types profiled by the BLUEPRINT consortium **(B)**. Relative gene expression in **(A)** represents gene-level TPM values scaled across all GTEx genes, and in **(B)** gene-level RPKM values scaled across genes with the top 25% of expression in the BLUEPRINT dataset. Each cell in the heatmap represents a cluster, with the gene-to-cluster assignments listed in **Supplementary Table S5A, B**, respectively. Abbreviations: FPKM, fragments per kilobase of transcript per million mapped reads, TPM, transcripts per million.

below 10^{-8} (traditionally taken as a "genome-wide significance level" through the Bonferroni correction) and assigned them to the nearest exon. Across the four GWAS, this approach prioritised 45 genes, of which 11 were also prioritised by COGS at a score threshold of 0.3 (**Supplementary Table S2**). The genes prioritised by both approaches included 8/23 loci with coding variants (*ABO, DPP9, IFNAR2, KANSL1, LZTFL1, OAS1, OAS3, SLC6A20*), and 3/22 with non-coding variants mapped to their nearest exons (*AP000295.9, PDCL3P4, RP11-304F15.3*). Genes identified by the nearest-exon approach exhibited a wide range of COGS scores (**Supplementary Figure S8A**). Unlike in the naive approach, COGS additionally incorporates data from promoter-interacting regions and has improved precision due to the use of statistical fine-mapping. Therefore, a limited overlap between these two approaches is expected.

Comparison With the COVID-19 HGI Gene Prioritisation Approach

The COVID-19 HGI consortium paper defined 13 genomic loci associated with infection or severe disease, and highlighted

47 putative gene targets across these loci. The genes highlighted in the consortium paper satisfied one or more of the following criteria: 1) being in close proximity to the lead variant, 2) overlapping disease-associated variants, 3) containing disease-associated coding variants (loss-of-function, missense), 4) being associated with an eQTL in LD with the lead variant, or 5) being prioritised by the OpenTargets V2G (Variant-to-Gene) algorithm (COVID-19 Host Genetics Initiative, 2021). HGI-prioritised genes showed a broad range of COGS scores (**Supplementary Figure S8B**), with 16 out of 47 HGI-prioritised genes showing scores above 0.3 (**Supplementary Table S3A**). For example, while HGI prioritised all three genes in the 2′-5′-Oligoadenylate synthetase (*OAS*) cluster, COGS prioritised *OAS3* (max COGS = 0.81) and *OAS1* (max COGS = 0.54), while *OAS2* had a subthreshold score (max COGS = 0.15).

At a COGS threshold of 0.3, a further 38/251 genes were prioritised within the 13 loci of genome-wide significance highlighted in the paper (**Supplementary Table S3A**). Notably, in the 21q22.11 locus we prioritised interferon A and B receptor subunit 1 (*IFNAR1*; max COGS = 0.91) in addition to the HGI-prioritised subunit 2 (*IFNAR2*; max COGS ~1); the products of these

two genes combine to form the type I interferon receptor (Piehler et al., 2012). In the 19q13.33 locus, the five HGI-prioritised genes had low COGS scores, whereas *BCAT2* (max COGS = 0.63) and *FTL* (max COGS = 0.41) were instead prioritised; of these, *FTL* (ferritin light chain) is reported to be anti-inflammatory (Zarjou et al., 2019). In the 19p13.3 locus, the Dipeptidyl peptidase 9 (*DPP9*) gene, which plays a key role in inflammasome regulation (Zhong et al., 2018), was confirmed with a COGS score of 1 (as well as two nearby non-coding genes: *DPP9-AS1* and *AC005783.1*). However, COGS also identified further seven distal gene targets including UBX domain protein 6 (*UBXN6*), which reportedly inhibits the degradation of COVID-19-implicated proteins IFNAR1 and TYK2 (Ketkar et al., 2021).

The remaining 197 out of 251 genes prioritised by COGS mapped outside of the 13 genome-wide significance loci (**Supplementary Table S3B**). These included such plausible candidates as LIF receptor (*LIFR*) and TNF receptor superfamily (*TNFSF*) members *10A/B*, *TNFSF8* and *15*, which have roles in cytokine signalling, as well as components of the PI3K/AKT signalling pathway (*LAMB4*, *THBS3*, *TLR4* and *YWHAE*), which was recently proposed as a therapeutic target in COVID-19 (Khezri, 2021).

Comparison With a Multiomics-Based Prioritisation

A recent study (Baranova et al., 2021) tested the colocalisation of COVID-19 HGI GWAS signals with expression and methylation quantitative trait loci using a combination of transcriptome-wide association study (TWAS) and Summary-based Mendelian randomisation (SMR). This approach prioritised 14 genes, five of which (*IFNAR2*, *MGC57346/LINC02210*, *OAS1*, *OAS3*, and *TYK2*) were also prioritised by COGS (score > 0.3). The remaining 9/14 genes had COGS scores ranging from zero to 0.249 (**Supplementary Table S4**).

Overall, while COGS analysis has confirmed the prioritisation of several genes found by the nearest-exon and alternative priorisation approaches, it also revealed large numbers of further candidates. The summary of all four prioritisation approaches is presented in **Supplementary Figure S8C** and in **Supplementary Table S4**.

Expression Patterns of COGS-Prioritised Genes

Tissue-specificity of the Prioritised Genes

To assess the gene expression patterns of the COGS-prioritised genes, we first took advantage of GTEx data across 54 non-diseased tissues (GTEx Consortium, 2020). In total, 218 genes were represented in this dataset. K-means clustering of scaled expression values segregated these genes into eight coherent clusters (**Figure 2A**; **Supplementary Table S5A**). Two large clusters (A6 and A8) containing 51 genes in total were characterised by their predominant expression in whole blood or EBV-transformed lymphocytes, respectively. This was expected from the involvement of well-characterised candidates such as *IFNAR1/2*, *OAS1/3* and *TYK2* in the immune function, as well as from the fact that the Javierre PCHi-C dataset was generated in blood cells. However, COGS

also prioritised multiple genes active in other tissues, likely driven by promoter-proximal and coding variants, as well as promoter contacts shared across tissues. Genes in two other clusters (A3 and A5; 45 genes in total, including synapse-associated genes *SYN2*, *SYT3* and *SHANK1*) were predominantly expressed in different parts of the brain, consistent with the common neurological symptoms and evidence of brain damage following SARS-CoV-2 infection (Marshall, 2021). Somewhat surprisingly, cluster A1 (31 genes in total) contained genes showing high expression in testis, including sperm-associated calcium channel subunit *CATSPERG* and signal peptide peptidase *SPPL2C* active in spermatids. While SARS-CoV-2 is known to infect testis (Ma et al., 2021) and the male sex is a known risk factor for COVID-19 severity (Docherty et al., 2020; Huang et al., 2020), the exact role of these genes in COVID-19 pathology remains to be elucidated. The remaining three clusters (clusters A2, A4, A7; 100 genes in total) were characterised by broader expression patterns across multiple tissues, including the lung, gut, skin and vasculature.

To obtain a finer-grained view of the prioritised genes expression patterns in the blood, we studied their expression in 27 primary blood cell types using data from the BLUEPRINT consortium (**Figure 2B**; **Supplementary Table S5B**) (Chen et al., 2014). We restricted this analysis to 55 genes showing top 25% expression levels in the BLUEPRINT dataset. K-means clustering of their scaled expression values yielded four distinct clusters containing between 10 and 17 genes each, characterised by predominant expression in T lymphocytes (cluster B1), erythroblasts (B2), macrophages (B3) and mature eosinophils (B4), respectively. Examples of genes in these clusters include effectors of TNF (*TNFSF8*, B1; *TNFRSF10B*, B2), toll-like receptor (*TLR4*, B3) and interferon signalling (*IFNAR1* and *IFNAR2*, clusters B3 and B4, respectively).

Jointly, these results suggest the involvement of a broad range of blood cells and solid tissues in COVID-19 pathology.

Comparison With Reported COVID-19-Regulated Genes

We asked if COGS preferentially prioritised genes that are known to change expression in response to COVID-19 infection. To address this question, we used the COGS scores of all annotated genes in a quantitative gene set enrichment analysis (GSEAPreranked) against 106 differential expression signature gene sets from The COVID-19 Drug and Gene Set Library (see Methods). All 106 COVID-19 gene sets had a positive Normalised Enrichment Score (NES), meaning that they were enriched at the top of the COGS-ranked gene list, with a mean FDR of 0.080 ± 0.125. This enrichment was significant at an FDR of 0.25 for 97 of the gene sets (**Figure 3A**; top plot and **Supplementary Table S5C**), indicating that the genes' COGS scores positively associate with their differential expression in COVID-19. The top two sets, as ranked by the Normalised Enrichment Score (NES), were lung organoids infected with COVID-19 *in vitro* (top-ranking

FIGURE 3 | Enrichment of COGS-prioritised genes in COVID-19-response gene sets. Quantitative GSEA analysis using the COGS score for each gene against gene sets from the COVID-19 Drug and Gene Set Library. Diagnostic plots produced by the GSEA software demonstrate the relationship between the normalised enrichment score (NES) and measures of significance [**(A)**, top plot] and the enrichment across COGS scores for the gene sets with the top two NES scores (A, middle and bottom plots) **(B)** Bubble-plot showing results for all gene sets. The "up" and "down" suffixes indicated the direction of differential expression in COVID-19 for the gene set in question.

genes *IFNAR2, CHI3L2, OAS3*) and natural killer (NK) cells from individuals with severe disease versus healthy (top-ranking genes *DPP9, SAFB2, SAFB*) (**Figure 3A** middle and bottom plots, respectively). We noted that the sets achieving the highest NES tended to contain upregulated, rather than downregulated, genes (**Figure 3B**), suggesting a role of many underpinning variants in controlling gene induction in response to infection. Overall, these results provide additional validation that the COGS approach prioritises genes with relevance to COVID-19.

The Biological Function of COGS-Prioritised Genes

To gain insight into the shared biological functions of the prioritised genes, we first performed KEGG pathway over-representation analysis (**Figure 4A**). We found that COGS-prioritised genes (max COGS score >0.3) were significantly enriched in pathways associated with response to influenza A and measles infection, as well as with inflammatory processes, including NOD-like receptor signaling (**Figure 4B**), necroptosis

and natural killer cell-mediated cytotoxicity. These enriched annotations were driven by a total of 11 COGS-prioritised genes with a high overlap between individual pathways (*FTL, IFNAR1/2, OAS1/3, PPP3C, TLR4, TNFRSF10 A/B, TYK2*, and *VAV3*; **Supplementary Table S6A**). We note that all five enriched pathways are druggable according to the KEGG database (Kanehisa and Goto, 2000), creating potential opportunities for drug repurposing for COVID-19 treatment. For example, the NOD-like receptor signaling pathway alone is currently targeted by 14 drugs indicated for various inflammatory diseases, with one of these drugs, a selective IRAK4 inhibitor Zimlovisertib, undergoing a clinical trial for COVID-19-induced pneumonia (https://clinicaltrials.gov/ct2/show/NCT04575610).

To further increase the sensitivity of pathway enrichment analysis, we again performed quantitative GSEA based on the COGS scores, this time against 50 Hallmark gene sets from the Molecular Signatures Database. Although none of the Hallmark sets were significantly enriched at an FDR of 0.25 (**Supplementary Table S6B**), the top sets ranked by NES included relevant inflammatory processes such as IL-6

FIGURE 4 | The biological functions of the prioritised genes. **(A)** Bubble plot showing the KEGG pathways enriched among COGS-prioritised genes (score >0.3). **(B)** Diagram of the NOD-like receptor signalling pathway, with the COGS-prioritised genes highlighted in red. **(C)** Bubble plot showing the results of a quantitative GSEA analysis using the COGS score for each gene against Hallmark gene sets.

signalling by STAT3, IL-2 signalling by STAT5, TNF-α signaling via NFκB, IFN-γ response and TGF-β signalling (**Figure 4C**), highlighting the roles of individual COGS-prioritised genes in these processes.

Jointly, these results support the notion that genetically-determined variation in the inflammatory response to viral infection plays a key role in COVID-19 susceptibility and severity.

Assessment of COGS Score Thresholds Based on Prioritisation of COVID-19-Differentially Expressed Genes

Since COGS-prioritised genes were enriched in gene sets associated with COVID-19 transcriptional response, we used this property to estimate the sensitivity and specificity of COGS analysis at a range of score thresholds. We focused on data from a recent COVID-19 host transcriptomics study reporting 11,170 differentially-expressed (DE) genes across PBMCs, lung and bronchoalveolar lavage samples (Daamen et al., 2021), of which 10,463 had a non-zero COGS score in our dataset. Assuming that this set of DE genes is enriched for true causal loci, we performed a precision-recall analysis of these genes at a range of COGS thresholds between 0 and 1 (**Supplementary Figure S8D**). As expected, increasing the COGS threshold increased the enrichment for DE genes (a proxy for specificity or "precision") among the prioritised candidates, but decreased their recall, as more DE genes ended up with subthreshold scores. Our predefined threshold of COGS score > 0.3 corresponded to a point at which the enrichment started to rise sharply (**Supplementary Figure S8D**), confirming that our choice of this threshold was reasonable for global downstream analyses. However, for more targeted selection of candidates (e.g., for small-scale perturbation experiments), using a higher COGS score threshold, which likely confers a higher specificity of the analysis at the expense of a lower sensitivity, may be warranted.

CONCLUSION

The COGS pipeline combining Bayesian fine-mapping of GWAS signals with PCHi-C-based prioritisation has provided 251 putative genes associated with COVID-19 severity, most of which were not prioritised using the naive nearest-exon approach and the strategies used in the original COVID-19 HGI GWAS publication. Most of these genes have no known biological function in COVID-19 to date, but are enriched in pathways associated with inflammatory response to viral infection. In conjunction with complementary prioritisation approaches and targeted validation experiments (Cano-Gamez and Trynka, 2020), these data will help to understand and tackle COVID-19 pathology.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article and **Supplementary Material**. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

This study did not generate primary data, and used only fully anonymised, publicly available data from human subjects. Therefore, ethical review and written informed consent were not required for this study under the local legislation and institutional requirements.

AUTHOR CONTRIBUTIONS

MT, EY, HR-J and MS designed the study and performed the analysis; OSB contributed computational tools and expert advice; EY, HR-J and MS wrote the manuscript; MS and HR-J supervised the work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene. 2021.745672/full#supplementary-material

REFERENCES

Baranova, A., Cao, H., and Zhang, F. (2021). Unraveling Risk Genes of COVID-19 by Multi-Omics Integrative Analyses. *Front. Med.* 8, 738687. doi:10.3389/fmed.2021.738687

Berisa, T., and Pickrell, J. K. (2016). Approximately Independent Linkage Disequilibrium Blocks in Human Populations. *Bioinformatics* 32, 283–285. doi:10.1093/bioinformatics/btv546

Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., et al. (2019). The NHGRI-EBI GWAS Catalog of Published Genome-wide Association Studies, Targeted Arrays and Summary Statistics 2019. *Nucleic Acids Res.* 47, D1005–D1012. doi:10.1093/nar/gky1120

Burren, O. S., Rubio García, A., Javierre, B.-M., Rainbow, D. B., Cairns, J., Cooper, N. J., et al. (2017). Chromosome Contacts in Activated T Cells Identify Autoimmune Disease Candidate Genes. *Genome Biol.* 18, 165. doi:10.1186/s13059-017-1285-0

Cairns, J., Freire-Pritchett, P., Wingett, S. W., Várnai, C., Dimond, A., Plagnol, V., et al. (2016). CHiCAGO: Robust Detection of DNA Looping Interactions in Capture Hi-C Data. *Genome Biol.* 17, 127. doi:10.1186/s13059-016-0992-2

Cano-Gamez, E., and Trynka, G. (2020). From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Front. Genet.* 11, 424. doi:10.3389/fgene.2020.00424

Chen, L., Kostadima, M., Martens, J. H. A., Canu, G., Garcia, S. P., Turro, E., et al. (2014). Transcriptional Diversity during Lineage Commitment of Human Blood Progenitors. *Science* 345, 1251033. doi:10.1126/science.1251033

COVID-19 Host Genetics Initiative (2021). Mapping the Human Genetic Architecture of COVID-19. *Nature*, 1–8. doi:10.1038/s41586-021-03767-x

Daamen, A. R., Bachali, P., Owen, K. A., Kingsmore, K. M., Hubbard, E. L., Labonte, A. C., et al. (2021). Comprehensive Transcriptomic Analysis of COVID-19 Blood, Lung, and Airway. *Sci. Rep.* 11, 7052. doi:10.1038/s41598-021-86002-x

de Wilde, A. H., Snijder, E. J., Kikkert, M., and van Hemert, M. J. (2018). Host Factors in Coronavirus Replication. *Curr. Top. Microbiol. Immunol.* 419, 1–42. doi:10.1007/82_2017_25

Docherty, A. B., Harrison, E. M., Green, C. A., Hardwick, H. E., Pius, R., Norman, L., et al. (2020). Features of 20 133 UK Patients in Hospital with Covid-19 Using the ISARIC WHO Clinical Characterisation Protocol: Prospective Observational Cohort Study. *BMJ* 369, m1985. doi:10.1136/bmj.m1985

Freire-Pritchett, P., Ray-Jones, H., Della Rosa, M., Eijsbouts, C. Q., Orchard, W. R., Wingett, S. W., et al. (2021). Detecting Chromosomal Interactions in Capture Hi-C Data with CHiCAGO and Companion Tools. *Nat. Protoc.* 16, 4144–4176. doi:10.1038/s41596-021-00567-5

GTEx Consortium (2020). The GTEx Consortium Atlas of Genetic Regulatory Effects across Human Tissues. *Science* 369, 1318–1330. doi:10.1126/science.aaz1776

Ho, J. S. Y., Mok, B. W.-Y., Campisi, L., Jordan, T., Yildiz, S., Parameswaran, S., et al. (2021). TOP1 Inhibition Therapy Protects against SARS-CoV-2-Induced Lethal Inflammation. *Cell* 184, 2618–2632. doi:10.1016/j.cell.2021.03.051

Hu, B., Guo, H., Zhou, P., and Shi, Z.-L. (2021). Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* 19, 141–154. doi:10.1038/s41579-020-00459-7

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020). Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan, China. *The Lancet* 395, 497–506. doi:10.1016/s0140-6736(20)30183-5

Javierre, B. M., Burren, O. S., Wilder, S. P., Kreuzhuber, R., Hill, S. M., Sewitz, S., et al. (2016). Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* 167, 1369–1384. doi:10.1016/j.cell.2016.09.037

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30. doi:10.1093/nar/28.1.27

Kanehisa, M., and Sato, Y. (2020). KEGG Mapper for Inferring Cellular Functions from Protein Sequences. *Protein Sci.* 29, 28–35. doi:10.1002/pro.3711

Ketkar, H., Harrison, A. G., Graziano, V. R., Geng, T., Yang, L., Vella, A. T., et al. (2021). UBX Domain Protein 6 Positively Regulates JAK-Stat1/2 Signaling. *J.I.* 206, 2682–2691. doi:10.4049/jimmunol.1901337

Khezri, M. R. (2021). PI3K/AKT Signaling Pathway: a Possible Target for Adjuvant Therapy in COVID-19. *Hum. Cel.* 34, 700–701. doi:10.1007/s13577-021-00484-5

Kuleshov, M. V., Stein, D. J., Clarke, D. J. B., Kropiwnicki, E., Jagodnik, K. M., Bartal, A., et al. (2020). The COVID-19 Drug and Gene Set Library. *Patterns* 1, 100090. doi:10.1016/j.patter.2020.100090

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cel Syst.* 1, 417–425. doi:10.1016/j.cels.2015.12.004

Ma, X., Guan, C., Chen, R., Wang, Y., Feng, S., Wang, R., et al. (2021). Pathological and Molecular Examinations of Postmortem Testis Biopsies Reveal SARS-CoV-2 Infection in the Testis and Spermatogenesis Damage in COVID-19 Patients. *Cell. Mol. Immunol.* 18, 487–489. doi:10.1038/s41423-020-00604-5

Marshall, M. (2021). COVID and the Brain: Researchers Zero in on How Damage Occurs. *Nature* 595, 484–485. doi:10.1038/d41586-021-01693-6

Mootha, V. K., Lindgren, C. M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., et al. (2003). PGC-1α-responsive Genes Involved in Oxidative Phosphorylation Are Coordinately Downregulated in Human Diabetes. *Nat. Genet.* 34, 267–273. doi:10.1038/ng1180

Piehler, J., Thomas, C., Garcia, K. C., and Schreiber, G. (2012). Structural and Dynamic Determinants of Type I Interferon Receptor Assembly and Their Functional Interpretation. *Immunol. Rev.* 250, 317–334. doi:10.1111/imr.12001

Ray-Jones, H., and Spivakov, M. (2021). Transcriptional Enhancers and Their Communication with Gene Promoters. *Cell. Mol. Life Sci.* doi:10.1007/s00018-021-03903-w

Schoenfelder, S., and Fraser, P. (2019). Long-range Enhancer-Promoter Contacts in Gene Expression Control. *Nat. Rev. Genet.* 20, 437–455. doi:10.1038/s41576-019-0128-0

Schoenfelder, S., Javierre, B.-M., Furlan-Magaril, M., Wingett, S. W., and Fraser, P. (2018). Promoter Capture Hi-C: High-Resolution, Genome-wide Profiling of Promoter Interactions. *JoVE* 136, 57320. doi:10.3791/57320

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: a Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. *Proc. Natl. Acad. Sci.* 102, 15545–15550. doi:10.1073/pnas.0506580102

van Berkum, N. L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L. A., et al. (2010). Hi-C: a Method to Study the Three-Dimensional Architecture of Genomes. *JoVE* 39, 1869. doi:10.3791/1869

Wakefield, J. (2009). Bayes Factors for Genome-wide Association Studies: Comparison withP-Values. *Genet. Epidemiol.* 33, 79–86. doi:10.1002/gepi.20359

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Heidelberg: Springer.

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16, 284–287. doi:10.1089/omi.2011.0118

Zarjou, A., Black, L. M., McCullough, K. R., Hull, T. D., Esman, S. K., Boddu, R., et al. (2019). Ferritin Light Chain Confers Protection against Sepsis-Induced Inflammation and Organ Injury. *Front. Immunol.* 10, 131. doi:10.3389/fimmu.2019.00131

FISH Going Meso-Scale: A Microscopic Search for Chromatin Domains

Antonina Maslova and Alla Krasikova*

Laboratory of Nuclear Structure and Dynamics, Cytology and Histology Department, Saint Petersburg State University, Saint Petersburg, Russia

The intimate relationships between genome structure and function direct efforts toward deciphering three-dimensional chromatin organization within the interphase nuclei at different genomic length scales. For decades, major insights into chromatin structure at the level of large-scale euchromatin and heterochromatin compartments, chromosome territories, and subchromosomal regions resulted from the evolution of light microscopy and fluorescence *in situ* hybridization. Studies of nanoscale nucleosomal chromatin organization benefited from a variety of electron microscopy techniques. Recent breakthroughs in the investigation of mesoscale chromatin structures have emerged from chromatin conformation capture methods (C-methods). Chromatin has been found to form hierarchical domains with high frequency of local interactions from loop domains to topologically associating domains and compartments. During the last decade, advances in super-resolution light microscopy made these levels of chromatin folding amenable for microscopic examination. Here we are reviewing recent developments in FISH-based approaches for detection, quantitative measurements, and validation of contact chromatin domains deduced from C-based data. We specifically focus on the design and application of Oligopaint probes, which marked the latest progress in the imaging of chromatin domains. Vivid examples of chromatin domain FISH-visualization by means of conventional, super-resolution light and electron microscopy in different model organisms are provided.

Keywords: chromatin domains, chromatin imaging, fluorescence *in situ* hybridization (FISH), FISH probes, fluorescent microscopy, topologically associating domains, genome compartments, Oligopaints

Abbreviations: (DOP)-PCR, degenerate oligonucleotide-primed polymerase chain reaction; (F)ISH, fluorescence *in situ* hybridization; (i)PALM, (interferometric) photoactivated localization microscopy; 3C, chromosome conformation capture; 3D-EMISH, three-dimensional serial block-face scanning electron microscopy combined with *in situ* hybridization; 5C, chromosome conformation capture carbon copy; BAC, bacterial artificial chromosome; ChIA-PET, chromatin interaction analysis by paired end tag sequencing; CRISPR/dCas9, clustered regularly interspaced short palindromic repeats/nuclease-deactivated CRISPR-associated protein 9; CTCF, CCCTC-binding factor; DNA-PAINT, DNA point accumulation for imaging of nanoscale topography; Hi-C, high-throughput chromosome conformation capture; Hi-M, high-throughput multiplexed sequential imaging; IHIs, inactive heterochromatic islands; IVT-RT, *in vitro* transcription and reverse transcription; MERFISH, multiplexed error-robust FISH; mESCs, mouse embryonic stem cells; Micro-C, micrococcal nuclease Hi-C; PAC, phage P1-derived artificial chromosomes; PcG, polycomb group; RASER-FISH, (resolution after single-strand exonuclease resection)-FISH; SIM, structured illumination microscopy; SRM, super-resolution microscopy; STORM, stochastic optical reconstruction microscopy; TAD, topologically associating domain.

INTRODUCTION: OVERVIEW OF THE TOOLKITS FOR CHROMATIN DOMAIN IMAGING

The term "chromatin" was coined by Walther Flemming at the end of the 19th century to designate structures stained by aniline dyes and confined within the cell nucleus (Paweletz, 2001). Together with significant improvements in sample preparation, detailed observations of chromatin behavior during cell division in different organisms and tissues made by Flemming, Walter Sutton, Karl Rabl, Theodor Bovery, and many other famous cytologists were possible through using light microscopes with state-of-the-art lenses, corrected for spherical and chromatic aberrations (Coleman, 1965). In the chromatin research timeline, the development of new microscopes and microscopy techniques together with chemical, biochemical, and later molecular biology methods for chromatin "contrasting" or "labeling" marked important milestones and defined the research trends for decades. The bulk of the methods, making chromatin details visible via a microscope, could be roughly classified into two categories: those that reveal protein components of the chromatin (mainly histones) and those that focus on DNA (reviewed in Lakadamyali and Cosma, 2015; Shao et al., 2017; Xu and Liu, 2019). Our current view on chromatin organization in the nucleus is drawn by implements from both protein and DNA detection. In this review, we will mainly concentrate on the evolution of DNA targeting techniques and tools, which fueled the recent success in imaging of subchromosomal chromatin domains at genomic length scales from several kilobases (kb) to several megabases (Mb). This most elusive "mesoscale" level of higher-order chromatin organization, mainly dissected by chromosome conformation capture technologies, is becoming open for microscopic examination.

At the beginning of the journey toward understanding the chromatin structure, conventional light microscopy was the only way to directly observe chromatin and chromosomes both in the nucleus and in spread (Cremer and Cremer, 2006). Interphase chromatin had been more readily detected when stained with common histological dyes and by Feulgen reaction, but finer details, other than the most intensely stained regions of chromosomes, remained indiscernible. Intensively stained heterochromatin was persistent and visible throughout the cell cycle in contrast to euchromatin, which decondensed in interphase (Heitz, 1929). Large masses of heterochromatin were generally observed at the nuclear periphery, near the nucleolus and in chromocenters (Barr and Bertram, 1949; Hsu et al., 1971). Despite limited instruments, early works provided the first evidence for large-scale chromatin structures and their non-random distribution within the nucleus (Comings, 1968). Later, fluorescent DNA dyes came into use as a straightforward and simple way to stain the chromatin and to unravel cell type-specific differences in its spatial arrangement (Latt, 1977; Ellison and Howard, 1981; Agard and Sedat, 1983; Solovei et al., 2009; Berchtold et al., 2011).

Chromatin ultrastructure at a nanometer scale has been intensively investigated by electron microscopy, which achieved a resolution three orders of magnitude higher than light microscopy (Woodcock and Horowitz, 1997; Daban, 2011). It was estimated that 10 nm chromatin fibers could account for only 6 fold linear DNA packaging, which forced the research toward deciphering other "levels" of higher-order chromatin compaction (Fussner et al., 2011). Transmission electron microscopy of thin sections of nuclei revealed chromatin filaments of larger size (from 30 to 130 nm) populating the nuclear volume and dense heterochromatic areas near the nuclear envelope and nucleolus (Belmont et al., 1989; Kuznetsova and Sheval, 2016). When interpreting chromatin structures, it should be borne in mind that chromatin compaction is highly sensitive to surrounding conditions (Albiez et al., 2006; Maeshima et al., 2019). For this reason, chromatin images taken by transmission electron microscopy are frequently criticized for possible artifacts caused by harsh sample preparation, including dehydration, contrasting with heavy metals, resin or plastic embedding and ultrathin sectioning (van Holde and Zlatanova, 1995; Mielańczyk et al., 2015). Efforts toward the preservation of chromatin ultrastructure and its nuclear environment stimulated the development of cryo-electron microscopy (Dubochet et al., 1988), serial microtome block-face scanning electron microscopy (Rouquette et al., 2009), focused ion beam milling combined with scanning electron microscopy (Hoang et al., 2017) and other techniques. Novel method of chromatin contrasting for electron microscopy, called ChromEM, can be effectively coupled with electron microscopy tomography (ChromEMT; Ou et al., 2017), transmission electron microscopy (ChromTEM) (Li et al., 2021) and scanning electron microscopy (ChromSTEM) (Huang et al., 2020; Li et al., 2021). ChromEMT demonstrated that in interphase nuclei and mitotic chromosomes chromatin is packed into disordered 5–24 nm granular chain with highly variable folding parameters and packing density.

Absence of underlying genetic sequence information remains a main obstacle for detailed investigation of the mesoscale chromatin domains, identified by electron microscopy. Indeed, when examining the ultrastructural image of chromatin, specific chromosomal regions are unidentifiable (Woodcock and Ghosh, 2010). Nowadays, to discern the ultrastructural organization of a certain genomic region, electron microscopy is combined with the identification of specific genomic sequences.

Fluorescent microscopes and the first prototypes of confocal laser scanning microscope came into emergence in the first half of the 20th century (Renz, 2013). However, their expansion and wide implication in chromatin studies started in 70th due to the appearance of a critical method, overcoming "sequence specificity" problem. Mary-Lou Pardue and Joseph Gall showed that DNA probes could effectively hybridize with complementary target DNA sequences in cytological preparations (Pardue and Gall, 1969). This technology, named nucleic acids *in situ* hybridization (ISH; John et al., 1969), transformed into fluorescence *in situ* hybridization (FISH), when hapten (or dye)-modified nucleotides and fluorescent streptavidin or antibody detection had been widely applied instead of radioisotope labeled probes (Langer-Safer et al., 1982; Manuelidis et al., 1982; Pinkel et al., 1986; Wiegant et al., 1991). Evolution in probe design, fluorochrome diversity, and versatility of labeling

protocols made ISH compatible with practically any microscopy technique including transmitted light microscopy, fluorescent, laser-scanning confocal and super-resolution light microscopy, as well as electron and correlation microscopy (Hutchison et al., 1982; Manuelidis, 1984; Rouquette et al., 2010; Weiland et al., 2011; Markaki et al., 2012; Jahn et al., 2016). Synergy of FISH and microscopy allowed to investigate individual gene loci and higher order genome organization relative to different nuclear compartments (Marshall, 2002), revived and expanded the theory of chromosome territories (Lichter et al., 1988; Cremer and Cremer, 2010), moved forward concepts of the dynamic nature of spatial genome organization and its close interdependence to genome functional state (Chubb and Bickmore, 2003; Kosak and Groudine, 2004; Misteli, 2005; Pueschel et al., 2016).

While FISH reveals DNA component of chromatin in fixed cells, immunofluorescent staining generally aims to detect chromatin proteins. Labeling of specific histone modifications, chromatin-associated proteins and components of chromatin-remodeling complexes allowed localizing structural and functional chromatin domains. Given the complexity of genome regulatory pathways, the development of FISH and immunofluorescent staining is directed to detect and visualize multiple targets in one experiment including the combination of FISH and immunofluorescent staining on the same preparation and automation of basic experimental procedures (Lin et al., 2015; Shachar et al., 2015; Huber et al., 2018). Before addressing FISH tools for mesoscale chromatin domain imaging, we will briefly focus on the delineation of these genome compartments by chromatin conformation capture.

CAPTURING AN IMAGE OF CHROMATIN DOMAINS

The emergence of the chromosome conformation capture (3C) technique in the first decade of the 21st century on the basis of nuclear ligation assay (Cullen et al., 1993) and the subsequent expansion of 3C technology had a great impact on the field of 3D genomics (Dekker et al., 2013; Denker and de Laat, 2016). 3C-derivate methods (Hi-C, 5C, ChIA-PET, Micro-C, etc.) allowed to dissect three-dimensional genome organization with resolution and throughput, unattainable by other approaches based on imaging (Fraser et al., 2015b; Goel and Hansen, 2020). Moreover, they facilitated the understanding of the functional significance of identified spatial genome folding due to the alignment of Hi-C data with other genome-wide landscapes (Sati and Cavalli, 2017). However, during the last decade, improving sequencing depth and "high throughput" power of C-based methods and the concurrent development of bioinformatics tools for the analysis of complex and big data have led to some ambiguities in the interpretation of results and terminological confusion (Marti-Renom et al., 2018; Ing-Simmons and Vaquerizas, 2019; Pal et al., 2019). The key initial steps of C-experiments – crosslinking and proximity-based ligation – were also pointed out as potential sources of biases and limitations (Gavrilov et al., 2015; Kempfer and Pombo, 2020). While recent updates in both data analysis and experimental procedures challenge some

of the biases (Brant et al., 2016; Belaghzal et al., 2017, 2021), the ability of C-methods to capture multimodal interactions of genomic loci, differentiate stable from short-lived interactions and examine cell-to-cell contact variability is still limiting (McCord et al., 2020).

Color-coded heatmaps of pairwise interactions of genomic loci, a common form of data presentation in high-throughput C-methods, provide information on intra- and interchromosomal interactions at different genomic length scales from dozens of megabases down to ∼200 bp (Ing-Simmons and Vaquerizas, 2019; Mota-Gómez and Lupiáñez, 2019; **Figure 1**). Resolution of these maps is important for the interpretation of Hi-C data. For example, early low-resolution Hi-C maps of the human genome demonstrated chromatin segregation into multi-megabase-sized compartments – active (A) and inactive (B) (Lieberman-Aiden et al., 2009; **Figure 1**). Higher resolution genomic heatmaps combined with epigenetic and transcriptomic data allowed to characterize finer subcompartments – two for A type (A1 and A2) and four for B type (B1, B2, B3 and B4) (Rao et al., 2014).

Another distinctive feature of Hi-C heatmaps in many species are local contact domains, known as topologically associating domains (TADs; Dekker and Heard, 2015). Since the first description of TADs in 2012, numerous facets of TAD structure, mechanisms of formation, dynamics during the cell cycle and ontogenesis, functional implications in gene regulation and genome folding have been deeply studied (Dixon et al., 2016; Szabo et al., 2019; Beagan and Phillips-Cremins, 2020). Paradoxically, the more data on TADs are accumulated, the more difficult it is to give a unified definition of TAD (de Wit, 2020). In the initial heatmaps of mammalian and *Drosophila* genomes, TADs were defined as diagonal domains of variable size (∼ 60 kb for *Drosophila* and ∼880 kb for mammals), where genomic loci belonging to the same TAD show higher interaction frequency than loci assigned to neighboring TADs (Dixon et al., 2012; Hou et al., 2012; Nora et al., 2012; **Figure 1**). Thus, the borders of TADs restrict genomic interactions between domains. However, even from these early maps it was clear that TADs comprise smaller substructures (subTADs), and the observed high-frequency contacts arise via chromatin looping (**Figure 1**). As with compartments, the development of Hi-C technology toward higher heatmap resolution enabled to detect finer details including discrete loops, insulated neighborhoods, enhancer-promoter contacts, etc. (Dowen et al., 2014; Rao et al., 2014; Krietenstein et al., 2020). TADs frequently appeared as assemblies of a number of nested domains and included into larger super-structures called TAD-cliques and meta-TADs (Fraser et al., 2015a; An et al., 2019; Collas et al., 2019). This hierarchy was also revealed by many computational domain-calling tools, but with variable correspondence in the identified domain borders (Weinreb and Raphael, 2016; Forcato et al., 2017; Zufferey et al., 2018). The matter of TAD detection is further complicated by the fact that some TADs, subTADs and loop domains arise via cohesin-mediated loops between convergent CTCF sites, manifested by off-diagonal "corner dots," while the other TADs and loop domains do not (Dixon et al., 2016; Beagan and Phillips-Cremins, 2020; de Wit, 2020; **Figure 1**). Unified classification of

FIGURE 1 | Chromatin domains identified by Hi-C. In Hi-C procedure chromatin of interphase chromosomes is cross-linked and restricted, closely positioned genomic loci are ligated and sequenced by high-throughput sequencing. The frequency of pairwise contacts of genomic loci is represented in the squire matrix or color-coded genomic interaction Hi-C heatmap, which is symmetric along the main diagonal (one halve is shown for simplicity). The A/B compartment segregation is reflected in checkboard-like pattern of the Hi-C heatmap. TADs are identified along the main diagonal of the heatmap as mega- to sub-megabase scale triangle domains with enriched interactions of genomic loci within a domain. Nested subTADs shown within the first TAD are the main features of mammalian Hi-C heatmaps. Corner-dots represent loops between TAD borders, mediated by CTCF and cohesin.

TADs and self-interacting domains could simplify comparison of data from multiple studies and orthogonal approaches, including microscopy.

While the functional significance of TADs in gene regulation has convincing experimental evidence, the physical nature and chromatin structural counterparts of TADs remain enigmatic (Szabo et al., 2018). Some data indicated that TADs could scarcely represent stable units of chromatin organization and appear in population-average Hi-C heatmaps as a mere statistical manifestations of different permitted chromatin conformations regulated by architectural proteins at the level of individual cells (Giorgetti et al., 2014; Dixon et al., 2016). It was clear that these issues could not be resolved by C-method alone and needed aid from microscopy-based methods. The demand of visualization of small neighboring genomic regions has stimulated the development of chromatin imaging by FISH approaches with broad involvement of confocal and super-resolution optical microscopy (Boettiger and Murphy, 2020; Cardozo Gizzi et al., 2020).

Hi-C – FISH PARADOX

From the very beginning of TAD studies, microscopy and FISH-based approaches were used as complementary methods to verify the patterns and "structures" seen in Hi-C heatmaps (Nora et al., 2012; Sexton et al., 2012; Rao et al., 2014). Intuitively, two genomic loci that exhibited higher contact frequency when analyzed by C-based approaches should be found closer in a nuclear space as determined by distance measurement. The inverse correlation between loci contact frequency and 3D-distance was indeed observed in a number of studies (Nora et al., 2012; Giorgetti et al., 2014). However, few loci escaped

this correlation, showing longer distances at high-interaction frequency sites (Williamson et al., 2014). Somewhat contradictory results from FISH-based and Hi-C methods provoked discussion in the field on how the data from these two orthogonal approaches could be cross-validated and reconciled (Dekker, 2016; Giorgetti and Heard, 2016; Fudenberg and Imakaev, 2017). As was noted, "Hi-C – FISH paradox" emerges from the intrinsic variability in physical proximity between two linearly distant genomic loci in single cells, so that their direct contact is a relatively rare event, still registered as a significant interaction by Hi-C in a million cell population (Finn et al., 2019; Shi and Thirumalai, 2019).

From this point of view, high-throughput power of Hi-C turns into a "drawback" as this approach is not able to predict chromatin folding and interactions in a particular cell nucleus. To overcome this limitation, significant efforts have been made in the development of single cell Hi-C and complementary techniques (Ulianov et al., 2017). What appeared from single cell Hi-C experiments is that individual pairwise interactions from different cells were highly variable. At the same time, cumulative heatmaps obtained from dozens or hundreds of analyzed cells generally recapitulated the patterns of conventional Hi-C maps (Nagano et al., 2013; Ramani et al., 2017; Stevens et al., 2017). Additionally, the "Hi-C – FISH paradox" could be solved by high-throughput FISH-imaging of multiple genomic loci in hundreds of cells, followed by averaging of pairwise distances between loci and generation of distance proximity matrices, which could be directly compared to Hi-C contact heatmaps (Boettiger and Murphy, 2020; Hu and Wang, 2021). Moreover, C-methods generally capture very close genomic contacts within nuclear space, which is defined by the paraformaldehyde crosslinking radius (presumably 10–100 nm). Thus, to obtain a better correlation between FISH and Hi-C data one could look in a

range of distances that are clearly beyond the diffraction limits of the conventional optical microscopy. FISH data obtained by super-resolution microscopy and specific probes demonstrate very close concordance with C-data (Szabo et al., 2020).

FISH AS A AN EFFICIENT APPROACH FOR CHROMATIN DOMAIN VISUALIZATION

The efficiency of FISH, as a direct metlhod of DNA-target visualization in fixed cells and tissues, relies on the targeting capacity of the probe and fluorescent dyes used for direct probe labeling or indirect probe detection (Beatty et al., 2002). FISH probes could detect targets from individual genes (few kb) to whole chromosomes and genomes. However, due to the high linear DNA packaging ratio in the interphase nuclei (1:300–1:3000), most individual genes are under the resolution limits of conventional fluorescent microscopy (∼200–250 nm) (Lawrence et al., 1990). This means that morphological details and distance measurements below this range are difficult and inconsistent, as any object would appear as a blurred point due to the diffraction of light. Visualization of genomic regions smaller than ∼10 kb is particularly demanding and usually requires extensive probe design or signal amplification (Schriml et al., 1999; Rogan et al., 2001; Yamada et al., 2011; Beliveau et al., 2015; Ni et al., 2017).

Another resolution-related and genome compaction issue is a genomic distance between simultaneously visualized genomic regions. Directly neighboring small genomic regions detected by FISH probes coupled to different dyes would apparently appear as one spot of co-localized signals. Depending on the epigenetic status of the visualized region (extended euchromatic or compact heterochromatic) 10–50 kb of linear genomic distance is required to discriminate regions as individual signals within the interphase nucleus to allow accurate 3D-distance measurements (Yokota et al., 1997). Thus, the rationale behind probe design is a key to comparative studies between Hi-C and FISH and to the visualization of structures seen in Hi-C maps. In further sections, we are reviewing the most common probes used for verification of C-based data and recent innovations in probe design.

While, theoretically, many probes could be used for simultaneous hybridization, due to the limited number of detection fluorochromes and the architecture of imaging systems, most often 1–3 genomic targets are visualized at one round of hybridization during FISH experiment. This major limitation of FISH is partly solved by applying combinatorial fluorochrome schemes for probe detection resulting in mixed colors or repeated sequential hybridization of samples with probes of interest (Ried et al., 1992; Hu and Wang, 2021). The current progress of FISH throughput in both the number of simultaneously detected probes and the number of cells analyzed in one experiment is tightly linked to the development of automated microfluidic systems (Huber et al., 2018). Apparently, probe and microscopy choice depends on the purposes of a particular experiment, the resolution and throughput that should be obtained (Gelali et al., 2018).

For verification of medium to low resolution Hi-C-derived data within and between megabase-sized contact domains, the resolution of conventional laser-scanning confocal microscopy is sufficient (**Figures 2A,B**). However, when assessing chromatin conformation within TADs, subTADs, and particular loop domains, one should consider methods of super-resolution fluorescent (SRM) or electron microscopy. In the last decade, stochastic optical reconstruction microscopy (STORM), photoactivated localization microscopy (PALM), structured illumination microscopy (SIM), focused ion beam scanning electron microscopy (FIB-SEM) were applied intensively in the exploration of chromatin domains within individual cells (Lakadamyali and Cosma, 2015; Birk, 2019; Szydlowski et al., 2019; Shim, 2021; Xie and Liu, 2021). Highlights and practical guidance for the application of these microscopy tools is beyond the scope of the present review and can be found elsewhere (Lambert and Waters, 2017; Schermelleh et al., 2019).

A microscopic image is the main source of information for imaging-based methods; therefore the requirements for image acquisition, equipment settings and image analysis pipelines should be particularly strict (Ronneberger et al., 2008). To obtain reliable data from fine-scale image analysis, possible distortions in the image and systematic errors should be carefully considered, eliminated or correctly adjusted (Ronneberger et al., 2008). The frequency of pairwise genomic loci interactions from C-data could, with some reservations discussed above, be correlated with the distances between the pair of loci from image-based data (**Figures 2B–D**). As such, the main statistics during image analysis were inferred from direct 2D or 3D-distance measurements between loci in a number of individual cells (Finn et al., 2017; **Figure 2E**). Independently of the form of the observed signal, i.e., dot-like or more extended irregular-shaped objects, usually program-assisted segmentation of thresholded signals is applied followed by calculation of centroid coordinates – a proxy of locus nuclear spatial position (Szabo et al., 2021; **Figures 2C,D**). Co-localization analysis between signals is also frequently performed and in some cases is more informative than distance distribution analysis, especially when FISH-targeted regions are genomically close or consecutive (Giorgetti and Heard, 2016; **Figure 2F**). As two-color and multicolor images may suffer from chromatic aberrations of optical systems, the chromatic shift between channels should be estimated and corrected in distance measurements (Kozubek and Matula, 2000). Custom-made scripts, plugins in free image software, and commercial packages are effectively implemented for image adjustment, object segmentation, and analysis. In experimental pipelines where multiple loci are imaged in hundreds of cells, image acquisition, error correction and measurements are fully automated (Su et al., 2020). Wide spectrum of other characteristics could be estimated during statistical analysis of FISH-images, including signal density, volume and 3D-shape, gyration radius of a signal, scaling exponent of power law dependence of genomic to physical distance over an extended genomic region or a chromosome, clustering of loci and proximity to nuclear landmarks, etc. (Boettiger et al., 2016; Wang et al., 2016; Liu et al., 2020; Szabo et al., 2021). The inclusion of microscopy-derived parameters together with C-method-derived

FIGURE 2 | Analysis of TAD properties by FISH-visualization of genomic regions. **(A)** Schematic heatmap of two neighboring TADs in chicken genome and relative positions of three BAC-based FISH probes, which are separated by ~500 kb from each other. **(B)** Confocal microscopy image (maximum projection) of chicken interphase nucleus after 3-color 3D-FISH with BAC-based FISH probes, FISH signals from two homologous chromosomes are enlarged. **(C)** 3D-surface reconstruction of FISH-signals using Imaris (Oxford Instruments) software. **(D)** Locations of signal mass centers (red, green and blue pixels) are calculated from the reconstruction. **(E)** Main statistics is inferred from pairwise 3D-distance measurements between mass centers of imaged regions in a cell sample and could be presented as boxplots or cumulative graphs. **(F)** Signal co-localization statistics is estimated by Pearson correlation coefficient (not shown) or Jaccard index of signal overlap, depicted as boxplot. Probes from the same TAD (green-blue pair) usually exhibit shorter 3D-distances and higher overlapping, than probes from neighboring TADs (red-blue pair). Nucleus is counterstained with DAPI. Scale bars, 5 μm.

parameters in data-driven polymer models of chromatin were shown to enhance the modeling results (Abbas et al., 2019).

FISH-PROBES FOR CHROMATIN DOMAIN VISUALIZATION

Clone-Based Probes

Probes, based on cloned sequences (PACs, BACs, fosmids, etc.) are the most common FISH probes and have been intensively used in interphase cytogenetics and genome architecture studies from the 1980th (Landegent et al., 1987; **Figure 3**, left column). BAC-clones contain inserts of genomic DNA of a particular species in a range of 50–300 kb, large enough for reliable visualization with fluorescent microscopy. Labeled probes are generated from BAC DNA via enzymatic incorporation of modified nucleotides (conjugated with either hapten or

fluorochrome) during nick translation, (DOP)-PCR or whole-genome amplification, including rolling-circle amplification (Bayani and Squire, 2004; Sharma and Meister, 2020). The enzymatic labeling produces shorter probes with a high labeling density of 20–40 dye/hapten-modified nucleotides per kb (Yu et al., 1994). Both the large size of the target and the high density of labels produce strong FISH signals needed for precise quantitative measurements **Figures 2B–D**. However, the quality of the clone-based probes greatly relies on the performance and activity of the enzymes used in the enzymatic labeling as different polymerases incorporate modified nucleotides with variable efficiency (Tasara et al., 2003; Anderson et al., 2005). Likewise, the density of probe labeling depends on the structure of the nucleotide-fluorochrome complex and fluctuates significantly, when different fluorochromes are utilized for the same enzymatic reactions (Zhu et al., 1994; Giller et al., 2003). This limitation is partially solved by applying two-step enzymatic probe labeling, when at the first step amynoallyl-modified nucleotides (5-(3-aminoallyl)-2′-deoxyuridine 5′-triphosphate) are efficiently incorporated into the probe DNA by polymerases and at the second step amino-reactive dyes (for example N-hydroxysuccinimide esters of fluorochromes) are used for binding to amino-groups (Cox and Singer, 2004; Bolland et al., 2013). Alternatively, enzyme-free nucleic acid labeling kits with platinum dye complexes are commercially available.

Being originated from relatively large genomic fragments, clone-based probes may exhibit moderate specificity due to the presence of repetitive sequences (Sealey et al., 1985). When this is the case, the overall specificity of the probes could be increased by adding preannealing of the labeled probe with competitor DNA (Cot-1 DNA) – the fast-renaturing repetitive DNA fraction of the same species, which is needed to suppress repetitive sequences within a probe (Lichter et al., 1988). Unique sequences of clone-based probes could be also enriched by Cot-1 and duplex-specific nuclease-assisted removal of repetitive sequences before enzymatic labeling (Swennenhuis et al., 2012). Considering above-mentioned issues, the protocols of labeling and FISH with particular clone-based probes may require significant efforts toward optimization. While BAC and PAC libraries with large genome coverage are available for many species (see, for example),[1] for some they are sparse or absent. Moreover, in poorly assembled genomes BAC-contigs could be placed incorrectly leading to the need for additional verification of chromosomal position for any particular BAC-clone. In summary, clone-based probes remain the probes of choice for imaging of relatively large genomic regions by both conventional microscopy and super-resolution microscopy due to the probe robustness, relatively low cost, versatile labeling and detection protocols.

PCR-Derived Probes

When the sequence of the DNA region to be visualized by FISH is known, FISH probes could be produced directly from genomic DNA via PCR with specific primers (**Figure 3**, middle column). Amplified products could be labeled by PCR. In this case,

---

[1] https://bacpacresources.org/
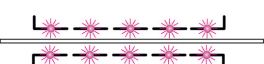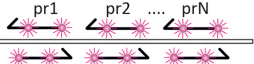
	Clone-based probes	PCR probes	Oligopaint probes
Probe design	genomic BAC, PAC, fosmid libraries dsDNA	pairs of region-specific primers dsDNA	synthetic oligonucleotide libraries dsDNA, ssDNA
Tools	BLAST, BACPAC Resource Centre	BLAT, BLAST, OligoMiner, scFISH, webFISH, PROBER, etc.	Chorus, OligoArray, OligoMiner, iFISH, PROBER, AnthOligo, Probe Dealer, etc.
Probe length	200-500 bp	200-1000 bp	45-150 nt
Min. target size	∼ 40 Kb (fosmids) ∼ 150 Kb (BAC-clones)	∼3 Kb (High definition-FISH)	∼1 Kb
Probe labeling	(DOP)-PCR or nick-translation with hapten/dye coupled nucleotides	PCR with hapten/dye coupled primers or nucleotides; nick-translation	PCR and/or in vitro transcription with hapten/dye coupled primers, secondary oligo probes coupled with hapten/dye
Obstacles	repeat-rich regions, availability of libraries, need for optimization of hybridization conditions, large target size	repeat-rich regions, medium to high cost	repeat-rich regions, high cost, uneven probe coverage (some regions)
Microscopy	epifluorescent, laser-scanning confocal, super-resolution fluorescent, electron, correlation microscopy		

FIGURE 3 | Types of probes used for FISH-visualization of Hi-C chromatin domains. Clone-based probes, PCR probes, and Oligopaint probes are compared in terms of design, labeling, and complications.

hapten, fluorochrome, or amine-modified nucleotide is added to the reaction. PCR amplicons could be also labeled by adding modified nucleotides with terminal deoxynucleotidyl transferase or by nick translation. PCR-generated probes are widely applied for repeated genomic targets (centromere and telomere repeats, ribosomal genes, etc.). However, PCR-generated probes for single-copy genomic targets of several kb in size may require enhanced detection protocols involving signal amplification (Bayani and Squire, 2004). One of the strategies is tyramide signal amplification, which allows identification of targets less than 1 kb (Schriml et al., 1999). This procedure has certain limitations when several targets are visualized simultaneously. Moreover, non-linear and hardly controlled signal amplification for multiple targets may compromise resolution and quantitative methods like co-localization analysis. PCR-generated probes, which are more than 1 kb in size, may require size optimization to improve penetration into the cell, which is achieved by probe digestion with restriction enzymes. Another straightforward strategy for increasing the visibility of single-copy genomic targets is to cover the whole region by smaller probes, produced by PCR with multiple pairs of primers. However, careful bioinformatic analysis of the region and primer selection should be done to exclude amplification of interspersed repeats. A number of tools for picking primers for PCR labeling have been suggested, including PROBER (Navin et al., 2006), webFISH (Nedbal et al., 2012), and scFISH (Rogan et al., 2001; **Figure 3**). For human and mouse genomes, there is a database of specific primers covering the whole genome but omitting repeated sequences. By using these primers, it is possible to produce FISH-probes from 100 to 200 bp

amplicons with a density of 80 amplicons per 100 kb (Bienko et al., 2013; Gelali et al., 2018). This high-definition FISH (HD-FISH) allows detecting 3 kb targets without signal amplification (Bienko et al., 2013).

Oligonucleotide Probes and Oligopaints

Oligonucleotides arrived on the FISH scene with the development of automated oligonucleotide synthesis. However, their use was limited to identifying repetitive sequences, spanning large genomic regions and usually found within the centromere, telomere, and nucleolus organizer regions of chromosomes (Matera and Ward, 1992). Presently, oligonucleotide probes are the probes of choice to detect genomic regions down to several kb, which are clearly smaller than the typical inserts in BAC clones (**Figure 2**, right column). Moreover, this type of probe performs equally well when used to visualize large targets from extended gene loci to whole chromosomes (Boyle et al., 2011; Jiang, 2019).

During the last decade, the development of cost-effective techniques for massive parallel oligo synthesis and accumulation of genome sequencing data have boosted the application of oligonucleotide-based FISH-probes and culminated in the development of Oligopaint technique (Beliveau et al., 2012). Oligopaints, pools of tens of thousands of oligonucleotides, have high specificity, controlled complexity and enable a versatile design matching various detection schemes and microscopy applications (Beliveau et al., 2017). The basic design of the Oligopaint includes a region complementary to the genome target tagged with non-genomic sequences at the 3′ and 5′ ends

(**Figure 4A**). At a minimum, these non-genomic tags contain sequences for primers, required to amplify (and sometimes simultaneously label) the entire Oligopaint library, since after synthesis the concentration of any certain oligonucleotide in the pool is usually low (femtomoles) (Murgha et al., 2014). Oligopaint 3′ and 5′ tag regions could be extended to comprise several primers allowing amplification of a certain oligo sublibrary, which may be useful and cost-effective when differential labeling of several smaller regions within a larger one is needed. Moreover, the non-genomic tags could be extended with additional sequences, depending on the labeling and detection schemes applied in a certain experiment (**Figure 4A**). Several strategies were introduced for the amplification, labeling, and modification of Oligopaint probes depending on the size of the visualized region, nature of the visualized target (DNA, RNA, or both) and species (**Figure 4B**; Beliveau et al., 2012; Chen et al., 2015; Wang et al., 2016; Fields et al., 2019).

The most straightforward method to obtain double-stranded Oligopaint probes from the synthesized library is to amplify the whole library with labeled primers, complementary to tag regions of Oligopaints (Beliveau et al., 2012; **Figure 4B.I**). This approach proved to be productive for the generation of whole chromosomal paints or probes to large chromosomal regions (Bi et al., 2020). Amplified double-stranded Oligopaint probes could be transformed into single-stranded form via introducing the nuclease digestion site into non-genomic tag for subsequent nicking, followed by denaturation and gel purification of the desired strand (**Figure 4B.II**; Beliveau et al., 2012, 2015; Murgha et al., 2014). However, when large amounts of single-stranded labeled probe are needed, a more convenient procedure has been introduced (Murgha et al., 2014; Chen et al., 2015). In this case, the whole library of Oligopaints is first amplified, transcribed *in vitro*, and then reverse-transcribed with specific labeled primers (see IVT-RT on **Figure 4B.III**). To direct enzymatic reactions, the sequences of the PCR-primers and T7 promoter for RNA polymerase should be included into the non-genomic tags of the Oligopaints either during *in silico* design or via touch-up PCR to the already amplified oligo libraries (**Figure 4B.III**). The IVT-RT method and its modifications have become widely applied to amplify and label Oligopaint libraries in different species from plants to human (Boettiger et al., 2016; Gelali et al., 2019; Jiang, 2019).

The recognized advantage of Oligopaint-based probes is the flexibility of design in terms of detection issues (**Figure 4C**). Fluorochrome or hapten-labeled primer(s), complementary to one or both tag-regions, are used during PCR amplification or *in vitro* transcription to obtain labeled oligonucleotides and visualize the target in one round of hybridization (**Figure 4C.I**). Another scheme relies on using 2-step hybridization: the first one with an unlabeled Oligopaint probe and the second one with dye-coupled detection oligo which is complementary to the tag sequences of Oligopaint (Beliveau et al., 2015; Gelali et al., 2018; Fields et al., 2019; **Figure 4C.II**). The second scheme is more versatile as it allows changing the color for visualization of the same probe by changing the label of the detection oligo, without the need of relabeling the entire library. It is widely applied when multiple locus-specific Oligopaint probes

are visualized in the same cell (Wang et al., 2016; Cardozo Gizzi et al., 2019; Fields et al., 2019). Complex detection schemes with activator and photoswitchable reporter dyes should be taken into consideration during design of the Oligopaint libraries for high-resolution imaging of small-scale genomic targets by single-molecule localization microscopy (SMLM) methods, like STORM (**Figure 4C.III**) or DNA point accumulation for imaging of nanoscale topography (DNA-PAINT; **Figure 4C.IV**; Boettiger et al., 2016; Beliveau et al., 2017; Bintu et al., 2018; Nir et al., 2018).

Construction of Oligopaint FISH probes for a certain genomic region starts with the generation of a database of oligonucleotides, tiling the primary sequence, and subsequent optimization of the oligonucleotide pool to ensure the needed specificity and complexity of the probes (Beliveau et al., 2012). There lay few pitfalls hindering the application of Oligopaints. Regions with high repeat content, erroneously or poorly assembled, could hardly be unambiguously covered by oligos and are excluded. Moreover, the application of Oligopaints is confined to popular model organisms with well-assembled genomes. Another difficulty concerns an extensive bioinformatic expertise to decide on the essential parameters needed to filter the initial oligonucleotide pool, even though multiple tools for designing custom arrays of oligonucleotides for genomic regions have been suggested, including OligoArray (Rouillard et al., 2003), PROBER (Navin et al., 2006), Chorus (Han et al., 2015), OligoMiner (Beliveau et al., 2018), iFISH (Gelali et al., 2019), AnthOligo (Jayaraman et al., 2020), ProbeDealer (Hu et al., 2020). The array-synthesized Oligopaint probes are expensive, compared to BAC- or PCR-based probes for the same-sized genomic regions; however, the cost of probes per hybridization could be comparable if high-throughput FISH is performed (see for discussion Beliveau et al., 2012; Boettiger and Murphy, 2020). Among many other parameters used for optimization of the Oligopaint library, the density of oligos per kb should be thoroughly streamlined. It not only influences the size of the Oligopaint library and therefore its cost, but also specifies the reliability of target detection during FISH. As a general principle – the smaller the target region to be visualized, the denser oligo coverage of the region is required to obtain a robust FISH-signal. Practically, the highest density of 15–20 oligos/kb is needed to detect regions from several kb to several dozen kb (Beliveau et al., 2012; Gelali et al., 2019), 10–15 oligos/kb sufficiently detect regions from one to several Mb, while only 0.1–5 oligos/kb are shown to be enough for dozen Mb-sized regions or whole chromosomal paints (Han et al., 2015; Rosin et al., 2018; Jiang, 2019; Bi et al., 2020).

In summary, FISH probes, based on pools of *in silico* designed synthetic Oligopaint libraries progressively displace clone-based probes in experimental designs where high-throughput single cell visualization of multiple and small genomic regions with maximal resolution is required (**Figure 2**). FISH with Oligopaint probes is an extremely rapidly evolving field in terms of Oligopaint design for multiplying the number of simultaneously targeted regions, protocol adjustments for detection of RNA and proteins, or both. For these reasons, Oligopaints are widely applied to assess the organization
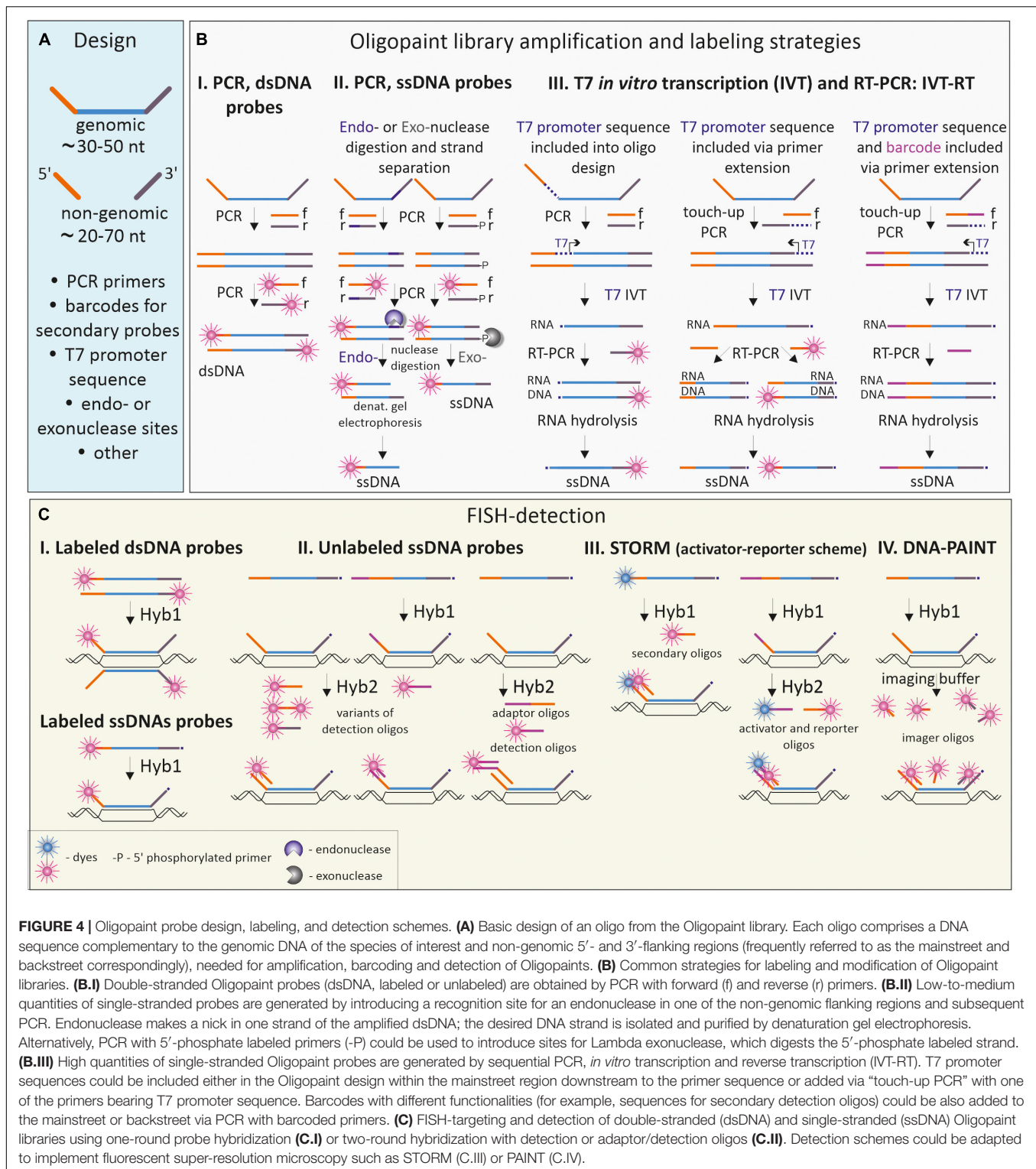
FIGURE 4 | Oligopaint probe design, labeling, and detection schemes. **(A)** Basic design of an oligo from the Oligopaint library. Each oligo comprises a DNA sequence complementary to the genomic DNA of the species of interest and non-genomic 5′- and 3′-flanking regions (frequently referred to as the mainstreet and backstreet correspondingly), needed for amplification, barcoding and detection of Oligopaints. **(B)** Common strategies for labeling and modification of Oligopaint libraries. **(B.I)** Double-stranded Oligopaint probes (dsDNA, labeled or unlabeled) are obtained by PCR with forward (f) and reverse (r) primers. **(B.II)** Low-to-medium quantities of single-stranded probes are generated by introducing a recognition site for an endonuclease in one of the non-genomic flanking regions and subsequent PCR. Endonuclease makes a nick in one strand of the amplified dsDNA; the desired DNA strand is isolated and purified by denaturation gel electrophoresis. Alternatively, PCR with 5′-phosphate labeled primers (-P) could be used to introduce sites for Lambda exonuclease, which digests the 5′-phosphate labeled strand. **(B.III)** High quantities of single-stranded Oligopaint probes are generated by sequential PCR, *in vitro* transcription and reverse transcription (IVT-RT). T7 promoter sequences could be included either in the Oligopaint design within the mainstreet region downstream to the primer sequence or added via "touch-up PCR" with one of the primers bearing T7 promoter sequence. Barcodes with different functionalities (for example, sequences for secondary detection oligos) could be also added to the mainstreet or backstreet via PCR with barcoded primers. **(C)** FISH-targeting and detection of double-stranded (dsDNA) and single-stranded (ssDNA) Oligopaint libraries using one-round probe hybridization **(C.I)** or two-round hybridization with detection or adaptor/detection oligos **(C.II)**. Detection schemes could be adapted to implement fluorescent super-resolution microscopy such as STORM (C.III) or PAINT (C.IV).

of TADs, A/B compartments and other chromatin domains in different model organisms (Boettiger and Murphy, 2020; Hu and Wang, 2021). Moreover, this type of FISH-probe is used to address questions of chromatin fiber organization within highly compacted metaphase chromosomes (Kubalová

et al., 2021). In further sections, we briefly review major insights from FISH-imaging, which together with 3C-based methods aided pieces to the puzzle of spatial organization of chromatin domains in human, mouse, *Drosophila* and other model genomes.

FISH-VISUALIZATION OF CHROMATIN DOMAINS IN MODEL ORGANISMS

Visualization of Chromatin Domains in Mammals

In the first papers, conceptualizing A/B compartments and TADs, FISH was used to visualize regions belonging to contact domains in mouse and human cell lines (Lieberman-Aiden et al., 2009; Dixon et al., 2012; Nora et al., 2012). Since then FISH-based visualization has become a "gold standard" not only to verify C-based data but to analyze the spatial architecture of particular chromatin domains (Giorgetti and Heard, 2016; Bintu et al., 2018). FISH signal evaluation gives an opportunity to test various domain properties (for instance, self-confinement, border-insulation, or large-scale association) within the nuclear context in a single cell. 3D distances between linearly equidistant genomic targets are shorter when measured in the same TAD than between neighboring TADs in numerous regions analyzed so far (Nora et al., 2012). Moreover, BAC-probes covering the whole TAD or several sequential TADs, such as those in the mouse *HoxD* gene cluster, were often discerned as separate globular domains, still variable in shape (Fabre et al., 2015). For *HoxD* gene cluster, it was also shown that the overall morphology of either the extended or more compacted *HoxD* regions does not necessarily correlate with the transcriptional state of the locus (Fabre et al., 2015). FISH-targeting of genes and regulatory sequences belonging to TAD within a α-*globin* locus in mouse embryonic stem cells (mESCs) and differentiating erythroblasts showed that distinct domain shapes and specific *cis*-contacts are established before transcriptional activation (Brown et al., 2018). Systematic studies assessing 3D-distances between BAC-probes to the same or different TADs across many chromosomal regions clearly demonstrated the local variability of chromatin folding at the level of TADs in individual cells (Finn et al., 2019). As it was shown for TADs, regions belonging to the same compartment tended to be closer in a nuclear space (Lieberman-Aiden et al., 2009). However, while these studies illuminated the spatial organization of particular genomic loci, they still gave little idea on the cytological equivalents of various contact domains. Recently, many questions regarding the presence of TADs, their physical parameters, spatial organization, and segregation of A/B compartments *in cis* and *in trans* as well as their relation to other nuclear domains, were addressed directly due to development of novel chromatin imaging technologies enabling tracing of the chromatin paths within the nucleus (Hu and Wang, 2021; **Figure 5**).

Chromatin tracing combines Oligopaint probes, multiplexed FISH-imaging of dozens or even hundreds of small genomic loci within a contiguous chromosome segment, and image analysis tools to visualize the chromatin paths from the subTAD to chromosomal level in individual cells (**Figure 5A**). Chromatin tracing also allows measuring pairwise spatial distances between multiple imaged loci and constructing heatmap matrices (both for a single cell and averaged between thousands of cells) similar to contact frequency maps in Hi-C or 5C. These heatmaps reflect the mean distances between the imaged loci or the frequency

of their proximity (Wang et al., 2016; Bintu et al., 2018; Nir et al., 2018; Su et al., 2020). Interestingly, on such distance- and proximity maps, generated for several chromosomes (20, 21, 22, X) in IMR90 human lung fibroblasts, there were distinct compartments that highly correlated with A and B compartments on Hi-C maps (Wang et al., 2016; Su et al., 2020). While the relative spatial distribution of loci belonging to different compartments varied between cells, the tendency for segregation of loci into A and B higher order domains were clearly observed. Similarly, in GM23248 human skin fibroblasts, super-resolution microscopy demonstrated that segments of the same type of compartments within the ∼8 Mb region on chromosome 19 clustered together and consisted of more distinct chromatin bundles (Nir et al., 2018).

Recent coupling of the chromatin tracing technology with RNA-multiplexed error-robust FISH (RNA-MERFISH) and immunofluorescence staining made it possible to simultaneously visualize chromosomal loci, RNA and nuclear domains (for example, nucleoli and lamina) in mouse embryonic liver cells (Liu et al., 2020) and human IMR90 cells (Su et al., 2020). This opened up wide opportunities to relate transcription and chromatin compartmentalization the individual genomic regions. As it was shown for mouse chromosome 19, its distance-derived compartment profile differed among certain liver cell types. Significant rise in the expression of genes harbored in particular TADs was associated with an increase in A-to-B compartment ratio of the locus; however, the increase in A-to-B ratio itself was not mandatory for changes in gene expression (Liu et al., 2020). In human cells, A-to-B ratio was higher around transcribed genes, nascent transcripts of which were visualized along chromosome 21 together with their chromatin walk (Su et al., 2020). Genome-wide examination of 50 kb loci and transcripts of their genes also showed that genes experiencing high transcriptional activation resided in A compartment (Su et al., 2020). In this latter state of the art study, introducing DNA-MERFISH method (**Figure 5B**), concurrent imaging and analysis of more than a thousand loci from all human chromosomes demonstrated extensive *trans* interactions of the loci. More specifically, interchromosomal and long-range intrachromosomal contacts occurred preferentially between genomic loci belonging to A-compartment, while short-range intrachromosomal contacts (in a scale below 70 Mb) occurred preferentially between genomic loci belonging to B-compartments. Multiple loci imaging also confirmed earlier observations on enrichment of genomic loci belonging to B-compartments near the nuclear lamina and nucleolus, and enrichment of genomic loci belonging to A-compartment in proximity to nuclear speckles (Chen et al., 2018; Quinodoz et al., 2018).

Taking into account the almost decade-long extremely intense investigation of genome architecture by C-methods, possibly the most relevant question to ask is whether TADs could be captured by microscopy as chromatin domains somehow insulated from neighboring chromatin? Several recent studies involving chromatin tracing of 5–50 kb regions and both super-resolution and diffraction-limited fluorescent microscopy succeeded in quantitative imaging of chromatin conformation at the level of TADs and other contact domains

FIGURE 5 | Multiplexed imaging of chromatin domains by chromatin tracing and DNA-multiplexed error-robust FISH. **(A)** The pipeline of chromatin tracing techniques with one-color/one locus in a round (shown for simplicity) and "adaptor oligo/detection oligo" detection scheme. A whole library of Oligopaint probes (primary library), targeting one or several chromatin domains, is divided into multiple sublibraries of oligos with unique mainstreets regions. After initial hybridization of primary probes, individual loci with locus-specific mainstreets are detected during sequential rounds of hybridization with locus-specific readout probes. Design of the adaptor oligo, containing regions, complementary to locus-specific mainstreets and to detection oligos, implements the detection scheme, where a universal dye-labeled detection oligo could be used for sequential imaging of all individual loci. Alternatively, mainstreet-specific dye-labeled oligos could be used (not shown on a scheme). To ensure detection of a particular locus at a time, readout probes are photobleached after imaging or washed out during the subsequent FISH-detection round by using additional small oligonucleotides for strand-displacement reactions (not shown on the scheme). **(B)** The pipeline of DNA-multiplexed error-robust FISH (DNA-MERFISH). A whole library of Oligopaint probes (encoding probes), targeting chromatin domains is divided into multiple sublibraries of oligos, bearing locus-specific "barcodes," encoded in the unique combination of mainstreet and/or backstreet regions (for simplicity shown by different colors of mainstreet regions only). After initial hybridization of the encoding probe, the individual units (bits) of the locus-specific barcodes are detected during sequential rounds of hybridization with "bit"-specific readout probes (one-color detection with reading one bit at a time is shown) until all bits are detected. To ensure a detection of a particular "bit" at a time, dyes are removed or photobleached after imaging. Imaging of all loci in chromatin tracing and decoding of all locus-specific barcodes allow to identify the spatial coordinates of individual loci and to reconstruct chromatin "folding" within the domain. Generally, due to the combinatorial format of individual locus identification, fewer rounds of readout hybridization are required for imaging of the same number of loci in DNA-MERFISH, compared to chromatin tracing.

(Bintu et al., 2018; Nir et al., 2018; Liu et al., 2020; Su et al., 2020; Beckwith et al., 2021). TAD-like chromatin domains of variable size, compactness, and degree of segregation were indeed observed in single cells (Bintu et al., 2018; Su et al., 2020). By analogy with population-averaged TADs, spatial distances between the foci within microscopically identified single-cell chromatin domains were shorter than between foci of neighboring domains. However, the borders separating one single-cell domain from the others were not permanent and their positions fluctuated from cell to cell (Bintu et al., 2018; Su et al., 2020). Surprisingly, these single-cell chromatin domains were insensitive to cohesin removal (Bintu et al., 2018). Another study used DNA stains, SIM, electron, and correlation microscopy to visualize the chromatin network and utilized denaturation-free

RASER (resolution after single-strand exonuclease resection)-FISH to map the positions of several TADs against chromatin substructures (Miron et al., 2020). As appeared, TADs could fall into chains of delineated chromatin nanodomains of 200–300 nm in size, which were also resistant to cohesin ablation (Miron et al., 2020).

Highly similar results were obtained in an independent study that addressed the internal organization of TADs and utilized Oligopaint-based super-resolution imaging of individual TADs in mESCs (Szabo et al., 2020). The authors suggested that chromatin nanodomains are true physical subunits of TADs, the formation of which is largely stochastic (i.e., variable number of nanodomains per TAD in single cells). Indeed, the histone deacetylase inhibitor trichostatin A disrupted

chromatin nanodomains, changing their volume and number (Szabo et al., 2020). While the size range and other properties of single-cell chromatin domains revealed by chromatin tracing and chromatin nanodomains observed by microscopy generally overlaps, how the two types of domains relate to each other remains to be elucidated.

Nevertheless, chromatin domain profiles, essentially corresponding to TAD profiles in Hi-C maps, emerged after averaging spatial distance (or proximity frequency) matrices between hundreds of cells (Bintu et al., 2018; Su et al., 2020). Furthermore, borders of the "averaged" single cell chromatin domains were marked by CTCF and cohesin binding sites and were sensitive to cohesin depletion. These experiments clearly showed that in individual cells chromatin folding at the level of TADs was highly variable, but certain regions exhibited higher probability of domain border formation. Some evidence indicates that the relationship between TADs and compartments in mammalian genomes is more complex than simply hierarchical and that the formation of TADs and compartments are guided by different mechanisms (Mirny et al., 2019). In line with this, it was shown that apart from single-cell chromatin domains of "pure" A- or B compartment type, a significant number of single-cell domains comprised different proportions of both types (Su et al., 2020). How compartmentalization communicates with loop extrusion-assisted TAD formation remains an expanding field of research.

Heterogeneity of chromatin-folding structures within distinct loop domains has also been demonstrated in human cell line by ISH combined with serial block-face scanning electron microscopy (3D-EMISH) (Trzaskoma et al., 2020), FISH with interferometric PALM (iPALM; Jufen Zhu et al., 2019) and loop tracing with DNA-PAINT in non-denaturing conditions (Beckwith et al., 2021). In the 3D-EMISH study, BAC-probes covering the 1.7-Mb region on chromosome 7 in lymphoblastoid GM12878 cells were detected with 1.4-nm-thick streptavidin-conjugated fluoronanogold, followed by analysis of density center distribution in reconstructed ultrastructural serial images of targeted chromatin region. While BAC-probes used in this study could not discriminate between three distinct CTCF-bordered loop domains, identified by ChIA-PET (Chromatin Interaction Analysis by Paired-End Tag sequencing) within this region, 3D-EMISH generally captured from one to four microscopically identified domains, which had highly variable structure and volume (Trzaskoma et al., 2020). Similarly, imaging-based models of a single 13 kb loop and its 10 kb flanking regions in T-cell receptor alpha locus, probed by FISH and iPALM, showed multiple loop conformations in single cells. Still, the pairwise distances for most of the conformations reproduced inverse correlation between frequency of interactions and distance when compared with Hi-C and ChIA-PET data (Jufen Zhu et al., 2019).

A recent study, combining non-denaturing RASER-FISH, DNA-PAINT and chromatin tracing at single-loop scale (kb to Mb region near the *Myc* locus), confirmed some conclusions from a loop extrusion model of TAD formation (Beckwith et al., 2021). Specifically, while the folding of chromatin fibers is intrinsically random and variable from cell to cell, structural

elements in CTCF-bound sites interfere with chromatin fiber random coiling and organize chromatin into loops and self-interacting domains, seen in cell population averages.

From impressive studies of chromatin domains by direct imaging it can be suggested that when the scope is shifted from population-based observations to single cells, stable patterns appear blurred and chromatin contacts – variable. While being spotted as compact globular chromatin domains, mammalian TADs and the loops that build them are something other than deterministically persistent structures with stable borders. Structural variability of chromatin organization could be linked to the variability of functional outcome and adds another level to genome regulation.

Visualization of Chromatin Domains in *Drosophila*

Being for more than a century a model organism in genetic laboratories, the fruit fly *Drosophila* has the most thoroughly characterized genome among invertebrates (Tweedie et al., 2009; Jennings, 2011). *Drosophila* genome, examined by C-methods in different cell types, possesses both TADs and compartments (Hou et al., 2012; Sexton et al., 2012; Rowley et al., 2017). This fact led to the conclusion that these levels of chromatin folding could be common for Metazoa. However, in some respect, *Drosophila* TADs are clearly distinct from mammalian ones. Notably, domain borders and contacts within TADs in *Drosophila* generally do not rely on looping interactions promoted by CTCF and cohesin (Ulianov et al., 2016; Matthews and White, 2019). In *Drosophila*, TAD formation is rather driven by transcription (Hou et al., 2012; Rowley et al., 2017), histone modifications (Sexton et al., 2012; Ulianov et al., 2016; El-Sharnouby et al., 2017), insulator elements and bound proteins (Ramírez et al., 2018; Wang et al., 2018; Bag et al., 2021). Along the *Drosophila* genome, large repressed TADs generally alternate with active regions (also known as boundary regions, inter-TADs or active TADs), occupied by smaller contact domains (subTADs, mini-domains) (Rowley et al., 2017; Ramírez et al., 2018; Wang et al., 2018; **Figure 6A**). A collective term "compartment domains" was coined for all identified *Drosophila* contact domains due to the correlation of TAD profile with epigenetic profiles and clear segregation of contact domains into two types, corresponding to either A or B compartments (Rowley et al., 2017).

Given the small size of both *Drosophila* genome and TADs/compartment domains themselves, efficient visualization of chromatin domains in the interphase nucleus has been achieved by super-resolution microscopy. In correspondence with the emerging picture of epigenetically specified compartment domains, microscopic observations identified repressed TADs as distinct nanometer-sized domains and active regions as extended chains of dotted subdomains (Cattoni et al., 2017; Szabo et al., 2018; **Figures 6A–C**). According to the spatial parameters of chromatin folding deduced from imaging, repressed TADs demonstrate a higher degree of chromatin compaction (Boettiger et al., 2016; Szabo et al., 2018). Interestingly, knockdown of several PcG-proteins, associated with repressed H3K27Me3 epigenetic domains in *Drosophila*,

FIGURE 6 | Imaging of TADs in Drosophila interphase cells and polytene chromosomes. **(A)** Schematic heatmap of two large consecutive TADs from X chromosome in male embryos and relative positions of PCR-derived and Oligopaint FISH probes. **(B)** Confocal microscopy images (z-slices) of several nuclei after 3-color FISH with genomically equidistant probes (red, green, blue), used for inter-probe 3D distance measurement (not shown) and verification of contacts within and between TADs (adapted with permission from Figure 3B in Szabo et al., 2018). In most cells probes from one TAD are closer in a nuclear space, than probes from neighboring TADs. **(C)** 3D-SIM images (maximum projections) of several nuclei after 2-color FISH with Oligopaint probes (magenta and cyan), targeting TAD1 and TAD2 correspondingly (adapted with permission from Figure 3F in Szabo et al., 2018). In most nuclei two large TADs appear as distinct globular nanocompartments. **(D)** Schematic heatmap of a region with one large TAD from chromosome 3R in salivary gland cells of wandering third instar larvae and relative positions of PCR-derived FISH probes. **(E)** Images of polytene chromosome 3R (fragment) after 2-color FISH with probes covering TAD center (red) and TAD borders (green) (adapted with permission from Figure 4C in Eagen et al., 2015). Probe to TAD is observed within a band, indicated by arrow and identifier, while probes to border regions are in the flanking interbands. Nuclei and chromosomes are counterstained with DAPI (B, gray; E-blue). Scale bars: **(B,C)** –1 μm, **(E)** –2 μm.

led to partial disassembly of repressed domains to more open conformations, indicating the role of PcG-proteins in domain maintenance (Boettiger et al., 2016).

Two cognate approaches to chromatin tracing combined with RNA detection (Hi-M and ORCA) allowed to visualize several adjacent TADs in *Drosophila* embryos (Cardozo Gizzi et al., 2019; Mateo et al., 2019). Multiplexed sequential imaging approach (Hi-M) allowed tracing of 21 loci within the ~350 kb region of two TADs, one of which contained genes essential for development and expressed during the early stages of zygotic genome activation. Averaged distance proximity maps constructed for this region highly correlated with Hi-C maps and followed the pattern of mitotic disappearance and reappearance of TADs during zygotic genome activation. Transcriptional activation of genes within microscopically identified TADs leads to perturbation of TAD structure (Cardozo Gizzi et al., 2019). In another study, optical reconstruction of chromatin architecture (ORCA) tiled 100–700 kb regions of the *bithorax* complex (*BX-C*) in several differentiating cell types of *Drosophila* embryos, discriminated by simultaneous mapping of 30 RNA species. This approach allowed to track cell type specific changes in the microscopically identified TAD patterns in *BX-C* locus and to correlate them with changes in epigenetic status accompanied by transcription activation. Upon sequential activation of *BX-C* genes, the inactive TAD substantially contracted and its boundary moved to the right, separating the still inactive H2K27Me3 and PcG-rich chromatin from the active regions. Importantly, smaller TADs appeared within the active region, bounding distinct genes

and their regulatory regions. The borders of these TADs were independent of PcG activity, but contained CTCF and CP190, indicating that CTCF may also play a role as an insulator of contact interactions at least in some of the chromatin domains in *Drosophila* (Mateo et al., 2019).

Apparently, the most illustrative "cytological" interpretation of TADs as structural chromatin domains is the banded pattern of polytene chromosomes in *Drosophila*. Early light and electron microscopy observations of polytene chromosomes from salivary glands and other tissues demonstrated that dense (black and gray) bands and more diffuse interbands alternate along the length of these extended interphase chromosomes (Kolesnikova, 2018). Persistent morphology of polytene chromosomes gave birth to the idea that certain principles of interphase chromosome folding can be encoded in polytene "barcodes" (Vatolina et al., 2011). Indeed, Hi-C of polytene chromosomes disclosed TADs as conserved genome architectural features between polytene and conventional somatic cell nuclei (Eagen et al., 2015). Substantial overlapping of TADs with polytene bands and interTAD regions with interbands has been demonstrated (Eagen et al., 2015; Ulianov et al., 2016; Stadler et al., 2017). FISH probes to TAD borders and internal regions perfectly mapped to interbands and bands correspondingly (Eagen et al., 2015; **Figures 6D,E**). Thus, *Drosophila genome* is characterized by a large degree of correlation between TADs/compartment domains seen in population-averaged Hi-C maps and microscopically visualized chromatin nanodomains in individual cells, as well as by correlation between TADs/compartment domains and

polytene bands. In line with this conclusions, drawn from imaging, single-cell Hi-C of *Drosophila* BG3 cell line showed that ~40% of TADs are conserved between individual cell nuclei (Ulianov et al., 2021). In comparison, in mammals, single-cell TADs are highly variable and comprise multiple nanodomains. Whether this discrepancy in nanoscale chromatin folding between *Drosophila* and mammals could be attributed to the loop extrusion mechanism or the basic differences in genome size and distribution of regulatory elements and genes remains to be elucidated. In this respect, imaging of chromatin topology in various species with differing genome structures seems to be of great value.

Visualization of Chromatin Domains in Other Model Organisms

In Hi-C heatmaps of other representatives of Vertebrata, including fish (*Danio rerio*) (Kaaij et al., 2018), amphibians (*Xenopus tropicalis*) (Niu et al., 2021) and birds (*Gallus gallus*) (Fishman et al., 2019), compartments, TADs and loops are readily discerned. Avian chromatin domains are of special interest in terms of FISH-visualization since certain cell types demonstrate dramatic changes in 3D-genome organization. Indeed, during chicken erythropoiesis, typical TADs disappear, while long-range interactions between distant genomic loci come into place (Fishman et al., 2019). Moreover, similar to polytene chromosomes found in *Drosophila* ovarian nurse cells, avian growing oocytes bear giant transcriptionally active lampbrush chromosomes with a distinct chromomere-loop structure (Gaginskaya et al., 2009). FISH-based approaches are now applied to establish a correspondence between meiotic lampbrush chromomeres and chromatin domains in the interphase nucleus (Krasikova et al., 2019; Zlotina et al., 2020).

TAD-like self-associating domains and/or compartments have been found in other widely studied model organisms from different taxa, including yeasts (*Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*), worms (*Caenorhabditis elegans*), plants and even prokaryotes (Rowley and Corces, 2016; Dong et al., 2017). These TAD-like domains, being much alike in appearance (triangles along the diagonal in Hi-C maps), are highly variable in size, chromosomal distribution, and functional significance and may be shaped by diverse factors and mechanisms (Dekker and Heard, 2015; Rada-Iglesias et al., 2018; Dong et al., 2020). For example, in *C. elegans*, X chromosomes of XX hermaphrodite animals consist of self-interacting TAD-like domains. Boundaries of many of these self-interacting domains are occupied by the dosage compensation complex (DCC), a condensin placed on *rex* sites (Crane et al., 2015). FISH-probes flanking TAD boundaries confirmed the insulation property of *rex* sites, which was disrupted upon DCC loss (Crane et al., 2015).

Large-scale interphase organization of plant chromosomes, segregation of eu- and heterochromatic chromosomal regions and their positioning relative to nuclear landmarks, such as nucleolus or nuclear periphery, have been meticulously probed by FISH and microscopy in several model plants (Pecinka et al., 2004; reviewed in Schubert and Shaw, 2011). Plant chromosome territories in nuclei of different species and/or tissues could display a large variety of spatial conformations from Rabl configuration to plant-specific "Rosette" configuration (Rodriguez-Granados et al., 2016). FISH with BAC-clone derived probes showed that in *Arabidopsis* each chromosome territory forms compact heterochromatic chromocenter "core" surrounded by (sub)megabase-sized euchromatic loops (Fransz et al., 2002), which could be involved in long-range interactions (Schubert et al., 2014). Hi-C studies of *Arabidopsis*, rice (*Oryza sativa*), and maize (*Zea mays*) interphase chromatin, revealed an important role of repeat content in shaping local chromatin domains (Dong et al., 2020). In contrast to plants with large genomes, like rice, that exhibit TAD-like domains (Ouyang et al., 2020), *Arabidopsis* features few domain-like structures. At the same time, rice and *Arabidopsis* are characterized by additional functional long-range interacting domains that may be plant-specific – inactive heterochromatic islands (IHIs) or KNOT engaged Elements (KEEs). FISH with probes to IHIs confirmed that KNOT is formed by IHIs that could belong to different chromosomes (Feng et al., 2014; Grob et al., 2014). Oligopaints are intensively used in comparative plant cytogenetics for the development of chromosome-specific painting probes and loci-specific probe sets (Jiang, 2019), and therefore may serve as a promising tool for deciphering fine-scale spatial chromatin architecture in plant nuclei. Given the diversity and peculiar chromatin domain structures in plant species, the application of novel methods of chromatin imaging for plant genomes will be of high priority (Dumur et al., 2019).

CONCLUSION AND PERSPECTIVES

We are currently witnessing large leaps forward in microscopy tools and FISH-based techniques, which attained genomic and spatial resolution unimaginable just several years ago. As a reminiscence of evolution in C-methods, the demand of ever-growing genomic resolution and higher throughput stimulate the development of chromatin imaging toward multiplexing the number of visualized genomic loci together with other nuclear landmarks, increase in the number of cells analyzed, streamlining and unifying of protocol and analysis pipelines. Apart from further expansion of multiplexed sequential FISH (Xiao et al., 2020; Takei et al., 2021), recent coupling of Oligopaint probes for targeting genomic loci and fluorescent *in situ* sequencing (OligoFISSEQ) demonstrated the potential for imaging more targets in fewer rounds of sequencing and with higher resolution, than chromatin tracing and DNA-MERFISH (Nguyen et al., 2020). The gap between imaging-based and sequencing-based methods of spatial genome probing is progressively shrinking. Combination of chromatin imaging by immunofluorescent staining or fluorescent protein tags and Hi-C on the same single cell offers hope for direct juxtaposition of data, obtained on the genome in exactly the same conformation (Lando et al., 2018). However, the convergence of multiplexed high-resolution FISH and Hi-C in one experiment faces some difficulties. Heat denaturation, fixation, and permeabilization may disturb fine-scale chromatin structures (below 1 Mb) (Solovei et al., 2002;

Markaki et al., 2012). Large efforts have been taken to develop probes for hybridization in more physiological and non-denaturing conditions, which maximally preserve the structure of small-scale chromatin domains (Hausmann et al., 2003; Schmitt et al., 2010). Among promising strategies are Cas9-mediated FISH (CASFISH; Deng et al., 2015), RNA-guided endonuclease *in situ* labeling (RGEN-ISL; Ishii et al., 2019) and RASER-FISH (Brown et al., 2018).

FISH, as Hi-C, is typically performed on fixed cells and tissues and captures only snapshots of the chromatin in action. Nevertheless, even in fixed cells, the observed heterogeneity of chromatin topologies at genomic length scales from tens kb to several Mb could mirror the constrained dynamics and plasticity of chromatin fibers (Hansen et al., 2018). Imaging of chromatin in living cells by expression of fluorescent protein-tagged histones revealed high dynamics and variability (in terms of size and shape) of multiple nucleosomal assemblages – chromatin "blobs," speculatively corresponding to self-interacting domains on Hi-C maps (Wachsmuth et al., 2016; Nozaki et al., 2017).

Tracing of individual chromatin contact domains *in vivo* would allow more precise analysis of chromatin fiber behavior, domain persistence time, stochastic and specific interactions, and other parameters, which are crucial for understanding the principles behind chromatin domain formation. Attaining this goal is tightly linked to the development of probes for *in vivo* labeling of genomic loci (both artificially inserted and endogenous), generally based on the operator-repressor methods (LacO/LacI, TetO/TetR), ANCHOR/ParB system, transcription activator-like effectors (TALEs) or clustered regularly interspaced short palindromic repeats (CRISPR)/nuclease-deactivated CRISPR-associated protein 9 (dCas9) technology (reviewed by Eykelenboom and Tanaka, 2020). CRISPR/dCas9 technology utilizes single guide RNAs (sgRNAs) to target complementary DNA locus and recruit dCas9 fused with fluorescent protein toward the DNA/RNA duplex (Chen et al., 2013). CRISPR/dCas9-based technology evolves exceedingly fast in terms of adapted labeling strategies to enhance signal-to-noise ratio and single locus visibility within cell nucleus (Wu et al., 2019). For instance, multicolor and high-resolution live cell tracking of loci and monitoring of inter-loci distances was achieved using either three dCas9 with different sgRNA binding specificity (Ma et al., 2015) or by engineering sgRNA to harbor RNA aptamers, recognized by cognate binding proteins (Clow et al., 2020).

Rapid advances in chromatin imaging and the spreading of these "hi-end" techniques within the scientific community hold promise for decoding the mesoscale spatial and temporal organization of the genome and its multifaceted regulatory capacity in the near future.

AUTHOR CONTRIBUTIONS

AM wrote the original draft, prepared the figures, and edited the manuscript. AK suggested concept of the article, critically revised the manuscript, and the figures. Both authors approved the final version of the manuscript.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Abbas, A., He, X., Niu, J., Zhou, B., Zhu, G., Ma, T., et al. (2019). Integrating Hi-C and FISH data for modeling of the 3D organization of chromosomes. *Nat. Commun.* 10:2049. doi: 10.1038/s41467-019-10005-6

Agard, D., and Sedat, J. (1983). Three-dimensional architecture of a polytene nucleus. *Nature* 302, 676–681. doi: 10.1038/302676a0

Albiez, H., Cremer, M., Tiberi, C., Vecchio, L., Schermelleh, L., Dittrich, S., et al. (2006). Chromatin domains and the interchromatin compartment form structurally defined and functionally interacting nuclear networks. *Chromosome Res.* 14, 707–733. doi: 10.1007/s10577-006-1086-x

An, L., Yang, T., Yang, J., Nuebler, J., Xiang, G., Hardison, R. C., et al. (2019). OnTAD: hierarchical domain structure reveals the divergence of activity among TADs and boundaries. *Genome Biol.* 20:282. doi: 10.1186/s13059-019-1893-y

Anderson, J. P., Angerer, B., and Loeb, L. A. (2005). Incorporation of reporter-labeled nucleotides by DNA polymerases. *Biotechniques* 38, 257–264. doi: 10.2144/05382RR02

Bag, I., Chen, S., Rosin, L. F., Chen, Y., Liu, C. Y., Yu, G. Y., et al. (2021). M1BP cooperates with CP190 to activate transcription at TAD borders and promote chromatin insulator activity. *Nat. Commun.* 12:4170. doi: 10.1038/s41467-021-24407-y

Barr, M. L., and Bertram, E. G. (1949). A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. *Nature* 163, 676–777. doi: 10.1038/163676a0

Bayani, J., and Squire, J. A. (2004). Fluorescence in situ hybridization (FISH). *Curr. Protoc. Cell Biol.* 22:22.4. doi: 10.1002/0471143030.cb2204s23

Beagan, J. A., and Phillips-Cremins, J. E. (2020). On the existence and functionality of topologically associating domains. *Nat. Genet.* 52, 8–16. doi: 10.1038/s41588-019-0561-1

Beatty, B., Mai, S., and Squire, J. (2002). *FISH: A Practical Approach*. New York: Oxford University Press.

Beckwith, K. S., Ødegård-Fougner, Ø, Morero, N. R., Barton, C., Schueder, F., Alexander, S., et al. (2021). Visualization of loop extrusion by DNA nanoscale tracing in single human cells. *bioRxiv* [Preprint] doi: 10.1101/2021.04.12.439407

Belaghzal, H., Dekker, J., and Gibcus, J. H. (2017). Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. *Methods* 123, 56–65. doi: 10.1016/j.ymeth.2017.04.004

Belaghzal, H., Borrman, T., Stephens, A. D., Lafontaine, D. L., Venev, S. V., Weng, Z., et al. (2021). Liquid chromatin Hi-C characterizes compartment-dependent chromatin interaction dynamics. *Nat. Genet* 53, 367–378. doi: 10.1038/s41588-021-00784-4

Beliveau, B. J., Joyce, E. F., Apostolopoulos, N., Yilmaz, F., Fonseka, C. Y., McCole, R. B., et al. (2012). Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl. Acad. Sci. U.S.A.* 109, 21301–21306. doi: 10.1073/pnas.1213818110

Beliveau, B. J., Boettiger, A. N., Avendaño, M. S., Jungmann, R., McCole, R. B., Joyce, E. F., et al. (2015). Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nat. Commun.* 6:7147. doi: 10.1038/ncomms8147

Beliveau, B. J., Boettiger, A. N., Nir, G., Bintu, B., Yin, P., Zhuang, X., et al. (2017). *In situ* super-resolution imaging of genomic DNA with OligoSTORM and OligoDNA-PAINT. *Methods Mol. Biol.* 1663, 231–252. doi: 10.1007/978-1-4939-7265-4_19

Beliveau, B. J., Kishi, J. Y., Nir, G., Sasaki, H. M., Saka, S. K., Nguyen, S. C., et al. (2018). OligoMiner provides a rapid, flexible environment for the design of genome-scale oligonucleotide in situ hybridization probes. *Proc. Natl. Acad. Sci. U.S.A.* 115, E2183–E2192. doi: 10.1073/pnas.1714530115

Belmont, A. S., Braunfeld, M. B., Sedat, J. W., and Agard, D. A. (1989). Large-scale chromatin structural domains within mitotic and interphase chromosomes *in vivo* and *in vitro*. *Chromosoma* 98, 129–143. doi: 10.1007/BF00291049

Berchtold, D., Fesser, S., Bachmann, G., Kaiser, A., Eilert, J. C., Frohns, F., et al. (2011). Nuclei of chicken neurons in tissues and three-dimensional cell cultures are organized into distinct radial zones. *Chromosome Res.* 19, 165–182. doi: 10.1007/s10577-010-9182-3

Bi, Y., Zhao, Q., Yan, W., Li, M., Liu, Y., Cheng, C., et al. (2020). Flexible chromosome painting based on multiplex PCR of oligonucleotides and its application for comparative chromosome analyses in *Cucumis*. *Plant J.* 102, 178–186. doi: 10.1111/tpj.14600

Bienko, M., Crosetto, N., Teytelman, L., Klemm, S., Itzkovitz, S., and van Oudenaarden, A. (2013). A versatile genome-scale PCR-based pipeline for high-definition DNA FISH. *Nat. Methods* 10, 122–124. doi: 10.1038/nmeth.2306

Bintu, B., Mateo, L. J., Su, J. H., Sinnott-Armstrong, N. A., Parker, M., Kinrot, S., et al. (2018). Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* 362:eaau1783. doi: 10.1126/science.aau1783

Birk, U. J. (2019). Super-resolution microscopy of chromatin. *Genes* 10:493. doi: 10.3390/genes10070493

Boettiger, A. N., Bintu, B., Moffitt, J. R., Wang, S., Beliveau, B. J., Fudenberg, G., et al. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* 529, 418–422. doi: 10.1038/nature16496

Boettiger, A., and Murphy, S. (2020). Advances in chromatin imaging at kilobase-scale resolution. *Trends Genet.* 36, 273–287. doi: 10.1016/j.tig.2019.12.010

Bolland, D. J., King, M. R., Reik, W., Corcoran, A. E., and Krueger, C. (2013). Robust 3D DNA FISH using directly labeled probes. *J. Vis. Exp.* 78:50587. doi: 10.3791/50587

Boyle, S., Rodesch, M. J., Halvensleben, H. A., Jeddeloh, J. A., and Bickmore, W. A. (2011). Fluorescence in situ hybridization with high-complexity repeat-free oligonucleotide probes generated by massively parallel synthesis. *Chromosome Res.* 19, 901–909. doi: 10.1007/s10577-011-9245-0

Brant, L., Georgomanolis, T., Nikolic, M., Brackley, C. A., Kolovos, P., van Ijcken, W., et al. (2016). Exploiting native forces to capture chromosome conformation in mammalian cell nuclei. *Mol. Syst. Biol.* 12:891. doi: 10.15252/msb.20167311

Brown, J. M., Roberts, N. A., Graham, B., Waithe, D., Lagerholm, C., Telenius, J. M., et al. (2018). A tissue-specific self-interacting chromatin domain forms independently of enhancer-promoter interactions. *Nat. Commun.* 9:3849. doi: 10.1038/s41467-018-06248-4

Cardozo Gizzi, A. M., Cattoni, D. I., Fiche, J. B., Espinola, S. M., Gurgo, J., Messina, O., et al. (2019). Microscopy-based chromosome conformation capture enables simultaneous visualization of genome organization and transcription in intact organisms. *Mol. Cell* 74, 212–222.e5. doi: 10.1016/j.molcel.2019.01.011

Cardozo Gizzi, A. M., Cattoni, D. I., and Nollmann, M. (2020). TADs or no TADS: lessons from single-cell imaging of chromosome architecture. *J. Mol. Biol.* 432, 682–693. doi: 10.1016/j.jmb.2019.12.034

Cattoni, D. I., Cardozo Gizzi, A. M., Georgieva, M., Di Stefano, M., Valeri, A., Chamousset, D., et al. (2017). Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. *Nat. Commun.* 8:1753. doi: 10.1038/s41467-017-01962-x

Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G. W., et al. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155, 1479–1491. doi: 10.1016/j.cell.2013.12.001

Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S., and Zhuang, X. (2015). RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 348:aaa6090. doi: 10.1126/science.aaa6090

Chen, Y., Zhang, Y., Wang, Y., Zhang, L., Brinkman, E. K., Adam, S. A., et al. (2018). Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J. Cell Biol.* 217, 4025–4048. doi: 10.1083/jcb.201807108

Chubb, J. R., and Bickmore, W. A. (2003). Considering nuclear compartmentalization in the light of nuclear dynamics. *Cell* 112, 403–406. doi: 10.1016/s0092-8674(03)00078-3

Clow, P. A., Jillette, N., Zhu, J. J., and Cheng, A. W. (2020). CRISPR-mediated multiplexed live cell imaging of nonrepetitive genomic loci. *bioRxiv* [Preprint]. doi: 10.1101/2020.03.03.974923

Cremer, T., and Cremer, C. (2006). Rise, fall and resurrection of chromosome territories: a historical perspective. Part I. The rise of chromosome territories. *Eur. J. Histochem.* 50, 161–176.

Cremer, T., and Cremer, M. (2010). Chromosome territories. *Cold Spring Harb. Perspect. Biol.* 2:a003889. doi: 10.1101/cshperspect.a003889

Coleman, W. (1965). Cell, Nucleus, and Inheritance: An Historical Study. *Proc. Am. Philosoph. Soc.* 109, 124–158.

Collas, P., Liyakat Ali, T. M., Brunet, A., and Germier, T. (2019). Finding friends in the crowd: three-dimensional cliques of topological genomic domains. *Front. Genet.* 10:602. doi: 10.3389/fgene.2019.00602

Comings, D. E. (1968). The rationale for an ordered arrangement of chromatin in the interphase nucleus. *Am. J. Hum. Genet.* 20, 440–460.

Cox, W. G., and Singer, V. L. (2004). Fluorescent DNA hybridization probe preparation using amine modification and reactive dye coupling. *Biotechniques* 36, 114–112. doi: 10.2144/04361RR02

Crane, E., Bian, Q., McCord, R. P., Lajoie, B. R., Wheeler, B. S., Ralston, E. J., et al. (2015). Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* 523, 240–244. doi: 10.1038/nature14450

Cullen, K. E., Kladde, M. P., and Seyfred, M. A. (1993). Interaction between transcription regulatory regions of prolactin chromatin. *Science* 261, 203–206. doi: 10.1126/science.8327891

Daban, J. R. (2011). Electron microscopy and atomic force microscopy studies of chromatin and metaphase chromosome structure. *Micron.* 42, 733–750. doi: 10.1016/j.micron.2011.05.002

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380. doi: 10.1038/nature11082

Dixon, J. R., Gorkin, D. U., and Ren, B. (2016). Chromatin domains: the unit of chromosome organization. *Mol. Cell.* 62, 668–680. doi: 10.1016/j.molcel.2016.05.018

de Wit, E. (2020). TADs as the caller calls them. *J. Mol. Biol.* 432, 638–642. doi: 10.1016/j.jmb.2019.09.026

Dekker, J. (2016). Mapping the 3D genome: Aiming for consilience. *Nat. Rev. Mol. Cell Biol.* 17, 741–742. doi: 10.1038/nrm.2016.151

Dekker, J., and Heard, E. (2015). Structural and functional diversity of Topologically Associating Domains. *FEBS Lett.* 589, 2877–2884. doi: 10.1016/j.febslet.2015.08.044

Dekker, J., Marti-Renom, M. A., and Mirny, L. A. (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat. Rev. Genet.* 14, 390–403. doi: 10.1038/nrg3454

Deng, W., Shi, X., Tjian, R., Lionnet, T., and Singer, R. H. (2015). CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells. *Proc. Natl. Acad. Sci. U.S.A.* 112, 11870–11875. doi: 10.1073/pnas.1515692112

Denker, A., and de Laat, W. (2016). The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev.* 30, 1357–1382. doi: 10.1101/gad.281964.116

Dong, P., Tu, X., Chu, P. Y., Lü, P., Zhu, N., Grierson, D., et al. (2017). 3D chromatin architecture of large plant genomes determined by local A/B compartments. *Mol. Plant.* 10, 1497–1509. doi: 10.1016/j.molp.2017.11.005

Dong, P., Tu, X., Liang, Z., Kang, B. H., and Zhong, S. (2020). Plant and animal chromatin three-dimensional organization: similar structures but different functions. *J. Exp. Bot.* 71, 5119–5128. doi: 10.1093/jxb/eraa220

Dowen, J. M., Fan, Z. P., Hnisz, D., Ren, G., Abraham, B. J., Zhang, L. N., et al. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* 159, 374–387. doi: 10.1016/j.cell.2014.09.030

Dubochet, J., Adrian, M., Chang, J. J., Homo, J. C., Lepanlt, J., McDowall, A. W., et al. (1988). Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* 21, 129–228. doi: 10.1017/s0033583500004297

Dumur, T., Duncan, S., Graumann, K., Desset, S., Randall, R. S., Scheid, O. M., et al. (2019). Probing the 3D architecture of the plant nucleus with microscopy approaches: challenges and solutions. *Nucleus* 10, 181–212. doi: 10.1080/19491034.2019.1644592

Eagen, K. P., Hartl, T. A., and Kornberg, R. D. (2015). Stable chromosome condensation revealed by chromosome conformation capture. *Cell* 163, 934–946. doi: 10.1016/j.cell.2015.10.026

Ellison, J. R., and Howard, G. C. (1981). Non-random position of the A-T rich DNA sequences in early embryos of Drosophila virilis. *Chromosoma* 83, 555–561. doi: 10.1007/BF00328279

El-Sharnouby, S., Fischer, B., Magbanua, J. P., Umans, B., Flower, R., Choo, S. W., et al. (2017). Regions of very low H3K27me3 partition the Drosophila genome into topological domains. *PLoS One* 12:e0172725. doi: 10.1371/journal.pone.0172725

Eykelenboom, J. K., and Tanaka, T. U. (2020). Zooming in on chromosome dynamics. *Cell Cycle* 19, 1422–1432. doi: 10.1080/15384101.2020.1757242

Fabre, P. J., Benke, A., Manley, S., and Duboule, D. (2015). Visualizing the HoxD Gene Cluster at the Nanoscale Level. *Cold Spring Harb. Symp. Quant. Biol.* 80, 9–16. doi: 10.1101/sqb.2015.80.027177

Feng, S., Cokus, S. J., Schubert, V., Zhai, J., Pellegrini, M., and Jacobsen, S. E. (2014). Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in Arabidopsis. *Mol. Cell.* 55, 694–707. doi: 10.1016/j.molcel.2014.07.008

Fields, B. D., Nguyen, S. C., Nir, G., and Kennedy, S. (2019). A multiplexed DNA FISH strategy for assessing genome architecture in Caenorhabditis elegans. *Elife* 8:e42823. doi: 10.7554/eLife.42823

Finn, E. H., Pegoraro, G., Shachar, S., and Misteli, T. (2017). Comparative analysis of 2D and 3D distance measurements to study spatial genome organization. *Methods* 123, 47–55. doi: 10.1016/j.ymeth.2017.01.007

Finn, E. H., Pegoraro, G., Brandão, H. B., Valton, A. L., Oomen, M. E., Dekker, J., et al. (2019). Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell* 176, 1502–1515.e10. doi: 10.1016/j.cell.2019.01.020

Fishman, V., Battulin, N., Nuriddinov, M., Maslova, A., Zlotina, A., Strunov, A., et al. (2019). 3D organization of chicken genome demonstrates evolutionary conservation of topologically associated domains and highlights unique architecture of erythrocytes' chromatin. *Nucleic Acids Res.* 47, 648–665. doi: 10.1093/nar/gky1103

Forcato, M., Nicoletti, C., Pal, K., Livi, C. M., Ferrari, F., and Bicciato, S. (2017). Comparison of computational methods for Hi-C data analysis. *Nat. Methods* 14, 679–685. doi: 10.1038/nmeth.4325

Fransz, P., de Jong, J. H., Lysak, M., Castiglione, M. R., and Schubert, I. (2002). Interphase chromosomes in Arabidopsis are organized as well defined chromocenters from which euchromatin loops emanate. *Proc. Natl. Acad. Sci. U.S.A.* 99, 14584–14589. doi: 10.1073/pnas.212325299

Fraser, J., Ferrai, C., Chiariello, A. M., Schueler, M., Rito, T., Laudanno, G., et al. (2015a). Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Mol. Syst. Biol.* 11:852. doi: 10.15252/msb.20156492

Fraser, J., Williamson, I., Bickmore, W. A., and Dostie, J. (2015b). An overview of genome organization and how we got there: from FISH to Hi-C. *Microbiol. Mol. Biol. Rev.* 79, 347–372. doi: 10.1128/MMBR.00006-15

Fudenberg, G., and Imakaev, M. (2017). FISH-ing for captured contacts: towards reconciling FISH and 3C. *Nat. Methods* 14, 673–678. doi: 10.1038/nmeth.4329

Fussner, E., Ching, R. W., and Bazett-Jones, D. P. (2011). Living without 30nm chromatin fibers. *Trends Biochem. Sci.* 36, 1–6. doi: 10.1016/j.tibs.2010.09.002

Gaginskaya, E., Kulikova, T., and Krasikova, A. (2009). Avian lampbrush chromosomes: a powerful tool for exploration of genome expression. *Cytogenet. Genome Res.* 124, 251–267. doi: 10.1159/000218130

Gavrilov, A., Razin, S. V., and Cavalli, G. (2015). In vivo formaldehyde crosslinking: it is time for black box analysis. *Brief. Funct. Genomics* 14, 163–165. doi: 10.1093/bfgp/elu037

Gelali, E., Custodio, J., Girelli, G., Wernersson, E., Crosetto, N., and Bienko, M. (2018). An application-directed, versatile DNA FISH platform for research and diagnostics. *Methods Mol. Biol.* 1766, 303–333. doi: 10.1007/978-1-4939-7768-0_17

Gelali, E., Girelli, G., Matsumoto, M., Wernersson, E., Custodio, J., Mota, A., et al. (2019). iFISH is a publically available resource enabling versatile DNA FISH to study genome architecture. *Nat. Commun.* 10:1636. doi: 10.1038/s41467-019-09616-w

Giller, G., Tasara, T., Angerer, B., Mühlegger, K., Amacker, M., and Winter, H. (2003). Incorporation of reporter molecule-labeled nucleotides by DNA polymerases. I. Chemical synthesis of various reporter group-labeled 2'-deoxyribonucleoside-5'-triphosphates. *Nucleic Acids Res.* 31, 2630–2635. doi: 10.1093/nar/gkg370

Giorgetti, L., Galupa, R., Nora, E. P., Piolot, T., Lam, F., Dekker, J., et al. (2014). Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* 157, 950–963. doi: 10.1016/j.cell.2014.03.025

Giorgetti, L., and Heard, E. (2016). Closing the loop: 3C versus DNA FISH. *Genome Biol.* 17:215. doi: 10.1186/s13059-016-1081-2

Goel, V. Y., and Hansen, A. S. (2020). The macro and micro of chromosome conformation capture. *Wiley Interdiscip. Rev. Dev. Biol.* 2020:e395. doi: 10.1002/wdev.395

Grob, S., Schmid, M. W., and Grossniklaus, U. (2014). Hi-C analysis in Arabidopsis identifies the KNOT, a structure with similarities to the flamenco locus of Drosophila. *Mol. Cell.* 55, 678–693. doi: 10.1016/j.molcel.2014.07.009

Han, Y., Zhang, T., Thammapichai, P., Weng, Y., and Jiang, J. (2015). Chromosome-specific painting in Cucumis species using bulked oligonucleotides. *Genetics* 200, 771–779. doi: 10.1534/genetics.115.177642

Hansen, A. S., Cattoglio, C., Darzacq, X., and Tjian, R. (2018). Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus* 9, 20–32. doi: 10.1080/19491034.2017

Hausmann, M., Winkler, R., Hildenbrand, G., Finsterle, J., Weisel, A., Rapp, A., et al. (2003). COMBO-FISH: specific labeling of nondenatured chromatin targets by computer-selected DNA oligonucleotide probe combinations. *Biotechniques* 35, 564–577. doi: 10.2144/03353rr03

Heitz, E. (1929). Heterochromatin, Chromocentren, Chromomeren. *Ber. Botan. Ges.* 47, 274–284.

Hoang, T. V., Kizilyaprak, C., Spehner, D., Humbel, B. M., and Schultz, P. (2017). Automatic segmentation of high pressure frozen and freeze-substituted mouse retina nuclei from FIB-SEM tomograms. *J. Struct. Biol.* 197, 123–134. doi: 10.1016/j.jsb.2016.10.005

Hou, C., Li, L., Qin, Z. S., and Corces, V. G. (2012). Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol. Cell.* 48, 471–484. doi: 10.1016/j.molcel.2012.08.031

Hsu, T. C., Cooper, J. E., Mace, M. L. Jr., and Brinkley, B. R. (1971). Arrangement of centromeres in mouse cells. *Chromosoma* 34, 73–87. doi: 10.1007/BF00285517

Hu, M., Yang, B., Cheng, Y., Radda, J. S. D., Chen, Y., Liu, M., et al. (2020). ProbeDealer is a convenient tool for designing probes for highly multiplexed fluorescence in situ hybridization. *Sci. Rep.* 10:22031. doi: 10.1038/s41598-020-76439-x

Hu, M., and Wang, S. (2021). Chromatin Tracing: Imaging 3D Genome and Nucleome. *Trends Cell Biol.* 31, 5–8. doi: 10.1016/j.tcb.2020.10.006

Huang, K., Li, Y., Shim, A. R., Virk, R. K. A., Agrawal, V., Eshein, A., et al. (2020). Physical and data structure of 3D genome. *Sci. Adv.* 6:eaay4055. doi: 10.1126/sciadv.aay4055

Huber, D., Voith von Voithenberg, L., and Kaigala, G. V. (2018). Fluorescence in situ hybridization (FISH): History, limitations and what to expect from micro-scale FISH? *Micro Nano Engin.* 1, 15–24. doi: 10.1016/j.mne.2018.10.006

Hutchison, N. J., Langer-Safer, P. R., Ward, D. C., and Hamkalo, B. A. (1982). In situ hybridization at the electron microscope level: hybrid detection by autoradiography and colloidal gold. *J. Cell Biol.* 95, 609–618. doi: 10.1083/jcb.95.2.609

Ing-Simmons, E., and Vaquerizas, J. M. (2019). Visualising three-dimensional genome organisation in two dimensions. *Development* 146:dev177162. doi: 10.1242/dev.177162

Ishii, T., Schubert, V., Khosravi, S., Dreissig, S., Metje-Sprink, J., Sprink, T., et al. (2019). RNA-guided endonuclease – in situ labelling (RGEN-ISL): a fast CRISPR/Cas9-based method to label genomic sequences in various species. *New Phytol.* 222, 1652–1661. doi: 10.1111/nph.15720

Jahn, M., Markert, S., Ryu, T., Ravasi, T., Stigloher, C., Hentschel, U., et al. (2016). Shedding light on cell compartmentation in the candidate phylum Poribacteria by high resolution visualisation and transcriptional profiling. *Sci. Rep.* 6:35860. doi: 10.1038/srep35860

Jayaraman, P., Mosbruger, T., Hu, T., Tairis, N. G., Wu, C., Clark, P. M., et al. (2020). AnthOligo: automating the design of oligonucleotides for capture/enrichment technologies. *Bioinformatics* 36, 4353–4356. doi: 10.1093/bioinformatics/btaa552

Jennings, B. H. (2011). Drosophila – a versatile model in biology & medicine. *Mater. Today* 14, 190–195. doi: 10.1016/S1369-7021(11)70113-4

Jiang, J. (2019). Fluorescence *in situ* hybridization in plants: recent developments and future applications. *Chromosome Res.* 27, 153–165. doi: 10.1007/s10577-019-09607-z

John, H. A., Birnstiel, M. L., and Jones, K. W. (1969). RNA-DNA hybrids at the cytological level. *Nature* 223, 582–587. doi: 10.1038/223582a0

Jufen Zhu, J., Parteka, Z., Lee, B., Szalaj, P., Wang, P., Jodkowska, K., et al. (2019). Super resolution imaging of a distinct chromatin loop in human lymphoblastoid cells. *bioRxiv* [Preprint] doi: 10.1101/621920

Kaaij, L. J. T., van der Weide, R. H., Ketting, R. F., and de Wit, E. (2018). Systemic Loss and Gain of Chromatin Architecture throughout Zebrafish Development. *Cell Rep.* 24, 1–10.e4. doi: 10.1016/j.celrep.2018.06.003

Kempfer, R., and Pombo, A. (2020). Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.* 21, 207–226. doi: 10.1038/s41576-019-0195-2

Kolesnikova, T. D. (2018). Banding Pattern of Polytene Chromosomes as a Representation of Universal Principles of Chromatin Organization into Topological Domains. *Biochemistry* 83, 338–349. doi: 10.1134/S0006297918040053

Kosak, S. T., and Groudine, M. (2004). Form follows function: The genomic organization of cellular differentiation. *Genes Dev.* 18, 1371–1384. doi: 10.1101/gad.1209304

Kozubek, M., and Matula, P. (2000). An efficient algorithm for measurement and correction of chromatic aberrations in fluorescence microscopy. *J. Microsc.* 200, 206–217. doi: 10.1046/j.1365-2818.2000.00754.x

Krasikova, A., Zlotina, A., Maslova, A., Starshova, P., and Kulikova, T. (2019). High-resolution mapping of A/B compartments and topologically associated domains on giant lampbrush chromosomes. *Biopolym. Cell.* 35:175. doi: 10.7124/bc.0009AF

Krietenstein, N., Abraham, S., Venev, S. V., Abdennur, N., Gibcus, J., Hsieh, T. S., et al. (2020). Ultrastructural Details of Mammalian Chromosome Architecture. *Mol. Cell* 78, 554–565.e7. doi: 10.1016/j.molcel.2020.03.003

Kubalová, I., Câmara, A. S., Cápal, P., Beseda, T., Rouillard, J.-M., Krause, G. M., et al. (2021). Helical metaphase chromatid coiling is conserved. *bioRxiv* [Preprint] doi: 10.1101/2021.09.16.460607

Kuznetsova, M. A., and Sheval, E. V. (2016). Chromatin fibers: from classical descriptions to modern interpretation. *Cell Biol. Int.* 40, 1140–1151. doi: 10.1002/cbin.10672

Lakadamyali, M., and Cosma, M. P. (2015). Advanced microscopy methods for visualizing chromatin structure. *FEBS Lett.* 589, 3023–3030. doi: 10.1016/j.febslet.2015.04.012

Lambert, T. J., and Waters, J. C. (2017). Navigating challenges in the application of superresolution microscopy. *J. Cell Biol.* 216, 53–63. doi: 10.1083/jcb.201610011

Landegent, J. E., Jansen in de Wal, N., Dirks, R. W., Baao, F., and van der Ploeg, M. (1987). Use of whole cosmid cloned genomic sequences for chromosomal localization by non-radioactive in situ hybridization. *Hum. Genet.* 77, 366–370. doi: 10.1007/BF00291428

Lando, D., Basu, S., Stevens, T. J., Riddell, A., Wohlfahrt, K. J., Cao, Y., et al. (2018). Combining fluorescence imaging with Hi-C to study 3D genome architecture of the same single cell. *Nat. Protoc.* 13, 1034–1061. doi: 10.1038/nprot.2018.017

Langer-Safer, P. R., Levine, M., and Ward, D. C. (1982). Immunological method for mapping genes on Drosophila polytene chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* 79, 4381–4385. doi: 10.1073/pnas.79.14.4381

Latt, S. A. (1977). Fluorescent probes of chromosome structure and replication. *Can. J. Genet. Cytol.* 19, 603–623. doi: 10.1139/g77-065

Lawrence, J. B., Singer, R. H., and McNeil, J. A. (1990). Interphase and metaphase resolution of different distances within the human dystrophin gene. *Science* 249, 928–932. doi: 10.1126/science.2203143

Li, Y., Eshein, A., Virk, R. K. A., Eid, A., Wu, W., Frederick, J., et al. (2021). Nanoscale chromatin imaging and analysis platform bridges 4D chromatin organization with molecular function. *Sci. Adv.* 7:eabe4310. doi: 10.1126/sciadv.abe4310

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293. doi: 10.1126/science.1181369

Lichter, P., Cremer, T., Borden, J., Manuelidis, L., and Ward, D. C. (1988). Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Hum. Genet* 80, 224–234. doi: 10.1007/BF01790090

Lin, J. R., Fallahi-Sichani, M., and Sorger, P. K. (2015). Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. *Nat. Commun.* 6:8390. doi: 10.1038/ncomms9390

Liu, M., Lu, Y., Yang, B., Chen, Y., Radda, J. S. D., Hu, M., et al. (2020). Multiplexed imaging of nucleome architectures in single cells of mammalian tissue. *Nat. Commun.* 11:2907. doi: 10.1038/s41467-020-16732-5

Ma, H., Naseri, A., Reyes-Gutierrez, P., Wolfe, S. A., Zhang, S., and Pederson, T. (2015). Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc. Natl. Acad. Sci. U.S.A.* 112, 3002–3007. doi: 10.1073/pnas.1420024112

Maeshima, K., Ide, S., and Babokhov, M. (2019). Dynamic chromatin organization without the 30-nm fiber. *Curr. Opin. Cell Biol.* 58, 95–104. doi: 10.1016/j.ceb.2019.02.003

Manuelidis, L., Langer-Safer, P. R., and Ward, D. C. (1982). High-resolution mapping of satellite DNA using biotin-labeled DNA probes. *J. Cell Biol.* 95, 619–625. doi: 10.1083/jcb.95.2.619

Manuelidis, L. (1984). Different central nervous system cell types display distinct and nonrandom arrangements of satellite DNA sequences. *Proc. Natl. Acad. Sci. U.S.A.* 81, 3123–3127. doi: 10.1073/pnas.81.10.3123

Markaki, Y., Smeets, D., Fiedler, S., Schmid, V. J., Schermelleh, L., Cremer, T., et al. (2012). The potential of 3D-FISH and super-resolution structured illumination microscopy for studies of 3D nuclear architecture: 3D structured illumination microscopy of defined chromosomal structures visualized by 3D (immuno)-FISH opens new perspectives for studies of nuclear architecture. *Bioessays* 34, 412–426. doi: 10.1002/bies.201100176

Marshall, W. F. (2002). Order and disorder in the nucleus. *Curr. Biol.* 12, R185–R192. doi: 10.1016/s0960-9822(02)00724-8

Marti-Renom, M. A., Almouzni, G., Bickmore, W. A., Bystricky, K., Cavalli, G., Fraser, P., et al. (2018). Challenges and guidelines toward 4D nucleome data and model standards. *Nat. Genet.* 50, 1352–1358. doi: 10.1038/s41588-018-0236-3

Mateo, L. J., Murphy, S. E., Hafner, A., Cinquini, I. S., Walker, C. A., and Boettiger, A. N. (2019). Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* 568, 49–54. doi: 10.1038/s41586-019-1035-4

Matera, A. G., and Ward, D. C. (1992). Oligonucleotide probes for the analysis of specific repetitive DNA sequences by fluorescence in situ hybridization. *Hum. Mol. Genet.* 1, 535–539. doi: 10.1093/hmg/1.7.535

Matthews, N. E., and White, R. (2019). Chromatin Architecture in the Fly: Living without CTCF/Cohesin loop extrusion?: Alternating chromatin states provide a basis for domain architecture in Drosophila. *Bioessays* 41:e1900048. doi: 10.1002/bies.201900048

McCord, R. P., Kaplan, N., and Giorgetti, L. (2020). Chromosome conformation capture and beyond: toward an integrative view of chromosome structure and function. *Mol. Cell.* 77, 688–708. doi: 10.1016/j.molcel.2019.12.021

Mielańczyk, Ł., Matysiak, N., Klymenko, O., and Wojnicz, R. (2015). "Transmission electron microscopy of biological samples," in *The Transmission Electron Microscope - Theory And Applications*, ed. M. Khan (New York, NY: Wiley), doi: 10.5772/60680

Mirny, L. A., Imakaev, M., and Abdennur, N. (2019). Two major mechanisms of chromosome organization. *Curr. Opin. Cell Biol.* 58, 142–152. doi: 10.1016/j.ceb.2019.05.001

Miron, E., Oldenkamp, R., Brown, J. M., Pinto, D. M. S., Xu, C. S., Faria, A. R., et al. (2020). Chromatin arranges in chains of mesoscale domains with nanoscale functional topography independent of cohesin. *Sci. Adv.* 6:eaba8811. doi: 10.1126/sciadv.aba8811

Misteli, T. (2005). Concepts in nuclear architecture. *Bioessays* 27, 477–487. doi: 10.1002/bies.20226

Mota-Gómez, I., and Lupiáñez, D. G. (2019). A (3D-nuclear) space odyssey: making sense of Hi-C maps. *Genes* 10:415. doi: 10.3390/genes10060415

Murgha, Y. E., Rouillard, J. M., and Gulari, E. (2014). Methods for the preparation of large quantities of complex single-stranded oligonucleotide libraries. *PLoS One* 9:e94752. doi: 10.1371/journal.pone.0094752

Nagano, T., Lubling, Y., Stevens, T. J., Schoenfelder, S., Yaffe, E., Dean, W., et al. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502, 59–64. doi: 10.1038/nature12593

Navin, N., Grubor, V., Hicks, J., Leibu, E., Thomas, E., Troge, J., et al. (2006). PROBER: oligonucleotide FISH probe design software. *Bioinformatics* 22, 2437–2438. doi: 10.1093/bioinformatics/btl273

Nedbal, J., Hobson, P. S., Fear, D. J., Heintzmann, R., and Gould, H. J. (2012). Comprehensive FISH probe design tool applied to imaging human immunoglobulin class switch recombination. *PLoS One* 7:e51675. doi: 10.1371/journal.pone.0051675

Nguyen, H. Q., Chattoraj, S., Castillo, D., Nguyen, S. C., Nir, G., Lioutas, A., et al. (2020). 3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing. *Nat. Methods* 17, 822–832. doi: 10.1038/s41592-020-0890-0

Ni, Y., Cao, B., Ma, T., Niu, G., Huo, Y., Huang, J., et al. (2017). Super-resolution imaging of a 2.5 kb non-repetitive DNA *in situ* in the nuclear genome using molecular beacon probes. *Elife* 6:e21660. doi: 10.7554/eLife.21660

Nir, G., Farabella, I., Pérez Estrada, C., Ebeling, C. G., Beliveau, B. J., Sasaki, H. M., et al. (2018). Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLoS Genet.* 14:e1007872. doi: 10.1371/journal.pgen.1007872

Niu, L., Shen, W., Shi, Z., Tan, Y., He, N., Wan, J., et al. (2021). Three-dimensional folding dynamics of the Xenopus tropicalis genome. *Nat. Genet.* 53, 1075–1087. doi: 10.1038/s41588-021-00878-z

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485:381ё385. doi: 10.1038/nature11049

Nozaki, T., Imai, R., Tanbo, M., Nagashima, R., Tamura, S., Tani, T., et al. (2017). Dynamic organization of chromatin domains revealed by super-resolution live-cell Imaging. *Mol. Cell.* 67, doi: 10.1016/j.molcel.2017.06.018 282-293.e7.

Ou, H. D., Phan, S., Deerinck, T. J., Thor, A., Ellisman, M. H., and O'Shea, C. C. (2017). ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* 357:eaag0025. doi: 10.1126/science.aag0025

Pal, K., Forcato, M., and Ferrari, F. (2019). Hi-C analysis: from data generation to integration. *Biophys. Rev.* 11, 67–78. doi: 10.1007/s12551-018-0489-1

Pardue, M. L., and Gall, J. G. (1969). Molecular hybridization of radioactive DNA to the DNA of cytological preparations. *Proc. Natl. Acad. Sci. U.S.A* 64, 600–604. doi: 10.1073/pnas.64.2.600

Paweletz, N. (2001). Walther Flemming: pioneer of mitosis research. *Nat. Rev. Mol. Cell Biol.* 2, 72–75. doi: 10.1038/35048077

Pecinka, A., Schubert, V., Meister, A., Kreth, G., Klatte, M., Lysak, M. A., et al. (2004). Chromosome territory arrangement and homologous pairing in nuclei of Arabidopsis thaliana are predominantly random except for NOR-bearing chromosomes. *Chromosoma* 113, 258–269. doi: 10.1007/s00412-004-0316-2

Pinkel, D., Straume, T., and Gray, J. W. (1986). Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc. Natl. Acad. Sci. U.S.A* 83, 2934–2938. doi: 10.1073/pnas.83.9.2934

Pueschel, R., Coraggio, F., and Meister, P. (2016). From single genes to entire genomes: the search for a function of nuclear organization. *Development* 143, 910–923. doi: 10.1242/dev.129007

Quinodoz, S. A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J. M., Detmar, E., et al. (2018). Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell* 174, 744–757.e24. doi: 10.1016/j.cell.2018.05.024

Ouyang, W., Xiong, D., Li, G., and Li, X. (2020). Unraveling the 3D Genome architecture in plants: present and future. *Mol. Plant.* 13, 1676–1693. doi: 10.1016/j.molp.2020.10.002

Rada-Iglesias, A., Grosveld, F. G., and Papantonis, A. (2018). Forces driving the three-dimensional folding of eukaryotic genomes. *Mol. Syst. Biol.* 14:e8214. doi: 10.15252/msb.20188214

Ramani, V., Deng, X., Qiu, R., Gunderson, K. L., Steemers, F. J., Disteche, C. M., et al. (2017). Massively multiplex single-cell Hi-C. *Nat. Methods* 14, 263–266. doi: 10.1038/nmeth.4155

Ramírez, F., Bhardwaj, V., Arrigoni, L., Lam, K. C., Grüning, B. A., Villaveces, J., et al. (2018). High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* 9:189. doi: 10.1038/s41467-017-02525-w

Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680. doi: 10.1016/j.cell.2014.11.021

Renz, M. (2013). Fluorescence microscopy-a historical and technical perspective. *Cytometry* 83, 767–779. doi: 10.1002/cyto.a.22295

Ried, T., Baldini, A., Rand, T. C., and Ward, D. C. (1992). Simultaneous visualization of seven different DNA probes by in situ hybridization using combinatorial fluorescence and digital imaging microscopy. *Proc. Natl. Acad. Sci. U.S.A* 89, 1388–1392. doi: 10.1073/pnas.89.4.1388

Rodriguez-Granados, N. Y., Ramirez-Prado, J. S., Veluchamy, A., Latrasse, D., Raynaud, C., Crespi, M., et al. (2016). Put your 3D glasses on: plant chromatin is on show. *J. Exp. Bot.* 67, 3205–3221. doi: 10.1093/jxb/erw168

Rogan, P. K., Cazcarro, P. M., and Knoll, J. H. (2001). Sequence-based design of single-copy genomic DNA probes for fluorescence in situ hybridization. *Genome Res.* 11, 1086–1094. doi: 10.1101/gr.171701

Ronneberger, O., Baddeley, D., Scheipl, F., Verveer, P. J., Burkhardt, H., Cremer, C., et al. (2008). Spatial quantitative analysis of fluorescently labeled nuclear structures: problems, methods, pitfalls. *Chromosome Res.* 16, 523–562. doi: 10.1007/s10577-008-1236-4

Rosin, L. F., Nguyen, S. C., and Joyce, E. F. (2018). Condensin II drives large-scale folding and spatial partitioning of interphase chromosomes in Drosophila nuclei. *PLoS Genet.* 14:e1007393. doi: 10.1371/journal.pgen.1007393

Rouillard, J. M., Zuker, M., and Gulari, E. (2003). OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res.* 31, 3057–3062. doi: 10.1093/nar/gkg426

Rouquette, J., Genoud, C., Vazquez-Nin, G. H., Kraus, B., Cremer, T., and Fakan, S. (2009). Revealing the high-resolution three-dimensional network of chromatin and interchromatin space: a novel electron-microscopic approach to reconstructing nuclear architecture. *Chromosome Res.* 17, 801–810. doi: 10.1007/s10577-009-9070-x

Rouquette, J., Cremer, C., Cremer, T., and Fakan, S. (2010). Functional nuclear architecture studied by microscopy: present and future. *Int. Rev. Cell Mol. Biol.* 282, 1–90. doi: 10.1016/S1937-6448(10)82001-5

Rowley, M. J., and Corces, V. G. (2016). The three-dimensional genome: principles and roles of long-distance interactions. *Curr. Opin. Cell Biol.* 40, 8–14. doi: 10.1016/j.ceb.2016.01.009

Rowley, M. J., Nichols, M. H., Lyu, X., Ando-Kuri, M., Rivera, I. S. M., Hermetz, K., et al. (2017). Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol. Cell.* 67, 837–852.e7. doi: 10.1016/j.molcel.2017.07.022

Sati, S., and Cavalli, G. (2017). Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma* 126, 33–44. doi: 10.1007/s00412-016-0593-6

Sealey, P. G., Whittaker, P. A., and Southern, E. M. (1985). Removal of repeated sequences from hybridization probes. *Nucl. Acids Res.* 13, 1905–1922. doi: 10.1093/nar/13.6.1905

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., et al. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* 148, 458–472. doi: 10.1016/j.cell.2012.01.010

Schermelleh, L., Ferrand, A., Huser, T., Eggeling, C., Sauer, M., Biehlmaier, O., et al. (2019). Super-resolution microscopy demystified. *Nat. Cell Biol.* 21, 72–84. doi: 10.1038/s41556-018-0251-8

Schmitt, E., Schwarz-Finsterle, J., Stein, S., Boxler, C., Müller, P., Mokhir, A., et al. (2010). Combinatorial Oligo FISH: directed labeling of specific genome domains in differentially fixed cell material and live cells. *Methods Mol. Biol.* 659, 185–202. doi: 10.1007/978-1-60761-789-1_13

Schriml, L. M., Padilla-Nash, H. M., Coleman, A., Moen, P., Nash, W. G., Menninger, J., et al. (1999). Tyramide signal amplification (TSA)-FISH applied to mapping PCR-labeled probes less than 1 kb in size. *Biotechniques* 27, 608–613. doi: 10.2144/99273pf01

Schubert, I., and Shaw, P. (2011). Organization and dynamics of plant interphase chromosomes. *Trends Plant Sci.* 16, 273–281. doi: 10.1016/j.tplants.2011.02.002

Schubert, V., Rudnik, R., and Schubert, I. (2014). Chromatin associations in Arabidopsis interphase nuclei. *Front. Genet.* 5:389. doi: 10.3389/fgene.2014.00389

Shachar, S., Voss, T. C., Pegoraro, G., Sciascia, N., and Misteli, T. (2015). Identification of gene positioning factors using high-throughput imaging mapping. *Cell* 162, 911–923. doi: 10.1016/j.cell.2015.07.035

Shao, S., Chang, L., Hou, Y., and Sun, Y. (2017). Illuminating the structure and dynamics of chromatin by fluorescence labeling. *Front. Biol.* 12:241–257. doi: 10.1007/s11515-017-1454-2

Sharma, R., and Meister, P. (2020). Generation of Inexpensive, Highly labeled probes for fluorescence *in situ* hybridization (FISH). *STAR Protoc.* 1:100006. doi: 10.1016/j.xpro.2019.100006

Shi, G., and Thirumalai, D. (2019). Conformational heterogeneity in human interphase chromosome organization reconciles the FISH and Hi-C paradox. *Nat. Commun.* 10:3894. doi: 10.1038/s41467-019-11897-0

Shim, S. H. (2021). Super-resolution microscopy of genome organization. *Genes Genom.* 43, 281–287. doi: 10.1007/s13258-021-01044-9

Solovei, I., Cavallo, A., Schermelleh, L., Jaunin, F., Scasselati, C., Cmarko, D., et al. (2002). Spatial preservation of nuclear chromatin architecture during three-dimensional fluorescence in situ hybridization (3D-FISH). *Exp. Cell Res.* 276, 10–23. doi: 10.1006/excr.2002.5513

Solovei, I., Kreysing, M., Lanctôt, C., Kösem, S., Peichl, L., Cremer, T., et al. (2009). Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell* 137, 356–368. doi: 10.1016/j.cell.2009.01.052

Stadler, M. R., Haines, J. E., and Eisen, M. B. (2017). Convergence of topological domain boundaries, insulators, and polytene interbands revealed by high-resolution mapping of chromatin contacts in the early Drosophila melanogaster embryo. *Elife* 6:e29550. doi: 10.7554/eLife.29550

Stevens, T., Lando, D., Basu, S., Atkinson, L. P., Cao, Y., Lee, S. F., et al. (2017). 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* 544, 59–64. doi: 10.1038/nature21429

Su, J. H., Zheng, P., Kinrot, S. S., Bintu, B., and Zhuang, X. (2020). Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell* 182, 1641–1659. doi: 10.1016/j.cell.2020.07.032

Swennenhuis, J. F., Foulk, B., Coumans, F. A., and Terstappen, L. W. (2012). Construction of repeat-free fluorescence in situ hybridization probes. *Nucleic Acids Res.* 40:e20. doi: 10.1093/nar/gkr1123

Szabo, Q., Jost, D., Chang, J. M., Cattoni, D. I., Papadopoulos, G. L., Bonev, B., et al. (2018). TADs are 3D structural units of higher-order chromosome organization in Drosophila. *Sci. Adv.* 4:eaar8082. doi: 10.1126/sciadv.aar8082

Szabo, Q., Bantignies, F., and Cavalli, G. (2019). Principles of genome folding into topologically associating domains. *Sci. Adv.* 5:eaaw1668. doi: 10.1126/sciadv.aaw1668

Szabo, Q., Donjon, A., Jerković, I., Papadopoulos, G. L., Cheutin, T., Bonev, B., et al. (2020). Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nat. Genet.* 52, 1151–1157. doi: 10.1038/s41588-020-00716-8

Szabo, Q., Cavalli, G., and Bantignies, F. (2021). Higher-order chromatin organization using 3D DNA fluorescent *in situ* hybridization. *Methods Mol. Biol.* 2157, 221–237. doi: 10.1007/978-1-0716-0664-3_13

Szydlowski, N. A., Go, J. S., and Hu, Y. S. (2019). Chromatin imaging and new technologies for imaging the nucleome. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 11:e1442. doi: 10.1002/wsbm.1442

Takei, Y., Yun, J., Zheng, S., Ollikainen, N., Pierson, N., White, J., et al. (2021). Integrated spatial genomics reveals global architecture of single nuclei. *Nature* 590, 344–350. doi: 10.1038/s41586-020-03126-2

Tasara, T., Angerer, B., Damond, M., Winter, H., Dörhöfer, S., Hübscher, U., et al. (2003). Incorporation of reporter molecule-labeled nucleotides by DNA polymerases. II. High-density labeling of natural DNA. *Nucleic Acids Res.* 31, 2636–2646. doi: 10.1093/nar/gkg371

Trzaskoma, P., Ruszczycki, B., Lee, B., Pels, K. K., Krawczyk, K., Bokota, G., et al. (2020). Ultrastructural visualization of 3D chromatin folding using volume electron microscopy and DNA in situ hybridization. *Nat. Commun.* 11:2120. doi: 10.1038/s41467-020-15987-2

Tweedie, S., Ashburner, M., Falls, K., Leyland, P., McQuilton, P., Marygold, S., et al. (2009). FlyBase: enhancing Drosophila gene ontology annotations. *Nucleic Acids Res.* 37, D555–D559. doi: 10.1093/nar/gkn788

Ulianov, S. V., Khrameeva, E. E., Gavrilov, A. A., Flyamer, I. M., Kos, P., Mikhaleva, E. A., et al. (2016). Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res.* 26, 70–84. doi: 10.1101/gr.196006.115

Ulianov, S. V., Tachibana-Konwalski, K., and Razin, S. V. (2017). Single-cell Hi-C bridges microscopy and genome-wide sequencing approaches to study 3D chromatin organization. *Bioessays* 39:10. doi: 10.1002/bies.201700104

Ulianov, S. V., Zakharova, V. V., Galitsyna, A. A., Kos, P. I., Polovnikov, K. E., Flyamer, I. M., et al. (2021). Order and stochasticity in the folding of individual Drosophila genomes. *Nat. Commun.* 12:41. doi: 10.1038/s41467-020-20292-z

van Holde, K., and Zlatanova, J. (1995). Chromatin higher order structure: chasing a mirage? *J. Biol. Chem.* 270, 8373–8376. doi: 10.1074/jbc.270.15.8373

Vatolina, T. Y., Boldyreva, L. V., Demakova, O. V., Demakov, S. A., Kokoza, E. B., Semeshin, V. F., et al. (2011). Identical functional organization of nonpolytene and polytene chromosomes in Drosophila melanogaster. *PLoS One.* 6:e25960. doi: 10.1371/journal.pone.0025960

Wachsmuth, M., Knoch, T. A., and Rippe, K. (2016). Dynamic properties of independent chromatin domains measured by correlation spectroscopy in living cells. *Epigenet. Chrom.* 9:57. doi: 10.1186/s13072-016-0093-1

Wang, Q., Sun, Q., Czajkowsky, D. M., and Shao, Z. (2018). Sub-kb Hi-C in *D. melanogaster* reveals conserved characteristics of TADs between insect and mammalian cells. *Nat. Commun.* 9:188. doi: 10.1038/s41467-017-02526-9

Wang, S., Su, J. H., Beliveau, B. J., Bintu, B., Moffitt, J. R., Wu, C. T., et al. (2016). Spatial organization of chromatin domains and compartments in single chromosomes. *Science* 353, 598–602. doi: 10.1126/science.aaf8084

Weiland, Y., Lemmer, P., and Cremer, C. (2011). Combining FISH with localisation microscopy: Super-resolution imaging of nuclear genome nanostructures. *Chromosome Res.* 19, 5–23. doi: 10.1007/s10577-010-9171-6

Weinreb, C., and Raphael, B. J. (2016). Identification of hierarchical chromatin domains. *Bioinformatics* 32, 1601–1609. doi: 10.1093/bioinformatics/btv485

Wiegant, J., Ried, T., Nederlof, P. M., van der Ploeg, M., Tanke, H. J., and Raap, A. K. (1991). In situ hybridization with fluoresceinated DNA. *Nucleic Acids Res.* 19, 3237–3241. doi: 10.1093/nar/19.12.3237

Williamson, I., Berlivet, S., Eskeland, R., Boyle, S., Illingworth, R. S., Paquette, D., et al. (2014). Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence *in situ* hybridization. *Genes Dev.* 28, 2778–2791. doi: 10.1101/gad.251694.114

Woodcock, C. L., and Horowitz, R. A. (1997). Electron microscopy of chromatin. *Methods* 12, 84–95. doi: 10.1006/meth.1997.0450

Woodcock, C. L., and Ghosh, R. P. (2010). Chromatin higher-order structure and dynamics. *Cold Spring Harb. Perspect. Biol.* 2:a000596. doi: 10.1101/cshperspect.a000596

Wu, X., Mao, S., Ying, Y., Krueger, C. J., and Chen, A. K. (2019). Progress and challenges for live-cell imaging of genomic loci using CRISPR-based platforms. *Genomics Proteomics Bioinform.* 17, 119–128. doi: 10.1016/j.gpb.2018.10.001

Xiao, L., Liao, R., and Guo, J. (2020). Highly Multiplexed Single-Cell In Situ RNA and DNA Analysis by Consecutive Hybridization. *Molecules* 25:4900. doi: 10.3390/molecules25214900

Xie, L., and Liu, Z. (2021). Single-cell imaging of genome organization and dynamics. *Mol. Syst. Biol.* 17:e9653. doi: 10.15252/msb.20209653

Xu, J., and Liu, Y. (2019). A guide to visualizing the spatial epigenome with super-resolution microscopy. *FEBS J.* 286, 3095–3109. doi: 10.1111/febs.14938

Yamada, N. A., Rector, L. S., Tsang, P., Carr, E., Scheffer, A., Sederberg, M. C., et al. (2011). Visualization of fine-scale genomic structure by oligonucleotide-based high-resolution FISH. *Cytogenet Genome Res.* 132, 248–254. doi: 10.1159/000322717

Yokota, H., Singer, M. J., van den Engh, G. J., and Trask, B. J. (1997). Regional differences in the compaction of chromatin in human G0/G1 interphase nuclei. *Chromosome Res.* 5, 157–166. doi: 10.1023/a:1018438729203

Yu, H., Chao, J., Patek, D., Mujumdar, R., Mujumdar, S., and Waggoner, A. S. (1994). Cyanine dye dUTP analogs for enzymatic labeling of DNA probes. *Nucleic Acids Res.* 22, 3226–3232. doi: 10.1093/nar/22.15.3226

Zhu, Z., Chao, J., Yu, H., and Waggoner, A. S. (1994). Directly labeled DNA probes using fluorescent nucleotides with different length linkers. *Nucleic Acids Res.* 22, 3418–3422. doi: 10.1093/nar/22.16.3418

Zlotina, A., Maslova, A., Pavlova, O., Kosyakova, N., Al-Rikabi, A., Liehr, T., et al. (2020). New insights Into chromomere organization provided by lampbrush chromosome microdissection and high-throughput sequencing. *Front. Genet.* 11:57. doi: 10.3389/fgene.2020.00057

Zufferey, M., Tavernari, D., Oricchio, E., and Ciriello, G. (2018). Comparison of computational methods for the identification of topologically associating domains. *Genome Biol.* 19:217. doi: 10.1186/s13059-018-1596-9

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

A Shift in Paradigms: Spatial Genomics Approaches to Reveal Single-Cell Principles of Genome Organization

Andres M. Cardozo Gizzi *

Centro de Investigación en Medicina Traslacional Severo Amuchastegui (CIMETSA), Instituto Universitario de Ciencias Biomédicas de Córdoba (IUCBC), CONICET, Córdoba, Argentina

The genome tridimensional (3D) organization and its role towards the regulation of key cell processes such as transcription is currently a main question in biology. Interphase chromosomes are spatially segregated into "territories," epigenetically-defined large domains of chromatin that interact to form "compartments" with common transcriptional status, and insulator-flanked domains called "topologically associating domains" (TADs). Moreover, chromatin organizes around nuclear structures such as lamina, speckles, or the nucleolus to acquire a higher-order genome organization. Due to recent technological advances, the different hierarchies are being solved. Particularly, advances in microscopy technologies are shedding light on the genome structure at multiple levels. Intriguingly, more and more reports point to high variability and stochasticity at the single-cell level. However, the functional consequences of such variability in genome conformation are still unsolved. Here, I will discuss the implication of the cell-to-cell heterogeneity at the different scales in the context of newly developed imaging approaches, particularly multiplexed Fluorescence in situ hybridization methods that enabled "chromatin tracing." Extensions of these methods are now combining spatial information of dozens to thousands of genomic *loci* with the localization of nuclear features such as the nucleolus, nuclear speckles, or even histone modifications, creating the fast-moving field of "spatial genomics." As our view of genome organization shifts the focus from ensemble to single-cell, new insights to fundamental questions begin to emerge.

Keywords: chromatin 3D architecture, chromosome conformation, topologically associated domain (TAD), fluorescence *in situ* cell hybridization (FISH), oligopaint, transcriptional regulation, genome organization, stochasticity

INTRODUCTION

In eukaryotes, DNA is arranged in a three-dimensional (3D) packaging within the nucleus. The genome hierarchical 3D organization conforms a key regulatory layer of gene expression and cell fate control (Bonev and Cavalli, 2016). Individual chromosomes are spatially partitioned into discrete "chromosome territories" (Cremer and Cremer, 2001; Bolzer et al., 2005; Cremer et al., 2006). Down from the chromosomal scale, the genome is partitioned into two types of structural units. On the one hand, active (A) and inactive (B) compartments are genomic regions spanning several mega-base pairs (Mb) which tend to engage in homotypic (A-A or B-B) rather than heterotypic contacts. On the

other hand, topologically associating domains (TADs) are defined as regions at the sub-Mb scale displaying higher intra-domain interactions and relatively insulated from neighboring domains.

The segregation of active and repressed chromatin was observed for the first time by Emil Heitz, who in 1928 suggested the terms "heterochromatin" and "euchromatin" (Passarge, 1979). A great deal about chromatin spatial organization has been learned thanks to the development of biochemical methods called chromatin conformation capture (3C) and 3C derivatives (Goel and Hansen, 2021; Jerkovic and Cavalli, 2021). 3C-based techniques rely on DNA crosslinking to fix the interacting sequences and nuclease fragmentation to retrieve the contact frequency of pairs of genomic positions. In particular, genome-wide maps of chromatin interaction have been obtained by sequencing-based high-throughput chromosome conformation capture techniques (Hi-C). Through initial Hi-C maps, it was found that domains sharing biochemical properties such as epigenetic marks and transcriptional status tend to interact with domains of the same type, to form A/B compartments (size ~ 1–3 Mb), which resemble euchromatin and heterochromatin, respectively (Lieberman-Aiden et al., 2009).

The other genome "structural unit," TADs, were discovered due to an increased genomic resolution of 3C-based methods (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012), with an average size between 185–900 kb in mammals (Dixon et al., 2012; Rao et al., 2014; Bonev et al., 2017) and 100–150 kb in *Drosophila* (Ulianov et al., 2016; Wang et al., 2018). TADs organization is, for the most part, stable between cell types or through differentiation (i.e., most TAD borders are invariant) (Dixon et al., 2015; Dixon et al., 2016). Furthermore, TADs borders coincide to a high degree with replication domain boundaries (Pope et al., 2014; Dixon et al., 2016; Ulianov et al., 2016). Even more importantly, *cis*-regulatory elements that direct transcription are mostly restricted to interactions within a TAD (Lupiáñez et al., 2015; Dixon et al., 2016). All in all, this points to a role of TADs as conserved genome "units of regulation" or even thought as physical globular domains present in most cells of a population. As we will see from single-cell techniques, the latter is an oversimplification.

Finally, the spatial compartmentalization of nuclear events is evidenced by the spatially defined localization of processes. The existence of diverse nuclear bodies, membraneless compartments with specific tasks, is a key aspect of the nuclear organization (Misteli, 2005; Mao et al., 2011). For example, nuclear speckles are subnuclear bodies that contain mRNA processing and splicing factors (Galganski et al., 2017). It has been shown that highly transcribed Pol II regions organize around nuclear speckles, whereas inactive genomic regions are frequently associated with the nuclear periphery (Guelen et al., 2008) or the nucleolus (Quinodoz et al., 2018). Inter-chromosomal contacts are organized around nuclear bodies to create a higher-order genome organization. Additionally, another principle of non-random nuclear architecture is the radial organization model where euchromatic regions (A compartment) organize centrally with respect to nuclear lamina whereas heterochromatin (B compartment) is associated with the nuclear periphery and perinucleolar regions (Buchwalter et al., 2019; Crosetto and Bienko, 2020). More importantly, the non-random organization of the genome has meaningful effects on its function and activity. As technology develops, both imaging- and sequencing-based, there is a notorious shift in paradigm: ensemble measurements are just simply not enough to understand the structure-function relationship. Here I will discuss the microscopy improvements that lead to new insights into the stochasticity in genome organization and its influence on the mechanisms involved.

INTRODUCING "SPATIAL GENOMICS"

Microscopy methods enable the visualization of genomic features in single cells (Xie and Liu, 2021). Fluorescence *in situ* hybridization (FISH) detects the physical position of targeted sequences by the annealing of labeled DNA or RNA probes. As genome-wide methods started to be widely used across many laboratories, single-cell 3D-DNA FISH was used as an orthogonal method to validate observations (Nora et al., 2012). Therefore, selected pairs of *loci* were used to measure physical distances and compare them with 3C contact frequencies (Giorgetti and Heard, 2016).

Two major FISH limitations can be identified when it comes to extending its throughput. The first is the probe design and production. Traditionally, FISH probes are derived from molecular cloning to vectors such as bacterial artificial chromosomes (BACs) (Roohi et al., 2008) and PCR-based methods like HD-FISH (Bienko et al., 2013). These methods are laborious and time-consuming, especially to produce multiple probes. Due to advances in high-throughput parallel chemical synthesis, it is now possible to construct FISH probes from oligonucleotides (oligos), termed Oligopaints (Beliveau et al., 2012; Beliveau et al., 2015). Oligo-based probes are selected bioinformatically and allow for great flexibility in terms of experimental design, targeting from a few kilobases (kb) to Mbs (Beliveau et al., 2018).

The other limitation is the color channels available to imaging, restricting FISH to 2–3 *loci* per experiment. An initial effort using a sequential color code trace a whole chromosomal arm (Lowenstein et al., 2004) although it has remained challenging to unambiguously identify multiple *loci*. Xiaowei Zhuang's lab developed the concept of sequential imaging of target *loci* combining the flexibility of Oligopaints with microfluidics in a regular widefield fluorescence microscope to accomplish the multiplexed detection of FISH probes (Wang et al., 2016).

The idea is to use a set of oligos (hereafter "barcode"), targeting a specific locus, that shares the same overhang region that is then recognized by a fluorescently labeled secondary oligo. After hybridizing primary probes to all target regions, barcode-specific secondary probes are injected to then imaged across multiple fields of view, photobleach and start a new hybridization cycle (**Figure 1A**). In each cycle, the barcodes appear as fluorescent spots whose centroid position is determined with nanometric precision (Boettiger and Murphy, 2020). Therefore,
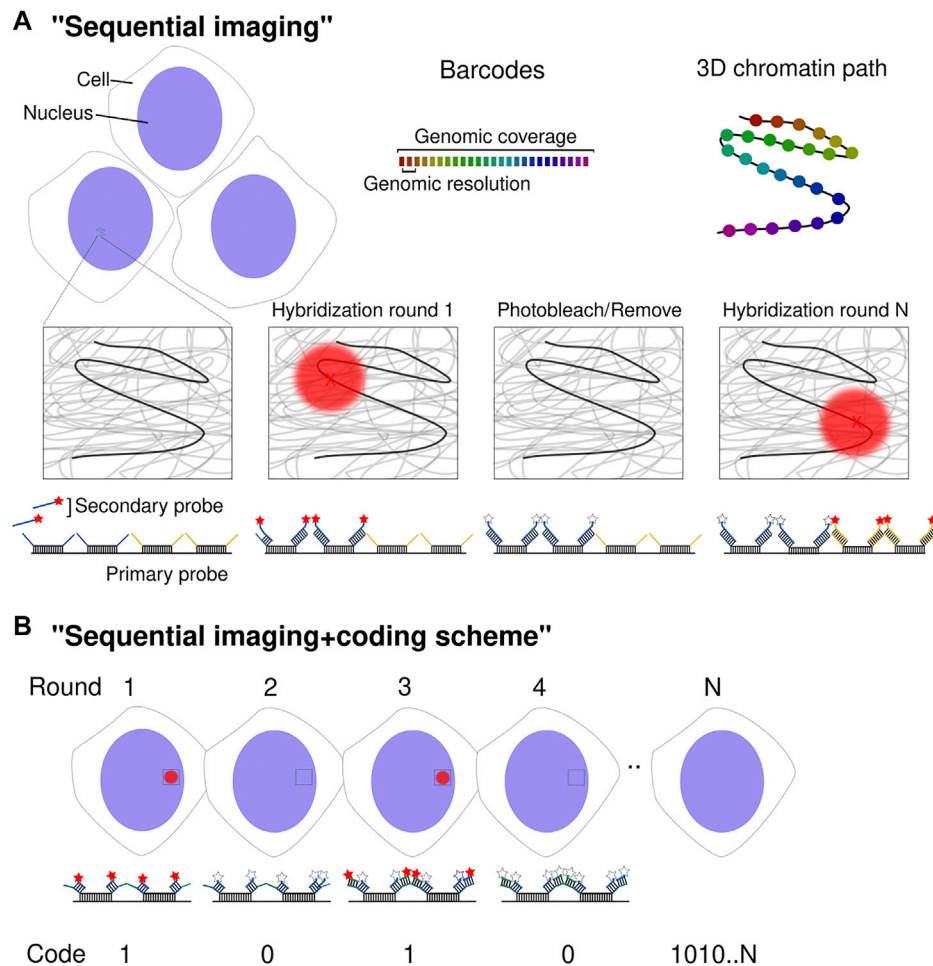
FIGURE 1 | Spatial genomics approaches. **(A)** Schematic diagram of multiplexed DNA barcode detection. DNA loci are detected sequentially through secondary readout probes complementary to barcode-specific overhang sequences. The centroid of diffraction-limited spots (red X) is determined with nanometric precision. After each hybridization and imaging round, fluorophores are either removed or photobleached before starting a new cycle. Upon completion of N rounds, the chromatin path is determined in individual chromosomes across thousands of cells. **(B)** Schematic diagram of the implementation of a coding scheme using N sequential cycles. Although the procedure to determine the chromatin path is the same as in **(A)**, a coding scheme is implemented. Every barcode is detected by more than one readout probe (two in this case) by the use of multiple overhang sequences per barcode. This leads to the detection of the same barcode in several imaging rounds. Detection is read as a "1" whereas no detection as "0". Post-signal processing allows decoding the position of 2^N barcodes.

the method enables a direct tracing of the chromatin path with a genomic coverage and resolution according to the design of the Oligopaint probes (i.e., size of the barcoded regions and the distance between barcodes).

The initial approach managed to image ~30 genomic *loci* covering a whole human chromosome with a genomic resolution around the Mb and revealing that at this scale there is a strong correlation between mean spatial distance with Hi-C contact frequency (Wang et al., 2016). Following this study, three papers appeared within a 6-months window, further developing the multiplexed methods using "chromatin tracing" (Bintu et al., 2018), "Hi-M" (Cardozo Gizzi et al., 2019) and "optical reconstruction of chromatin architecture" (ORCA) (Mateo et al., 2019) and achieved a resolution of 2–30 kb at the sub-TAD scale to cover ~20–70 regions. At this scale, it was found that TADs, discovered by 3C-based methods, indeed

appeared when averaging the population chromatin spatial conformation (see below *Stochasticity in Genome Organization*). Furthermore, it was then possible to establish, in the same cells, the transcriptional status by imaging RNA species (Cardozo Gizzi et al., 2019; Mateo et al., 2019). In *Drosophila*, it was shown that active transcription is associated with the unfolding of the gene-containing TAD at the ensemble level (Cardozo Gizzi et al., 2019). In eukaryotic cells, transcription is controlled by *cis*-regulatory elements (CREs) such as enhancers, silencers and promoters. By using contiguous barcodes to achieve a resolution of ~2 kb, it was possible to study CRE hubs that regulate gene expression. It was found that enhancer-promoter (E-P) distance was only a weak predictor of transcription (Mateo et al., 2019) and that distal CRE hubs are formed before gene activation (or even TADs) and may reinforce transcriptional repression (Espinola et al., 2021). Additionally,

FIGURE 2 | Chromatin organization is variable between cells. **(A)** Ensemble-averaged median spatial distance map, color-code from red to blue according to the scale bar. Three eTADs are clearly visible. **(B)** Single-cell spatial distance maps indicate the heterogeneity in chromatin 3D architecture. **(C)** Chromatin path representation of the single-cell distance maps, color-coded according to the genomic coordinate scale bar. Representation based on spatial genomic approaches.

and 98 "0" bits to image 1,041 genomic *loci* employing 50 hybridization rounds in two channels. In this study, a particular genomic locus would be decoded after being detected (*on* or "1") in a particular spatial localization in two out of 100 different hybridization cycles. In Takei et al. (2021a) 2,460 genomic loci were imaged using 80 hybridization rounds in two channels.

A different spatial genomics approach is the combination of microscopy and sequencing by adapting and improving fluorescence RNA *in situ* sequencing (IGS) or FISSEQ technology (Lee et al., 2014). Recent developments of IGS have permitted both targeted (Oligo-FISSEQ) (Nguyen et al., 2020) or untargeted approaches (Payne et al., 2021). Oligo-FISSEQ uses barcoded Oligopaints targeting multiple genomic regions that are sequenced *in situ* whereas untargeted IGS uses Tn5 transposase to randomly incorporate DNA sequencing adaptors into fixed DNA, achieving a resolution of ~1 Mb genome-wide. Finally, combining chromating tracing with multimodal RNA- and immuno-labeling (Liu et al., 2020; Su et al., 2020; Payne et al., 2021; Takei et al., 2021a) enables the profiling of genome conformation, nuclear bodies, gene expression and epigenetic status in the same cell.

In the last 3 years, this revolution kickstarted a new field. These very recent developments put us within range of genome-wide spatial maps of chromatin organization, complementing the best of genomics and microscopy fields. More and more labs are developing and implementing "spatial genomics" approaches even if at the present the methodology employs custom-made setups and requires an in-house knowledge of automated image analysis. From these approaches, the different contributions of heterogeneity to chromosome architecture at different scales are being sorted out.

STOCHASTICITY IN GENOME ORGANIZATION

Genome organization has a large degree of variability at the single-cell level (Finn and Misteli, 2019) and the 3D segregation of chromosomes shows a clear variability between cells. Accordingly, the relative position of a particular chromosome to each other is not "predefined" yet the "chromosome territories" are physical structures present in all cells within a population. This is not the case for A/B compartments or TADs that arise from averaging multiple cell conformations in mammalian cells. In other words, they are statistical features of genome organization not necessarily present from cell to cell. Here I will discuss the evidence supporting this claim, mainly obtained from spatial genomics techniques unless stated otherwise.

The segregation of active and inactive chromatin by the preferred contacts between chromatin of the same class is observed in single cells, that display their chromosomes in a "polarized fashion" in interphase human fibroblasts (Wang et al., 2016; Nir et al., 2018) and *C. elegans* embryos (Sawh et al., 2020). This indicates that compartments, or regions of active/inactive chromatin, are localized side-by-side with various degrees of intermixing. Consistent with genome-wide studies, chromatin

the simultaneous detection of RNA can also be used as a proxy to perform cell-type specific studies (Mateo et al., 2019; Liu et al., 2020; Espinola et al., 2021).

The "coding scheme" concept was later introduced to deliver throughput orders of magnitude higher. It was adapted from multiplexed error-robust FISH (MERFISH) (Chen et al., 2015) or sequential FISH (seqFISH) (Lubeck et al., 2014; Shah et al., 2018; Eng et al., 2019), initially developed for RNA *in situ* detection. Xiaowei Zhuang's DNA-MERFISH and Long Cai's seqFISH+ were developed in parallel and consist of embedding a particular barcode with more than one class of readout sequence, constituting a binary code. In other words, two to five different overhang sequences are added in each barcode, that will be then detected with multiple readout fluorescent oligos (**Figure 1B**). The "1" or "0" value of each bit corresponds to the presence or not of a particular barcode in a hybridization round. This allows for 2^N genomic positions to be imaged in N rounds of hybridization. The vast majority of possible encoded barcodes are not used to implement an error detection and correction scheme. Su et al. (2020) employed 100-binary barcodes with two "1" bits

tracing experiments found a spatial correlation between nucleoli and nuclear lamina with B-compartment regions (Liu et al., 2020; Su et al., 2020) or between speckle with A-compartment regions (Liu et al., 2020). Furthermore, the degree of segregation between compartments showed a gradual establishment during the cell cycle, increasing from G1 to G2/S phase (Su et al., 2020) as previously seen by Hi-C (Abramo et al., 2019). However, individual chromosomes display a high level of variation, from the extreme complete segregation of A- and B-clusters to a highly intermingling configuration (Liu et al., 2020; Su et al., 2020).

Microscopy reports determined a low contact probability (1–10%) of long-range associations between any pair of loci and a modest two-fold increase within TADs (Cattoni et al., 2017; Finn et al., 2019). The single-cell contact maps frequently exhibit TAD-like structures, as seen in multiple chromatin tracing studies (Bintu et al., 2018; Mateo et al., 2019; Su et al., 2020; Payne et al., 2021; Takei et al., 2021b). These are local physical domains of enhanced contact that are well separated from one another. The physical properties of domains, such as size or degree of insulation, displayed a large heterogeneity (Boettiger et al., 2016; Nir et al., 2018; Szabo et al., 2018; Luppino et al., 2020; Szabo et al., 2020). This is consistent with the high variability in TAD formation observed in single-cell sequencing-based biochemical methods (reviewed in Ulianov et al., 2017; Galitsyna and Gelfand, 2021), such as Hi-C (Flyamer et al., 2017; Nagano et al., 2017; Stevens et al., 2017; Tan et al., 2018), ChIA-Drop (Zheng et al., 2019) or scSPRITE (Arrastia et al., 2021). Consistently, the boundaries of such domains do not necessarily correspond to ensemble-averaged TADs (eTADs; Bohrer and Larson, 2021) (**Figure 2**).

In mammals, Hi-C-defined eTADs are frequently flanked by pairs of CTCF binding sites in convergent orientation and serve as anchors for chromatin loops (Nora et al., 2012; Rao et al., 2014). TAD-like domain boundaries were preferentially positioned at CTCF and cohesin binding sites, belonging to eTADs boundaries, peaking at ~15% probability. However, all other genomic loci within a TAD shared a boundary probability of ~5–7% (Bintu et al., 2018; Su et al., 2020). In contrast, *Drosophila* TADs, whose borders are not enriched in CTCF, are much more stable from cell-to-cell, observed both by microscopy (Szabo et al., 2018) and single-cell Hi-C (Ulianov et al., 2021); although the reasons are under investigation (Ulianov and Razin, 2021).

A very graphical example of TAD architecture at the single-cell level is this is the organization of the inactivated X chromosome, used as a model for chromosome organization (Galupa and Heard, 2018). In mammalian females, the two copies of X chromosomes display a very different transcriptional and epigenetic landscape. At the ensemble level, the inactivated X chromosome (Xi) displays only two mega domains with the boundary located at macrosatellite DXZ4. Strikingly, both the active X chromosome and Xi show TAD-like domains at the single-cell level (Cheng et al., 2021; Takei et al., 2021b).

The role of TADs in transcription regulation is still an open question, but evidence supports both a role on facilitating CREs communication within the TAD and on blocking enhancer-promoter contacts between TADs (Furlong and Levine, 2018;

Cavalheiro et al., 2021). However, the stochastic nature of TADs (and compartments) questions the real influence of TADs on transcription modulation (Bohrer and Larson, 2021). The timescales involved in chromosome organization and transcription is a dimension that needs to be considered, and that is not being addressed by FISH or sequencing-based methods in fixed cells (Nollmann et al., 2021). The live-cell tracking of loci gives information on the dynamic nature of regulatory DNA contacts such as E-P interactions (Brandão et al., 2021), and thus can bring understanding into the role of 3D genome organization in CREs regulation (Schoenfelder and Fraser, 2019).

Cell-to-cell variability within a phenotypically indistinguishable population has also been found in the transcriptome and epigenome (DNA methylation and histone modification profile) (Golov et al., 2016). For example, results by microscopy indicate that mRNA levels of targeted genes fluctuate from cell to cell due to the intrinsically stochastic, infrequent events of gene activation (Raj et al., 2006). Transcriptional activation can reposition genes in space (Zink et al., 2004), possibly by the action of RNA polymerases (Heinz et al., 2018; Brandão et al., 2019). Moreover, chromatin marks exhibit high variability between cells (Takei et al., 2021a), such as the intensities of H3K4me3 histone mark at different gene bodies (Woodworth et al., 2021), that at some point may regulate chromatin compartmentalization (Wang et al., 2019) (see below). Moreover, H3K4me3 histone mark intensities at different gene bodies show great heterogeneity (Woodworth et al., 2021) or that chromatin marks exhibit high variability in embryonic stem cells (Takei et al., 2021a). Considering that transcriptional activation can reposition genes in space (Zink et al., 2004) by the action of RNA polymerases (Heinz et al., 2018; Brandão et al., 2019) or that histone modifications may regulate chromatin compartmentalization (Wang et al., 2019) (see below). Therefore, the variability in gene expression and/or epigenetic status could have a direct effect on the observed stochasticity in genome conformation at the compartment and TAD levels (Ulianov and Razin, 2021). The influence and interdependence between genome organization, epigenomics and transcription is a central question in cellular biology. In the next section, I will address this by dissecting the current knowledge on the cellular processes directing them.

MOLECULAR MECHANISMS OF SPATIAL ORGANIZATION

The two types of 3D organization found in mammalian chromosomes form by independent mechanisms (Schwarzer et al., 2017; Nuebler et al., 2018). In contrast to what was once thought, there is no hierarchy between compartments and TADs, but rather a competition between two different organization modes. The self-organization principle of the genome (Misteli, 2020) indicates that chromatin of the same type tends to interact in the space and this is directly related to the polymeric nature of the genome, although the mechanism remains elusive. Polymer physics has modeled the genome as consecutive blocks of alternating active/inactive chromatin (block copolymers), that

assemble to generate the observed compartmentalization (Jost et al., 2014; Hildebrand and Dekker, 2020). It has been proposed that such compartments can arise through polymer phase separation mediated by associations of chromatin domains of similar epigenetic and/or transcriptional state (Rowley et al., 2017; Cook and Marenduzzo, 2018; Erdel and Rippe, 2018). Furthermore, a recent Hi-C study of outstanding sequencing deep revealed that median size of A/B compartments intervals is only 12.5 kb, and that even kilobase-sized domains show enhanced interactions with regions of the same class (Gu et al., 2021). However, the molecular bases of these associations are unknown.

Although spatial genetics approaches established that interactions between compartments vary from cell to cell, B-B domain contact frequencies were higher than A-A domains at distances below 75 Mb but not at higher genomic distances (Su et al., 2020), consistent with Hi-C studies in mammalian (Abramo et al., 2019) or *Drosophila* cells (Ulianov et al., 2021). The latter indicates that the mechanism of compartment segregation differs according to chromatin type. Accordingly, different players have been proposed, such as HP1alpha-mediated heterochromatin phase segregation (Larson et al., 2017; Strom et al., 2017) or clustering of active transcription sites (Hilbert et al., 2021). Based on polymer simulations, it was proposed that interactions between heterochromatin regions control compartmentalization over euchromatin contacts or the interaction of heterochromatin with the nuclear lamina (Falk et al., 2019). Recently, the role of homotypic repetitive elements and their RNA products has also been suggested as a mechanism of chromatin organization (Lu et al., 2021).

One spatial genomics study was able to establish a "chromatin profile" based on the multiplexed detection of several histone marks at specific DNA locus that matched ChIP-seq (Shen et al., 2012) or SPRITE measurements (Quinodoz et al., 2018) at 1-Mb resolution, but in this case with single-cell information. This analysis found "fixed" loci that, despite the variability in genome organization, are consistently associated with particular hallmarks (e.g., nuclear speckles, H3K9me3 mark, etc.) in most of the cells (Takei et al., 2021b). The existence of such "anchoring" points on each chromosome restricts their possible conformations. Because the spatial organization of nuclear bodies is cell-type dependent, they postulate that the nuclear architecture arises from the interaction between fixed or dominant loci with them. Moreover, related cell types have similar A/B compartment organization but very different nucleolar and lamina associations (Liu et al., 2020).

The loop-extrusion mechanism (Alipour and Marko, 2012) is to date the most accepted model of TAD formation in mammalian genomes (Nuebler et al., 2018). It postulates that the ring-shaped cohesin complex acts as a molecular motor actively extruding DNA and forming increasingly long chromatin loops that are stalled at convergent CTCF sites (Sanborn et al., 2015; Fudenberg et al., 2016; Fudenberg et al., 2017). Once bound to chromatin, the cohesin ring stochastically detaches from it, giving rise to highly dynamic structures (Hansen et al., 2017). This paradigm explains the Hi-C data showing the existence of chromatin loops between eTAD boundaries that present CTCF and cohesin complexes (Rao et al., 2014; Bonev et al., 2017). Moreover, when cohesin-loading factor Nipbl is removed from mouse cells, eTAD organization is lost (Schwarzer et al., 2017). However, chromatin tracing indicates that in single cells pairs of eTADs boundaries do not show a smaller physical distance distribution compared to control loci (Su et al., 2020) but rather there is a progressive looping anchored at the stronger CTCF binding site that progresses to more and more downstream loci (Beckwith et al., 2021). More strikingly, TAD-like domains persist upon cohesin depletion, although the boundary positions are randomized (Bintu et al., 2018). In line with this, even genomic regions that do now display eTADs form domain-like structures indicating that the folding of chromatin into this architecture is an intrinsic characteristic and that loop extrusion is a regulator of this process.

The process of compartmentalization and TAD formation shapes the genome architecture and changes the chromatin accessibility of genes and regulatory elements, modulating the functional output of genomes (Rowley and Corces, 2018). Among different cell types, the general principles of single-cell genome organization delineated above are conserved. However, cell-type specific spatial configurations delineate the functional differences (Liu et al., 2020; Takei et al., 2021b). Based on microscopy observations, we have proposed through the concept of "modulated stochasticity" that subtle changes in interaction frequencies give rise to measurable differences in genome architecture and could have a meaningful role in gene regulation (Cattoni et al., 2017). Complementary, nuclear structures such as speckles, which in practice act as chromatin scaffolds, might define different cell types and states. The stochasticity of genome architecture is a consequence of its polymeric nature, and it is modulated by several mechanisms mediated by proteins that interact through the sequence information. These mechanisms include, but are not restricted to, the processes of compartmentalization and loop extrusion. In general, sequences encode information for specific protein binding whose abundance and action will generate/regulate contacts between genomic loci.

CONCLUSION AND FUTURE PERSPECTIVES

In this review, I have summarized the technological improvements and recent discoveries of spatial genomic approaches. These advances, together with single-cell sequencing methods, are shifting the focus from ensemble measurements to the organization of genomes at the single-cell level to account for the observed high degree of stochasticity and heterogeneity.

Genome organization is shaped from its polymeric nature together with biological processes such as loop-extrusion, which are both stochastic in nature. The question that emerges is what is the biological relevance of such variable organization. In other words, how the genome architecture shapes transcription: the structure/function *conundrum*.

Maybe the important point here is it not anymore whether genome conformation is cause or consequence of genome function but rather what is the relationship between them. Furthermore, epigenetics and gene expression display a high degree of cell-to-cell variability. In order to reveal the contribution of each aspect to the function of genomes, new technologies capable of simultaneous detection of transcriptional output, epigenetic state and 3D conformation *in the same cell* will have to emerge. Undoubtedly, live-cell measurements, currently limited in scope, will also be necessary to understand the temporal aspects of genome organization. More importantly, despite the efforts, the function and activity of TADs and nuclear compartments continue to be unresolved. How are the specific genomic interactions generated if such heterogeneity is present? Moreover, how stochasticity is modulated to allow for precise spatio-temporal regulation of gene expression? Further developments in microscopy, genome-wide approaches and polymer simulations hold promise for the understanding of these key questions.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Abramo, K., Valton, A.-L., Venev, S. V., Ozadam, H., Fox, A. N., and Dekker, J. (2019). A Chromosome Folding Intermediate at the Condensin-To-Cohesin Transition during Telophase. *Nat. Cell Biol.* 21, 1393–1402. doi:10.1038/s41556-019-0406-2

Alipour, E., and Marko, J. F. (2012). Self-Organization of Domain Structures by DNA-Loop-Extruding Enzymes. *Nucleic Acids Res.* 40, 11202–11212. doi:10.1093/nar/gks925

Arrastia, M. V., Jachowicz, J. W., Ollikainen, N., Curtis, M. S., Lai, C., Quinodoz, S. A., et al. (2021). Single-Cell Measurement of Higher-Order 3D Genome Organization with scSPRITE. *Nat. Biotechnol.* doi:10.1038/s41587-021-00998-1

Beckwith, K. S., Ødegård-Fougner, Ø., Morero, N. R., Barton, C., Schueder, F., Alexander, S., et al. (2021). Visualization of Loop Extrusion by DNA Nanoscale Tracing in Single Human Cells. *bioRxiv.* doi:10.1101/2021.04.12.439407

Beliveau, B. J., Boettiger, A. N., Avendaño, M. S., Jungmann, R., McCole, R. B., Joyce, E. F., et al. (2015). Single-Molecule Super-Resolution Imaging of Chromosomes and *In Situ* Haplotype Visualization Using Oligopaint FISH Probes. *Nat. Commun.* 6, 7147. doi:10.1038/ncomms8147

Beliveau, B. J., Joyce, E. F., Apostolopoulos, N., Yilmaz, F., Fonseka, C. Y., McCole, R. B., et al. (2012). Versatile Design and Synthesis Platform for Visualizing Genomes with Oligopaint FISH Probes. *Proc. Natl. Acad. Sci.* 109, 21301–21306. doi:10.1073/pnas.1213818110

Beliveau, B. J., Kishi, J. Y., Nir, G., Sasaki, H. M., Saka, S. K., Nguyen, S. C., et al. (2018). OligoMiner Provides a Rapid, Flexible Environment for the Design of Genome-Scale Oligonucleotide *In Situ* Hybridization Probes. *Proc. Natl. Acad. Sci. USA* 115, E2183–E2192. doi:10.1073/pnas.1714530115

Bienko, M., Crosetto, N., Teytelman, L., Klemm, S., Itzkovitz, S., and van Oudenaarden, A. (2013). A Versatile Genome-Scale PCR-Based Pipeline for High-Definition DNA FISH. *Nat. Methods* 10, 122–124. doi:10.1038/nmeth.2306

Bintu, B., Mateo, L. J., Su, J.-H., Sinnott-Armstrong, N. A., Parker, M., Kinrot, S., et al. (2018). Super-Resolution Chromatin Tracing Reveals Domains and Cooperative Interactions in Single Cells. *Science* 362, eaau1783. doi:10.1126/science.aau1783

Boettiger, A., and Murphy, S. (2020). Advances in Chromatin Imaging at Kilobase-Scale Resolution. *Trends Genet.* 36, 273–287. doi:10.1016/j.tig.2019.12.010

Boettiger, A. N., Bintu, B., Moffitt, J. R., Wang, S., Beliveau, B. J., Fudenberg, G., et al. (2016). Super-Resolution Imaging Reveals Distinct Chromatin Folding for Different Epigenetic States. *Nature* 529, 418–422. doi:10.1038/nature16496

Bohrer, C. H., and Larson, D. R. (2021). The Stochastic Genome and its Role in Gene Expression. *Cold Spring Harb. Perspect. Biol.* 13, a040386. doi:10.1101/cshperspect.a040386

Bolzer, A., Kreth, G., Solovei, I., Koehler, D., Saracoglu, K., Fauth, C., et al. (2005). Three-Dimensional Maps of All Chromosomes in Human Male Fibroblast Nuclei and Prometaphase Rosettes. *PLoS Biol.* 3, e157. doi:10.1371/journal.pbio.0030157

Bonev, B., and Cavalli, G. (2016). Organization and Function of the 3D Genome. *Nat. Rev. Genet.* 17, 661–678. doi:10.1038/nrg.2016.112

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G. L., Lubling, Y., et al. (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* 171, 557–572.e24. doi:10.1016/j.cell.2017.09.043

Brandão, H. B., Gabriele, M., and Hansen, A. S. (2021). Tracking and Interpreting Long-Range Chromatin Interactions with Super-Resolution Live-Cell Imaging. *Curr. Opin. Cell Biol.* 70, 18–26. doi:10.1016/j.ceb.2020.11.002

Brandão, H. B., Paul, P., van den Berg, A. A., Rudner, D. Z., Wang, X., and Mirny, L. A. (2019). RNA Polymerases as Moving Barriers to Condensin Loop Extrusion. *Proc. Natl. Acad. Sci. USA* 116, 20489–20499. doi:10.1073/pnas.1907009116

Buchwalter, A., Kaneshiro, J. M., and Hetzer, M. W. (2019). Coaching from the Sidelines: the Nuclear Periphery in Genome Regulation. *Nat. Rev. Genet.* 20, 39–50. doi:10.1038/s41576-018-0063-5

Cardozo Gizzi, A. M., Cattoni, D. I., Fiche, J.-B., Espinola, S. M., Gurgo, J., Messina, O., et al. (2019). Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Mol. Cell* 74, 212–222.e5. doi:10.1016/j.molcel.2019.01.011

Cattoni, D. I., Cardozo Gizzi, A. M., Georgieva, M., Di Stefano, M., Valeri, A., Chamousset, D., et al. (2017). Single-Cell Absolute Contact Probability Detection Reveals Chromosomes Are Organized by Multiple Low-Frequency yet Specific Interactions. *Nat. Commun.* 8, 1753. doi:10.1038/s41467-017-01962-x

Cavalheiro, G. R., Pollex, T., and Furlong, E. E. (2021). To Loop or Not to Loop: what Is the Role of TADs in Enhancer Function and Gene Regulation? *Curr. Opin. Genet. Develop.* 67, 119–129. doi:10.1016/j.gde.2020.12.015

Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S., and Zhuang, X. (2015). RNA Imaging. Spatially Resolved, Highly Multiplexed RNA Profiling in Single Cells. *Science* 348, aaa6090. doi:10.1126/science.aaa6090

Cheng, Y., Liu, M., Hu, M., and Wang, S. (2021). TAD-Like Single-Cell Domain Structures Exist on Both Active and Inactive X Chromosomes and Persist under Epigenetic Perturbations. *Genome Biol.* 22, 309. doi:10.1186/s13059-021-02523-8

Cook, P. R., and Marenduzzo, D. (2018). Transcription-Driven Genome Organization: a Model for Chromosome Structure and the Regulation of Gene Expression Tested through Simulations. *Nucleic Acids Res.* 46, 9895–9906. doi:10.1093/nar/gky763

Cremer, T., and Cremer, C. (2001). Chromosome Territories, Nuclear Architecture and Gene Regulation in Mammalian Cells. *Nat. Rev. Genet.* 2, 292–301. doi:10.1038/35066075

Cremer, T., Cremer, M., Dietzel, S., Müller, S., Solovei, I., and Fakan, S. (2006). Chromosome Territories - a Functional Nuclear Landscape. *Curr. Opin. Cell Biol.* 18, 307–316. doi:10.1016/j.ceb.2006.04.007

Crosetto, N., and Bienko, M. (2020). Radial Organization in the Mammalian Nucleus. *Front. Genet.* 11. doi:10.3389/fgene.2020.00033

Dixon, J. R., Gorkin, D. U., and Ren, B. (2016). Chromatin Domains: The Unit of Chromosome Organization. *Mol. Cell* 62, 668–680. doi:10.1016/j.molcel.2016.05.018

Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., et al. (2015). Chromatin Architecture Reorganization during Stem Cell Differentiation. *Nature* 518, 331–336. doi:10.1038/nature14222

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012). Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. *Nature* 485, 376–380. doi:10.1038/nature11082

Eng, C.-H. L., Lawson, M., Zhu, Q., Dries, R., Koulena, N., Takei, Y., et al. (2019). Transcriptome-Scale Super-Resolved Imaging in Tissues by RNA seqFISH+. *Nature* 568, 235–239. doi:10.1038/s41586-019-1049-y

Erdel, F., and Rippe, K. (2018). Formation of Chromatin Subcompartments by Phase Separation. *Biophysical J.* 114, 2262–2270. doi:10.1016/j.bpj.2018.03.011

Espinola, S. M., Götz, M., Bellec, M., Messina, O., Fiche, J.-B., Houbron, C., et al. (2021). Cis-Regulatory Chromatin Loops Arise before TADs and Gene Activation, and Are Independent of Cell Fate during Early *Drosophila* Development. *Nat. Genet.* 53, 477–486. doi:10.1038/s41588-021-00816-z

Falk, M., Feodorova, Y., Naumova, N., Imakaev, M., Lajoie, B. R., Leonhardt, H., et al. (2019). Heterochromatin Drives Compartmentalization of Inverted and Conventional Nuclei. *Nature* 570, 395–399. doi:10.1038/s41586-019-1275-3

Finn, E. H., and Misteli, T. (2019). Molecular Basis and Biological Function of Variability in Spatial Genome Organization. *Science* 365. doi:10.1126/science.aaw9498

Finn, E. H., Pegoraro, G., Brandão, H. B., Valton, A.-L., Oomen, M. E., Dekker, J., et al. (2019). Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell* 176, 1502–1515.e10. doi:10.1016/j.cell.2019.01.020

Flyamer, I. M., Gassler, J., Imakaev, M., Brandão, H. B., Ulianov, S. V., Abdennur, N., et al. (2017). Single-Nucleus Hi-C Reveals Unique Chromatin Reorganization at Oocyte-To-Zygote Transition. *Nature* 544, 110–114. doi:10.1038/nature21711

Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A., and Mirny, L. A. (2017). Emerging Evidence of Chromosome Folding by Loop Extrusion. *Cold Spring Harb. Symp. Quant. Biol.* 82, 45–55. doi:10.1101/sqb.2017.82.034710

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L. A. (2016). Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* 15, 2038–2049. doi:10.1016/j.celrep.2016.04.085

Furlong, E. E. M., and Levine, M. (2018). Developmental Enhancers and Chromosome Topology. *Science* 361, 1341–1345. doi:10.1126/science.aau0320

Galganski, L., Urbanek, M. O., and Krzyzosiak, W. J. (2017). Nuclear Speckles: Molecular Organization, Biological Function and Role in Disease. *Nucleic Acids Res.* 45, 10350–10368. doi:10.1093/nar/gkx759

Galitsyna, A. A., and Gelfand, M. S. (2021). Single-Cell Hi-C Data Analysis: Safety in Numbers. *Brief. Bioinform.* doi:10.1093/bib/bbab316

Galupa, R., and Heard, E. (2018). X-Chromosome Inactivation: A Crossroads between Chromosome Architecture and Gene Regulation. *Annu. Rev. Genet.* 52, 535–566. doi:10.1146/annurev-genet-120116-024611

Giorgetti, L., and Heard, E. (2016). Closing the Loop: 3C versus DNA FISH. *Genome Biol.* 17, 215. doi:10.1186/s13059-016-1081-2

Goel, V. Y., and Hansen, A. S. (2021). The Macro and Micro of Chromosome Conformation Capture. *Wiley Interdiscip. Rev. Dev. Biol.* 10, e395. doi:10.1002/wdev.395

Golov, A. K., Razin, S. V., and Gavrilov, A. A. (2016). Single-Cell Genome-Wide Studies Give New Insight into Nongenetic Cell-To-Cell Variability in Animals. *Histochem. Cell Biol.* 146, 239–254. doi:10.1007/s00418-016-1466-z

Gu, H., Harris, H., Olshansky, M., Eliaz, Y., Krishna, A., Kalluchi, A., et al. (2021). Fine-Mapping of Nuclear Compartments Using Ultra-Deep Hi-C Shows that Active Promoter and Enhancer Elements Localize in the Active A Compartment Even when Adjacent Sequences Do Not. *bioRxiv*, 462599. doi:10.0310.1101/2021.10.03.462599

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M. B., Talhout, W., et al. (2008). Domain Organization of Human Chromosomes Revealed by Mapping of Nuclear Lamina Interactions. *Nature* 453, 948–951. doi:10.1038/nature06947

Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R., and Darzacq, X. (2017). CTCF and Cohesin Regulate Chromatin Loop Stability with Distinct Dynamics. *Elife* 6. doi:10.7554/eLife.25776

Heinz, S., Texari, L., Hayes, M. G. B., Urbanowski, M., Chang, M. W., Givarkes, N., et al. (2018). Transcription Elongation Can Affect Genome 3D Structure. *Cell* 174, 1522–1536.e22. doi:10.1016/j.cell.2018.07.047

Hilbert, L., Sato, Y., Kuznetsova, K., Bianucci, T., Kimura, H., Jülicher, F., et al. (2021). Transcription Organizes Euchromatin via Microphase Separation. *Nat. Commun.* 12, 1–12. doi:10.1038/s41467-021-21589-3

Hildebrand, E. M., and Dekker, J. (2020). Mechanisms and Functions of Chromosome Compartmentalization. *Trends Biochem. Sci.* 45, 385–396. doi:10.1016/j.tibs.2020.01.002

Jerkovic´, I., and Cavalli, G. (2021). Understanding 3D Genome Organization by Multidisciplinary Methods. *Nat. Rev. Mol. Cell Biol.* 22, 511–528. doi:10.1038/s41580-021-00362-w

Jost, D., Carrivain, P., Cavalli, G., and Vaillant, C. (2014). Modeling Epigenome Folding: Formation and Dynamics of Topologically Associated Chromatin Domains. *Nucleic Acids Res.* 42, 9553–9561. doi:10.1093/nar/gku698

Larson, A. G., Elnatan, D., Keenen, M. M., Trnka, M. J., Johnston, J. B., Burlingame, A. L., et al. (2017). Liquid Droplet Formation by HP1α Suggests a Role for Phase Separation in Heterochromatin. *Nature* 547, 236–240. doi:10.1038/nature22822

Lee, J. H., Daugharthy, E. R., Scheiman, J., Kalhor, R., Yang, J. L., Ferrante, T. C., et al. (2014). Highly Multiplexed Subcellular RNA Sequencing *In Situ*. *Science* 343, 1360–1363. doi:10.1126/science.1250212

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* 326, 289–293. doi:10.1126/science.1181369

Liu, M., Lu, Y., Yang, B., Chen, Y., Radda, J. S. D., Hu, M., et al. (2020). Multiplexed Imaging of Nucleome Architectures in Single Cells of Mammalian Tissue. *Nat. Commun.* 11, 2907. doi:10.1038/s41467-020-16732-5

Lowenstein, M. G., Goddard, T. D., and Sedat, J. W. (2004). Long-Range Interphase Chromosome Organization in *Drosophila*: A Study Using Color Barcoded Fluorescence *In Situ* Hybridization and Structural Clustering Analysis. *MBoC* 15, 5678–5692. doi:10.1091/mbc.e04-04-0289

Lu, J. Y., Chang, L., Li, T., Wang, T., Yin, Y., Zhan, G., et al. (2021). Homotypic Clustering of L1 and B1/Alu Repeats Compartmentalizes the 3D Genome. *Cell Res* 31, 613–630. doi:10.1038/s41422-020-00466-6

Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M., and Cai, L. (2014). Single-Cell *In Situ* RNA Profiling by Sequential Hybridization. *Nat. Methods* 11, 360–361. doi:10.1038/nmeth.2892

Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., et al. (2015). Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* 161, 1012–1025. doi:10.1016/j.cell.2015.04.004

Luppino, J. M., Park, D. S., Nguyen, S. C., Lan, Y., Xu, Z., Yunker, R., et al. (2020). Cohesin Promotes Stochastic Domain Intermingling to Ensure Proper Regulation of Boundary-Proximal Genes. *Nat. Genet.* 52, 840–848. doi:10.1038/s41588-020-0647-9

Mao, Y. S., Zhang, B., and Spector, D. L. (2011). Biogenesis and Function of Nuclear Bodies. *Trends Genet.* 27, 295–306. doi:10.1016/j.tig.2011.05.006

Mateo, L. J., Murphy, S. E., Hafner, A., Cinquini, I. S., Walker, C. A., and Boettiger, A. N. (2019). Visualizing DNA Folding and RNA in Embryos at Single-Cell Resolution. *Nature* 568, 49–54. doi:10.1038/s41586-019-1035-4

Misteli, T. (2005). Concepts in Nuclear Architecture. *BioEssays* 27, 477–487. doi:10.1002/bies.20226

Misteli, T. (2020). The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell* 183, 28–45. doi:10.1016/j.cell.2020.09.014

Nagano´, T., Lubling, Y., Várnai, C., Dudley, C., Leung, W., Baran, Y., et al. (2017). Cell-Cycle Dynamics of Chromosomal Organization at Single-Cell Resolution. *Nature* 547, 61–67. doi:10.1038/nature23001

Nguyen, H. Q., Chattoraj, S., Castillo, D., Nguyen, S. C., Nir, G., Lioutas, A., et al. (2020). 3D Mapping and Accelerated Super-Resolution Imaging of the Human Genome Using *In Situ* Sequencing. *Nat. Methods* 17, 822–832. doi:10.1038/s41592-020-0890-0

Nir, G., Farabella, I., Pérez Estrada, C., Ebeling, C. G., Beliveau, B. J., Sasaki, H. M., et al. (2018). Walking along Chromosomes with Super-Resolution Imaging, Contact Maps, and Integrative Modeling. *PLoS Genet.* 14, e1007872. doi:10.1371/journal.pgen.1007872

Nollmann, M., Bennabi, I., Götz, M., and Gregor, T. (2021). The Impact of Space and Time on the Functional Output of the Genome. *Cold Spring Harb. Perspect. Biol.*, a040378. doi:10.1101/cshperspect.a040378

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., et al. (2012). Spatial Partitioning of the Regulatory Landscape of the X-Inactivation centre. *Nature* 485, 381–385. doi:10.1038/nature11049

Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N., and Mirny, L. A. (2018). Chromatin Organization by an Interplay of Loop Extrusion and Compartmental Segregation. *Proc. Natl. Acad. Sci. USA* 115, E6697–E6706. doi:10.1073/pnas.1717730115

Passarge, E. (1979). Emil Heitz and the Concept of Heterochromatin: Longitudinal Chromosome Differentiation Was Recognized Fifty Years Ago. *Am. J. Hum. Genet.* 31, 106–115.

Payne, A. C., Chiang, Z. D., Reginato, P. L., Mangiameli, S. M., Murray, E. M., Yao, C.-C., et al. (2021). *In Situ* genome Sequencing Resolves DNA Sequence and Structure in Intact Biological Samples. *Science* 371, eaay3446. doi:10.1126/science.aay3446

Pope, B. D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., et al. (2014). Topologically Associating Domains Are Stable Units of Replication-Timing Regulation. *Nature* 515, 402–405. doi:10.1038/nature13986

Quinodoz, S. A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J. M., Detmar, E., et al. (2018). Higher-Order Inter-Chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* 174, 744–757.e24. doi:10.1016/j.cell.2018.05.024

Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y., and Tyagi, S. (2006). Stochastic mRNA Synthesis in Mammalian Cells. *PLoS Biol.* 4, e309. doi:10.1371/journal.pbio.0040309

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* 159, 1665–1680. doi:10.1016/j.cell.2014.11.021

Roohi, J., Cammer, M., Montagna, C., and Hatchwell, E. (2008). An Improved Method for Generating BAC DNA Suitable for FISH. *Cytogenet. Genome Res.* 121, 7–9. doi:10.1159/000124374

Rowley, M. J., and Corces, V. G. (2018). Organizational Principles of 3D Genome Architecture. *Nat. Rev. Genet.* 19, 789–800. doi:10.1038/s41576-018-0060-8

Rowley, M. J., Nichols, M. H., Lyu, X., Ando-Kuri, M., Rivera, I. S. M., Hermetz, K., et al. (2017). Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol. Cell* 67, 837–852.e7. doi:10.1016/j.molcel.2017.07.022

Sanborn, A. L., Rao, S. S. P., Huang, S.-C., Durand, N. C., Huntley, M. H., Jewett, A. I., et al. (2015). Chromatin Extrusion Explains Key Features of Loop and Domain Formation in Wild-Type and Engineered Genomes. *Proc. Natl. Acad. Sci. USA* 112, E6456–E6465. doi:10.1073/pnas.1518552112

Sawh, A. N., Shafer, M. E. R., Su, J.-H., Zhuang, X., Wang, S., and Mango, S. E. (2020). Lamina-Dependent Stretching and Unconventional Chromosome Compartments in Early *C. elegans* Embryos. *Mol. Cell* 78, 96–111.e6. doi:10.1016/j.molcel.2020.02.006

Schoenfelder, S., and Fraser, P. (2019). Long-Range Enhancer-Promoter Contacts in Gene Expression Control. *Nat. Rev. Genet.* 20, 437–455. doi:10.1038/s41576-019-0128-0

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., et al. (2017). Two Independent Modes of Chromatin Organization Revealed by Cohesin Removal. *Nature* 551, 51–56. doi:10.1038/nature24281

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., et al. (2012). Three-Dimensional Folding and Functional Organization Principles of the *Drosophila* Genome. *Cell* 148, 458–472. doi:10.1016/j.cell.2012.01.010

Shah, S., Takei, Y., Zhou, W., Lubeck, E., Yun, J., Eng, C.-H. L., et al. (2018). Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. *Cell* 174, 363–376.e16. doi:10.1016/j.cell.2018.05.035

Shen, Y., Yue, F., McCleary, D. F., Ye, Z., Edsall, L., Kuan, S., et al. (2012). A Map of the Cis-Regulatory Sequences in the Mouse Genome. *Nature* 488, 116–120. doi:10.1038/nature11243

Stevens, T. J., Lando, D., Basu, S., Atkinson, L. P., Cao, Y., Lee, S. F., et al. (2017). 3D Structures of Individual Mammalian Genomes Studied by Single-Cell Hi-C. *Nature* 544, 59–64. doi:10.1038/nature21429

Strom, A. R., Emelyanov, A. V., Mir, M., Fyodorov, D. V., Darzacq, X., and Karpen, G. H. (2017). Phase Separation Drives Heterochromatin Domain Formation. *Nature* 547, 241–245. doi:10.1038/nature22989

Su, J.-H., Zheng, P., Kinrot, S. S., Bintu, B., and Zhuang, X. (2020). Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell* 182, 1641–1659.e26. doi:10.1016/j.cell.2020.07.032

Szabo, Q., Jost, D., Chang, J. M., Cattoni, D. I., Papadopoulos, G. L., Bonev, B., et al. (2018). TADs Are 3D Structural Units of Higher-Order Chromosome Organization in *Drosophila*. *Sci. Adv.* 4, eaar8082. doi:10.1126/sciadv.aar8082

Szabo, Q., Donjon, A., Jerković, I., Papadopoulos, G. L., Cheutin, T., Bonev, B., et al. (2020). Regulation of Single-Cell Genome Organization into TADs and Chromatin Nanodomains. *Nat. Genet.* 52, 1151–1157. doi:10.1038/s41588-020-00716-8

Takei, Y., Yun, J., Zheng, S., Ollikainen, N., Pierson, N., White, J., et al. (2021a). Integrated Spatial Genomics Reveals Global Architecture of Single Nuclei. *Nature* 590, 344–350. doi:10.1038/s41586-020-03126-2

Takei, Y., Zheng, S., Yun, J., Shah, S., Pierson, N., White, J., et al. (2021b). Single-Cell Nuclear Architecture across Cell Types in the Mouse Brain. *Science*, eabj1966. doi:10.1126/science.abj1966

Tan, L., Xing, D., Chang, C.-H., Li, H., and Xie, X. S. (2018). Three-Dimensional Genome Structures of Single Diploid Human Cells. *Science* 361, 924–928. doi:10.1126/science.aat5641

Ulianov, S. V., Khrameeva, E. E., Gavrilov, A. A., Flyamer, I. M., Kos, P., Mikhaleva, E. A., et al. (2016). Active Chromatin and Transcription Play a Key Role in Chromosome Partitioning into Topologically Associating Domains. *Genome Res.* 26, 70–84. doi:10.1101/gr.196006.115

Ulianov, S. V., and Razin, S. V. (2021). The Two Waves in Single-Cell 3D Genomics. *Semin. Cell Dev. Biol.* doi:10.1016/j.semcdb.2021.05.021

Ulianov, S. V., Tachibana-Konwalski, K., and Razin, S. V. (2017). Single-Cell Hi-C Bridges Microscopy and Genome-Wide Sequencing Approaches to Study 3D Chromatin Organization. *Bioessays* 39, 1700104. doi:10.1002/bies.201700104

Ulianov, S. V., Zakharova, V. V., Galitsyna, A. A., Kos, P. I., Polovnikov, K. E., Flyamer, I. M., et al. (2021). Order and Stochasticity in the Folding of Individual *Drosophila* Genomes. *Nat. Commun.* 12, 41. doi:10.1038/s41467-020-20292-z

Wang, L., Gao, Y., Zheng, X., Liu, C., Dong, S., Li, R., et al. (2019). Histone Modifications Regulate Chromatin Compartmentalization by Contributing to a Phase Separation Mechanism. *Mol. Cell* 76, 646–659.e6. doi:10.1016/j.molcel.2019.08.019

Wang, Q., Sun, Q., Czajkowsky, D. M., and Shao, Z. (2018). Sub-kb Hi-C in *D. melanogaster* Reveals Conserved Characteristics of TADs between Insect and Mammalian Cells. *Nat. Commun.* 9, 188–8. doi:10.1038/s41467-017-02526-9

Wang, S., Su, J.-H., Beliveau, B. J., Bintu, B., Moffitt, J. R., Wu, C.-T., et al. (2016). Spatial Organization of Chromatin Domains and Compartments in Single Chromosomes. *Science* 353, 598–602. doi:10.1126/science.aaf8084

Woodworth, M. A., Ng, K. K. H., Halpern, A. R., Pease, N. A., Nguyen, P. H. B., Kueh, H. Y., et al. (2021). Multiplexed Single-Cell Profiling of Chromatin States at Genomic Loci by Expansion Microscopy. *Nucleic Acids Res.* 49, e82. doi:10.1093/nar/gkab423

Xie, L., and Liu, Z. (2021). Single-Cell Imaging of Genome Organization and Dynamics. *Mol. Syst. Biol.* 17, e9653. doi:10.15252/msb.20209653

Zheng, M., Tian, S. Z., Capurso, D., Kim, M., Maurya, R., Lee, B., et al. (2019). Multiplex Chromatin Interactions with Single-Molecule Precision. *Nature* 566, 558–562. doi:10.1038/s41586-019-0949-1

Zink, D., Amaral, M. D., Englmann, A., Lang, S., Clarke, L. A., Rudolph, C., et al. (2004). Transcription-Dependent Spatial Arrangements of CFTR and Adjacent Genes in Human Cell Nuclei. *J. Cell Biol.* 166, 815–825. doi:10.1083/jcb.200404107

Understanding 3D Genome Organization and Its Effect on Transcriptional Gene Regulation Under Environmental Stress in Plant: A Chromatin Perspective

Suresh Kumar[1]*, Simardeep Kaur[1], Karishma Seem[1], Santosh Kumar[2] and Trilochan Mohapatra[3]

[1]Division of Biochemistry, ICAR-Indian Agricultural Research Institute, New Delhi, India, [2]Decode Genomics Private Limited, New Delhi, India, [3]Indian Council of Agricultural Research, New Delhi, India

The genome of a eukaryotic organism is comprised of a supra-molecular complex of chromatin fibers and intricately folded three-dimensional (3D) structures. Chromosomal interactions and topological changes in response to the developmental and/or environmental stimuli affect gene expression. Chromatin architecture plays important roles in DNA replication, gene expression, and genome integrity. Higher-order chromatin organizations like chromosome territories (CTs), A/B compartments, topologically associating domains (TADs), and chromatin loops vary among cells, tissues, and species depending on the developmental stage and/or environmental conditions (4D genomics). Every chromosome occupies a separate territory in the interphase nucleus and forms the top layer of hierarchical structure (CTs) in most of the eukaryotes. While the A and B compartments are associated with active (euchromatic) and inactive (heterochromatic) chromatin, respectively, having well-defined genomic/ epigenomic features, TADs are the structural units of chromatin. Chromatin architecture like TADs as well as the local interactions between promoter and regulatory elements correlates with the chromatin activity, which alters during environmental stresses due to relocalization of the architectural proteins. Moreover, chromatin looping brings the gene and regulatory elements in close proximity for interactions. The intricate relationship between nucleotide sequence and chromatin architecture requires a more comprehensive understanding to unravel the genome organization and genetic plasticity. During the last decade, advances in chromatin conformation capture techniques for unravelling 3D genome organizations have improved our understanding of genome biology. However, the recent advances, such as Hi-C and ChIA-PET, have substantially increased the resolution, throughput as well our interest in analysing genome organizations. The present review provides an overview of the historical and contemporary perspectives of chromosome conformation capture technologies, their applications in functional genomics, and the constraints in predicting 3D genome organization. We also discuss the future perspectives of understanding high-order chromatin organizations in deciphering transcriptional

regulation of gene expression under environmental stress (4D genomics). These might help design the climate-smart crop to meet the ever-growing demands of food, feed, and fodder.

1 INTRODUCTION

A eukaryotic genome comprises several chromosomes, which vary along their length, contain supra-molecular complexes of chromatin fibers, and are intricately folded in a three-dimensional (3D) structure. The genome is not randomly positioned in the nucleus, but it is packed into higher-order chromatin structures that play important functional roles. Understanding the organization of the nuclear genome is seeking significant attention nowadays, as several processes like DNA replication, transcription, genome integrity, etc. involved in growth, development, and stress tolerance are regulated through the nuclear genome organization. Eukaryotic genome organization can be observed at three levels i) linear genome: the nucleotide sequence deciphered by DNA sequencing, ii) epigenome: representing the additional information added due to the modified bases and/or histone proteins which help regulate gene expression, and iii) 3D structure of the genome: representing the arrangement of chromatins/chromosomes in the nucleus (Bonev and Cavalli, 2016). These genome-level organizations are being studied with the help of recent advances in imaging and molecular biology techniques. To understand 3D genome structure, techniques like chromosome conformation capture (3C), chromosome conformation capture-on-chip (4C), chromosome conformation capture carbon copy (5C), chromatin interaction analysis by paired-end tag (ChIA-PET) sequencing, high-throughput chromosome conformation capture (Hi-C), and their derivatives are being used.

In eukaryotes, chromatin is packed into nucleosomes wherein histone proteins make up the largest component. DNA wrapped around a histone octamer (two units of each of the four core histones H2A, H2B, H3, and H4) sealed by a linker histone (H1) builds the structural constituent nucleosome to form chromatin. The chromatin-related research is progressing with unprecedented speed and resolution, deciphering the complex and dynamic chromatin architecture during cellular processes including DNA replication, recombination, repair, transcription, mitosis, and meiosis. Chromatin structures are highly dynamic, which undergo cyclic compaction and de-compaction during the cell cycle, cell differentiation, developmental processes, and defense responses. Chromatin accessibility to the regulatory elements like RNA polymerase II (RNA Pol-II) is affected by chromatin compaction/de-compaction, which fine-tunes the regulation of gene expression (Dixon et al., 2015; Boltsis et al., 2021). Differentiated cells have different cellular functions, and a different set of genes are expressed under different environmental conditions which require varying 3D genome architecture (Dixon et al., 2015). Changing environmental conditions (stresses)

interfere with several cellular processes, which might require modulation in chromatin architecture to adjust the gene expression in response to the stress (Sun L. et al., 2020). Nucleotide sequence alone does not carry the entire regulatory information, as interactions among the chromosomes and topological changes in response to the developmental and/or environmental stimuli affect the expression of genes. Transient rearrangement of chromatin architecture (the compact heterochromatin or loosely-packed euchromatin) and modulation in chromatin composition upon stress exposure are being demonstrated in animals and plants (Lupianez et al., 2015; Li X. et al., 2019; Fei et al., 2019; Sun L. et al., 2020). Interaction of distal regulatory elements with the promoter through physical proximity mediated by the chromatin structural proteins like CCCTC-binding factor (CTCF) and cohesin to regulate the transcription process is being reported in animals (Eser et al., 2017; Robson et al., 2019).

Being sessile, plants face numerous abiotic and biotic stresses throughout their life. Our understanding of chromatin organization in model species has advanced significantly in the past decade (Sexton et al., 2012; Bonev and Cavalli, 2016; Grob and Grossniklaus, 2017). Highly condensed chromatin, such as heterochromatin, prevents accessibility of the transcriptional machinery (transcription factors, polymerases, and other nuclear proteins) to the gene. An environmental signal may cause some alterations in chromatin architecture which make the gene accessible to transcriptional machinery. Such chromatin remodeling includes shifting or removal of histones (Perrella et al., 2020), the introduction of histone variants (Dai et al., 2017; Wang et al., 2020), or post-translational modifications of histone proteins, etc. (Eberharter and Becker, 2002; Clapier et al., 2017). Studies show the hierarchical organization of genomes, wherein chromosome territories (CTs) are at the top of the hierarchical structure, followed by the chromosome compartments, topologically associating domains (TADs) and gene body/ chromatin loops (Pecinka et al., 2004; Amano et al., 2009; Zhang and Wang, 2021). 3D genomics helps to decipher the spatial chromatin configurations and investigate their regulatory roles in gene expression (Gonzalez-Sandoval and Gasser, 2016).

Despite the absence of insulator protein CTCF in plants, TADs have rarely been observed in Arabidopsis. TAD-like domains and motifs at the TAD boundaries have been identified in rice (Liu et al., 2017). Moreover, cohesins subunits have also been identified in rice (Zhang et al., 2006; Tao et al., 2007; Gong et al., 2011). However, it is still not clear whether these cohesins have similar functions in plants. Inactive heterochromatic islands (IHIs) or KNOT engaged elements (KEEs) were reported to be present within euchromatin and exhibit strong long-range interactions in Arabidopsis (Feng et al., 2014; Grob et al.,

2014; Grob and Grossniklaus, 2019), rice (Dong et al., 2018), and Brassica (Ting et al., 2019). Therefore, future investigations on the identification of CTCF-like insulator proteins, KNOT, KEEs and their functions in plants would be required.

Due to the sessile nature of plants, they deploy highly evolved mechanisms to manage their growth and development under varying environmental conditions (abiotic and biotic stresses). During the last few decades, linear genomes and epigenomes of eukaryotes have been extensively studied towards understanding the regulation of gene expression. It is now evident that the information and function of a genome are modulated under varying environmental conditions not only by the epigenetic modifications in the linear DNA sequence but also by altering the 3D chromatin organization within the nucleus (Dogan and Liu, 2018; Grob, 2020). Gene activities are controlled/regulated by alterations in chromatin architecture *via* DNA methylation (Kumar et al., 2018; Kumar and Mohapatra, 2021), histone modifications (Rowley et al., 2017; Park et al., 2018), and chromatin remodelers (Peterson and Workman, 2000; Bhadouriya et al., 2021). Different chromatin remodelers such as CHD, INO80, ISWI, and Switch/Sucrose non-fermenting (SWI/SNF) have been reported to act upon chromatin under diverse environmental stresses to convert transcriptionally inactive chromatin to the transcriptionally active state. Chromatin architecture at the promoter region is more crucial for determining the level of gene expression (Tannenbaum et al., 2018; Barragán-Rosillo et al., 2021). Advances in chromatin visualization, NGS, and 3C-based techniques have accumulated evidence for chromosome architecture, chromatin domains/loops and different epigenetic modifications to be correlated with transcriptional activities (Zhang and Wang, 2021). Studies suggest a functional correlation among the changes in nuclear organization, stressful conditions, and the level of gene expression. Tight coiling of chromatin (a default state) restricts transcriptional expression of the gene, which gets expressed when the nearby chromatin is loosened (remodeled) (Bannister and Kouzarides, 2011). The accumulating datasets on epigenomics (Kumar and Mohapatra, 2021) and the evident roles of genome architecture on the regulation of gene expression (Zhang and Wang, 2021) indicate that 3D genomics would be an important player in deciphering the key regulators. Developmental and environmental stimuli affect epigenetic landscape and chromatin architecture, which are dynamic and modulate gene expression to cope with stress (Bhadouriya et al., 2021). Some of the transcriptional repressors communicate with chromatin remodeler, directly or indirectly, or alter the chromatin structure. Some of these modifications may get transmitted through cell division, and help cope with the stress on reoccurrence (Gallusci et al., 2017). However, further validation of the transmission of stress-induced changes in chromatin architecture and their role in stress tolerance would be required.

This review presents the recent advances in 3D genomics methods and focuses on understanding the 3D genome organization of plants with reference to the available knowledge of nuclear genome organization in the animal system. We also discuss the developments in chromosome conformation capture technologies, their relevance in understanding genome structure (genome assembly) and functions. Future perspectives of 3D genomics, with special reference to its application in plant/crop improvement, and the constraints currently being faced are also discussed.

2 UNDERSTANDING NUCLEAR GENOME ORGANIZATION

The eukaryotic genome is not randomly positioned in the nucleus, but it is packaged in a higher-order chromatin structure that plays important role in genome structure/functions. The spatial organization of chromatins allows an additional layer of regulatory information for transcriptional gene regulation that is far from the encoded information in the 1D genomic sequence (Lanctot et al., 2007). To explore the regulatory information of such organizational elements, the 3D structure of the genome has to be deciphered. Interacting nucleosomes make chromatin fiber, which physically interacts with the *cis*-acting elements to form chromatin loops (Crevillén et al., 2013). Structural proteins (CTCF, cohesin), transcription factors (TFs), and heterochromatin-binding proteins stabilize chromatin loops that form TADs (Wang Q. et al., 2018). TADs further interact to form chromatin compartments, which merge to constitute CTs (Lieberman-Aiden et al., 2009; Rao et al., 2014). Understanding 3D genome organization also demands to consider the sub-nuclear components like nuclear bodies (nucleolus, nuclear speckles, and Cajal bodies) and nuclear periphery (Mao et al., 2011; Rowley et al., 2017).

Genome organization is eminently dynamic, as it changes with the progression of the cell cycle, developmental transition (photomorphogenesis, flowering), and environmental cues (Kaiserli et al., 2018). In germinating Arabidopsis seedling, chromocenters were reported to be produced which could be visualized as large, bright spots on nuclear staining with DAPI (Bourbousse et al., 2015). Large chromatin regions associated with the nuclear periphery to form a network of lamina-associated domains (LADs), were reported in mammalian cells (Guelen et al., 2008). Some of the chromatin domains are also associated with the nucleolar periphery of nucleolus to form nucleolus-associated chromatin domains (NADs) (Nemeth et al., 2010; van Koningsbruggen et al., 2010). Although most of the 3D information on genome organization (e.g., LADs, NADs, TADs, etc.) in animals is comparable to that of plants, our knowledge of plant chromatin architecture is still in its infancy. Active and repressed chromatin regions are separated from each other in animals, and some of the nuclear compartments like nuclear-periphery and nucleolar-periphery are enriched with heterochromatin (repressed chromatin) (van Steensel and Belmont, 2017; Bersaglieri and Santoro, 2019). The chromatin domains localized at the nuclear/nucleolar periphery in Arabidopsis have been recently identified (Hu et al., 2019; Sun L. et al., 2020).

2.1 Deciphering 3D Genome Organization
Experimentation with 3D genome organization reveals that chromosomes occupy distinct nuclear spaces in the eukaryotic
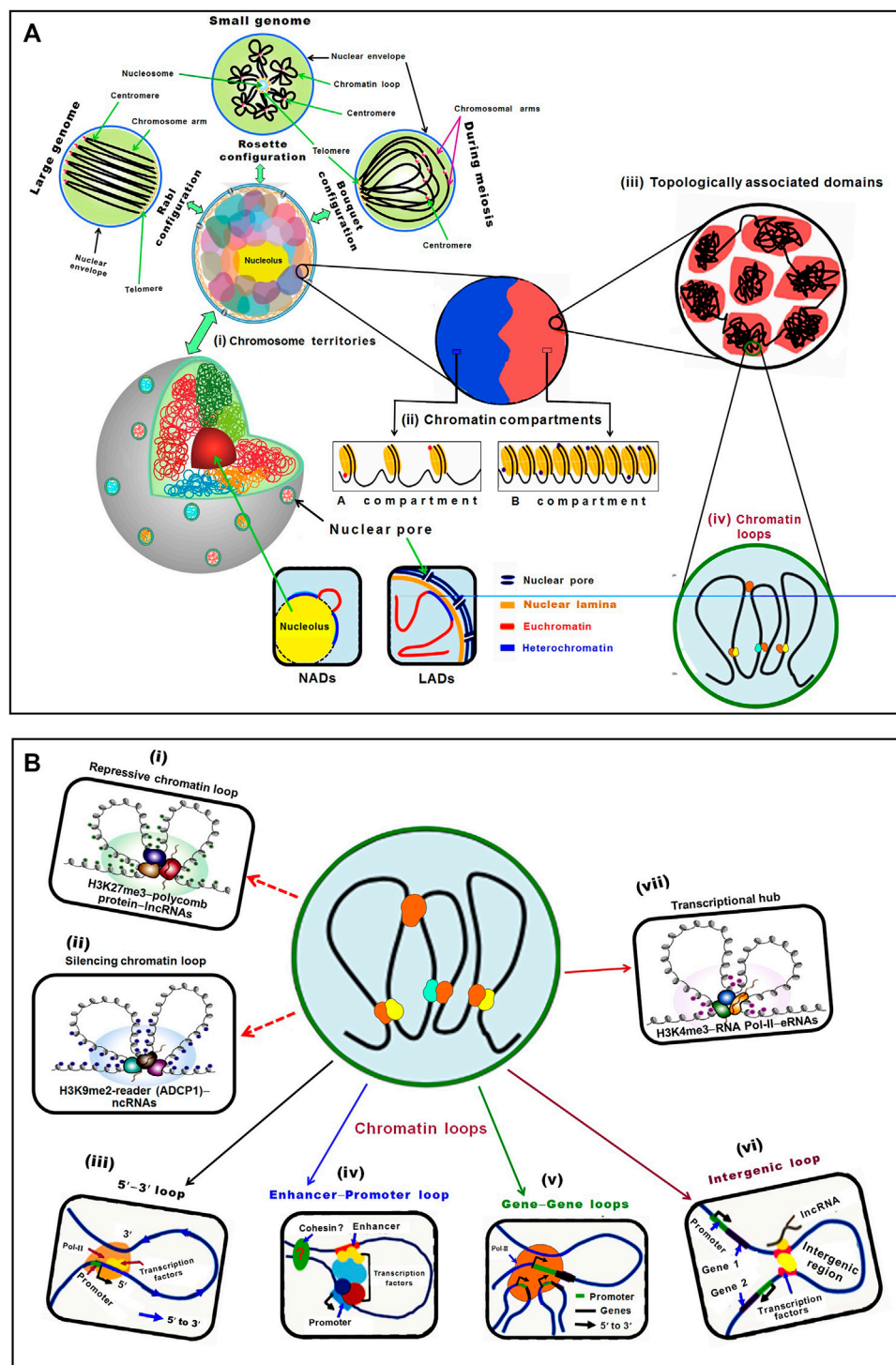
FIGURE 1 | Schematic representation of plant chromatin organization in the nucleus. **(A)** Hierarchical chromatin organization can be studied mainly at four levels: chromosome territory, chromatin compartments, topologically associating domain (TAD), and chromatin loops. (i) Chromosomes occupy specific territories in the nucleus. In different territories, chromosomes show different morphology, such as Rabl, Rosette, and Bouquet configuration. In Rabl configuration, telomeres and centromeres of chromosomes cluster at two different poles in the nucleus, particularly in plants with larger genomes. In the Rosette configuration, the nucleolus is surrounded by telomeres, while heterochromatin and centromeres are clustered together but euchromatin oozes out freely in the nucleus to form a rosette-like configuration, observed in plants with smaller genome like Arabidopsis. Bouquet configuration is a transient chromatin configuration observed during meiosis in different organisms, including plants, wherein telomeres of the chromosomes are co-localized on a specific site of nuclear periphery, while the rest of the chromatin remains dispersed in the nuclear space. Nucleolus-associated domains (NADs) are chromatin regions that interact with the nucleolus, while the lamina-associated domains

(Continued)

FIGURE 1 | (LADs) are associated with the lamina of the nuclear envelope. Chromosome territories are further divided into (ii) A and B compartments, which correspond to euchromatic and heterochromatic regions, respectively. While the A compartment is constituted of high gene density, activating epigenetic modifications, and active transcriptional activity, the B compartment possesses lesser genes, low transcriptional activity, repressive epigenetic modifications, and higher transposon density. (iii) Topologically associated domains (TADs) are relatively independent local units/regions where chromatins interact with each other at a higher frequency than with the surrounding regions. (iv) Number of factors/modifications/readers is involved in the formation of chromatin loops that connects regulatory elements to their target loci in plants. **(B)** Lower level chromatin interactions (chromatin loops) establish regulatory networks between the distant elements through their physical proximity. The regulatory function of chromatin loops comes due to the formation of (i) heterochromatin/repressive loop by histone modifiers–H3K27me3–polycomb protein–lncRNAs, while (ii) silencing chromatin loop is formed by H3K9me2-reader (ADCP1)–ncRNAs. (iii) Different regions (5′–3′ gene looping) of the same gene, (iv) an enhancer and promoter (enhance–promoter loop) of a gene, (v) the different co-regulated genes (gene-gene loops), (vi) non-coding genomic regions (intergenic loop), and (vii) transcriptional hub/loop formed by H3K4me3 modifiers, RNA Pol-II and eRNAs.

nucleus (Parada and Misteli, 2002). Additionally, chromosomes can show different configurations including Rab1, Rosette, and Bouquet configuration (Fransz et al., 2002; Harper et al., 2004; Grob and Grossniklaus, 2017). Advances in high-throughput 3C techniques and their derivatives help decipher the chromosomal interactions and address the complicated interplay between local chromatin organization and genome functions (Sexton and Cavalli, 2015). Individual chromosomes occupy separate CTs in the nucleus during interphase, which is the top hierarchical structure in most eukaryotic genomes (Pecinka et al., 2004). Further, chromosomes can be divided into A and B compartments wherein the A compartment is associated with high gene density/active transcriptional activity; the B compartment has higher transposon density and repressive epigenetic modifications. In mammals, TADs are enriched with chromatin loops on the hundreds of kilobase (Kb) scale which link promoters and *cis*-regulatory elements to modulate gene expression by recruiting TFs (Li et al., 2012; Rao et al., 2014; Rao et al., 2017). Moreover, compartmentalization within TAD protects the promoters from making ectopic contact with distant enhancers (Szabo et al., 2019). In contrast, the plant TADs might play different roles by making regulatory contact between enhancers and promoters occurring across TAD boundaries (Dong et al., 2017; Stam et al., 2019). Similarly, other structures like frequently interacting regions (FIREs), transcriptional hubs, and repressive loops have also been observed in plants while exploring the hierarchical chromatin interactions with the help of Hi-C (Dogan and Liu, 2018; Dong et al., 2018).

2.1.1 Chromosome Territories

Chromosome territory (CT) was directly visualized by cytological and microscopy techniques (Lichter et al., 1988; Pinkel et al., 1988). During mitosis interphase, each chromosome occupies an exclusive and limited domain in the nucleus called chromosome territory (CT) (Fransz and de Jong, 2011). Initial study on genome organization in Drosophila using Hi-C indicated sub-compartmentalization of chromosome arms (Sexton et al., 2012), and each chromosome occupies a space in the nucleus (Schubert et al., 2014) (**Figure 1**). CT is further subdivided into chromosomal arm territories, and a contact matrix of chromosome arms is more intricate than the contact matrix of the chromosome. Only weaker interaction in the pericentromeric regions, while strong interactions were reported between pericentromeric heterochromatin and telomeres in Arabidopsis (Grob et al., 2014). Studies on the genome of different crop plants

(rice, maize, sorghum, tomato, and foxtail millet) revealed interactions between the adjacent loci and helped to understand the chromatin architecture (Dong et al., 2017). Intra- and inter-chromosomal interactions between the euchromatic arms resulted in the identification of CTs in all five plant species. While intense chromosomal interactions indicated frequent interactions between the chromosomal arms and centromeric regions of the chromosomes in the Hi-C map of maize, such interactions were not detected in rice and foxtail millet (Dong et al., 2020b). However, a signal for such interactions was reported in barley (Mascher et al., 2017).

In plants, chromosomes show different morphology including Rabl (in the honor of Carl Rabl), Rosette, and Bouquet configuration. In Rabl configuration, the chromosomes are folded at the centromere making a polarized separation of centromeres and telomeres (**Figure 1A**). Such configuration is observed in diverse organisms (animals, yeasts, and plants) (Huang Y. et al., 2020; Zhang and Wang, 2021). The existence of different chromosome configurations within an organism suggested its specificity in different cell types. The emergence of single-cell 3D genomics techniques would help to assess the linkage between chromatin organization and cell identity. In Rosette configuration, pericentromeric heterochromatin forms a condensed chromocenter from which euchromatic loops emerge out (Fransz et al., 2002). Traditionally, such configuration was attributed to the small genome of plant like Arabidopsis; however, having comparable genome size sorghum does not present this chromosome configuration (Muller et al., 2019). Moreover, yeasts having even smaller genomes than Arabidopsis presents Rabl configuration (Muller et al., 2019), indicating that genome size is not the determinant of chromosome configuration in the nucleus. In Bouquet conformation, telomeres cluster at the nuclear periphery while the chromatins emanate in the nucleoplasm (**Figure 1A**), which has been described in different plant species including rice, maize, and wheat during meiosis (Tiang et al., 2012; Zhang et al., 2017; Dogan and Liu, 2018; Hurel et al., 2018). Bouquet configuration appears to be a universal and transient feature of meiotic cells in plants, yeast, and animals (Huang Y. et al., 2020).

Recent studies reveal that gene expression is associated with chromatin positioning (CT) in the nucleus. Some of the chromatin domains are associated with the nucleolar periphery and named nucleolus-associated chromatin domains (NADs) (Pontvianne and Grob, 2020). In addition, the nuclear periphery is enriched with repressed chromatin associated with lamin fibers, and named as lamina-associated domains (LADs)

(Gonzalez-Sandoval and Gasser, 2016). In plants, transcriptionally inactive LADs and NADs have been detected. In Arabidopsis, the LAD-specific protein, crowded nuclei 1 (CRWN1), has been reported to interact with polycombs1 (PWO1) to mediate chromatin tethering at the nuclear periphery (Poulet et al., 2017; Hu et al., 2019; Pontvianne and Liu, 2020). However, only limited research on LADs and NADs in plants have been carried out due to inadequate knowledge of the proteins required for the formation of these chromatin domains (Pontvianne and Liu 2020; Sakamoto et al., 2020).

2.1.2 Global and Local A and B Compartments
Chromosomal compartments are formed due to the genome-wide interactions between TADs and epigenetic signatures, which have been discovered in both animals and plants while analysing Hi-C data. Two spatial compartments, namely A and B compartments, of chromosomes, have been reported. While A compartment is associated with open/active chromatin, the B compartment is associated with closed/inactive chromatin (Lieberman-Aiden et al., 2009; Dong et al., 2017) (**Figure 1A**). Apart from the A/B compartments, other compartment-like domains termed structural domains (SDs) have been reported in Arabidopsis. While the less compact euchromatin contains loose structural domains (LSDs), heterochromatin contains closed structural domains (CSDs) and is enriched in repressive epigenetic marks (Grob and Grossniklaus, 2017). The A compartment enriched with euchromatic activation histone marks, and the B compartment containing heterochromatic repressive epigenetic marks around the pericentromeric region have been reported in rice (Dong et al., 2018). Comparative analysis of rice, maize, and millet tissues using *in situ* Hi-C technique reported the existence of global A/B compartments across the tissues, while the local A/B compartment was reported to be dynamic and tissue-specific associated with differential expression of genes (Zhou et al., 2019; Dong et al., 2020a).

2.1.3 Topologically Associating Domains (TADs)
In the eukaryotic genome, TADs are the independent local/structural units and the regions of high chromatin inter-connectivity. The A/B compartments can be further segmented as TADs which are 0.1–1.0 Mb in size (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012; Rao et al., 2014). While the mammalian TADs are highly conserved in different tissues across the species (Dixon et al., 2012; Sexton and Cavalli, 2015; Vietri Rudan, et al., 2015; Grob and Grossniklaus, 2017), plant TADs are not conserved across the species (Dong et al., 2017). TADs are contiguous regions with more frequent chromatin interactions within the region than those with the other region in mammalian genomes (Dixon et al., 2012) (**Figure 1A**). TADs allow long-range chromatin interaction for target specificity of the remote *cis*-regulatory elements in plant and the human genome (Jin et al., 2013; Rao et al., 2014; Dogan and Liu, 2018). TADs are not reported in Arabidopsis because of the small genome size, as prominent TADs could not be detected in the species having smaller (<400 Mb) genomes (Dong et al., 2017; Stam et al., 2019). However, the effect of genome size on TAD formation is still under debate (Zhang and Wang, 2021). It is also speculated that

TADs are displayed in plants having lower gene density/larger genome size (Dogan and Liu, 2018).

Animals TAD boundaries are reported to be bound by the insulator protein CCCTC-binding factor (CTCF) and specific epigenetic marks (Dixon et al., 2012; Sexton et al., 2012; Rao et al., 2014; Tang et al., 2015), which affect chromatin functions and transcriptional activity (Eser et al., 2017) through promoter-enhancer interactions. In plants, CTCF homologue has not been identified, which indicate that it might not be required for the formation of TAD boundary (Pontvianne and Grob, 2020). Growing evidence suggests that cohesin couple with CTCF in TAD establishment in mammals (Fudenberg et al., 2016; Rao et al., 2017; Nuebler et al., 2018). Cohesins are conserved between animals and plants (Zhang et al., 2004); however, the cohesins have similar functions in plants is still not clear (Ouyang et al., 2020). It would be interesting to identify CTCF-like insulator proteins in plants involved in the formation of TAD boundaries.

In embryonic stem cells of mice, high-resolution (allele-specific 4C) mapping indicated that TAD is constituted of metaTADs and subTADs, which are dynamic to form active and inactive nuclear compartments (Wijchers et al., 2016). Wijchers et al. (2016) suggested that *trans*-associated factors (SUV39H1, or EZH2) influence 3D compartmentalization independent of their *cis*-effect on local chromatin composition and activity. In Arabidopsis, several local structural features like positive strips which interact frequently with the neighboring chromatin were observed (Wang et al., 2015). Such positive strips enriched with repressive histone marks like H3K27me3 were reported in Arabidopsis (Liu Z. et al., 2016). In plants, TAD-like domains lack co-expression behavior and do not possess a conserved biological function as observed in mammals (Dekker and Heard, 2015). Moreover, the TADs rich in GC motifs and positively correlated with transcriptional activation/gene expression were observed in rice and cotton (Liu et al., 2017).

Similarly, TAD-like domains enriched in and associated with highly expressed genes were observed in maize, tomato, foxtail millet, and sorghum (Dong et al., 2017). Hi-C analysis of diploid and tetraploid cotton suggested the existence of intra-chromosomal interactions and TAD-like regions (Wang M. et al., 2018). In rice, TADs showed increased sequence variation and meiotic recombination compared to that observed in the inter-TAD regions (Golicz et al., 2020). In wheat, the existence of TAD-like structures (termed as intergenic condensed spacers, ICONS) was reported (Concia et al., 2020). Therefore, it would be interesting to examine whether the occurrence of TADs/TAD-like structures is linked with larger genome size. In general, these observations support the hypothesis that plant genomes are packaged into TAD-like structures by yet to be identified molecular mechanism(s). The *cis*-regulatory elements of target genes form chromatin loops to control gene expression.

2.1.4 Chromatin Loops
Another level of hierarchical genome organization that plays important role in transcriptional regulation of gene expression

is the chromatin loop. The chromatin loops are formed due to physical interaction between *cis*-acting elements and the genes that are brought into close spatial proximity, which are vital for gene regulation (Bonev and Cavalli, 2016) (**Figure 1B**). In plants, chromatin loops are formed between distal regulatory elements and promoters to exert their function by providing the opportunity for enhancers to contact with their genes located at tens of kilobase-pair away (Dogan and Liu, 2018; Li E. et al., 2019). In maize, the first chromatin loop was observed between the promoter and regulatory sequences at *b1* locus (Louwers et al., 2009), and it was reported between 5′ and 3′ flanking regions of flowering locus C (FLC) in Arabidopsis (Crevillén et al., 2013). Moreover, chromatin loops of varying sizes (Kb to Mb, small as well as large loops) have been reported in the eukaryotic genome (Rao et al., 2014). Generally, transcription start site loops with the downstream region and transcription termination site loops with the upstream region. The formation of such loops enhances promoter−enhancer interaction to initiate the transcription process. Loop structure has also been reported in the formation of rosette-like structure in the heavy chain of immunoglobulin that is required for V(D)J recombination (Ebert et al., 2015). *In situ* Hi-C analysis revealed extensive chromatin loops in the regions enriched with epigenetic marks and active genes in the larger genomes like maize and tomato, while such loops are absent in the smaller genome (Dong et al., 2017). Spatial organization of the regulatory elements revealed by the construction of high-resolution chromatin interaction maps in maize deciphered the role of chromatin loops in gene expression (Peng et al., 2019). The active and repressed chromatin regions are separated from each other in animals, and some compartments in the nucleus, like nuclear and nucleolar periphery, are enriched with repressed chromatin (van Steensel and Belmont, 2017; Bersaglieri and Santoro, 2019).

Histone modifiers (e.g., H3K9me reader ADCP1) and non-coding RNAs (ncRNAs) are involved in the formations of repressive chromatin loop [**Figure 1B**(i)], while H3K27me3 and long non-coding RNAs (lncRNAs) are associated with silenced chromatin loop formation [**Figure 1B** (ii)] (Stam et al., 2019; Kantidze and Razin, 2020). The 5′−3′ loop bring together the 5′ and 3′ termini of the same gene [**Figure 1B** (iii)], an enhancer-promoter loop occurs between the promoter and distant enhancer of a gene [**Figure 1B** (iv)], gene—loop is formed between different co-regulated genes [**Figure 1B** (v)], an intergenic loop is comprised of the intergenic region (Rodriguez-Granados et al., 2016; Huang Y. et al., 2020) [**Figure 1B** (vi)], and transcriptional hub is formed by certain activation histone marks/modifiers (e.g., readers, writers, mediators of H3K4me3), RNA Pol-II, and RNAs (Ouyang et al., 2020) [**Figure 1B** (vii)].

RNAs and multivalent proteins play vital roles in the formation of chromatin loops. Enhancer RNAs (eRNAs) combined with mediator and RNA Pol-II were reported to promote the formation of enhancer-promoter loops to modulate transcriptional activity of the target genes in human cell lines (Lai et al., 2013; Pefanis et al., 2015) [**Figure 1B** (vii)]. In Arabidopsis, a Mediator subunit (MED25) was reported to affect the dynamics of chromatin looping between the promoter and

enhancer to trigger transcriptional programming in the jasmonic acid signaling pathway (Wang et al., 2019). Activating histone marks (H3K4me3 and H3K36me3) and histone variants (H2A.Z, and H2Bub1) were observed at the *FLC* locus which is bound by histone modifiers like H3K4 methyltransferase and H3K36 methyltransferase (Li Z. et al., 2018). Binding of these 'writers' to *FLC* results in the formation of 5′-to-3′ loop.

3 3D GENOME MAPPING TECHNIQUES

The advances/improvements in bioimaging and biochemical methods over the last few decades have unveiled 3D genome architectures in animals and plants at a rapid speed (Sexton and Cavalli, 2015; Ouyang et al., 2020). The 3D genome mapping approaches can be broadly divided into two categories. The first category of approaches is based on cytology/microscopy, which utilizes fluorescent dye to label DNA/chromatin and/or visualization of the spatial chromatin organization using a microscope (Probst, 2018). Combining microscopy with fluorescent *in situ* hybridization (FISH) boosted the progress in understanding how the spatial organization of CTs affects gene expression within the nucleus (Zhang and Wang, 2021). The second group of approaches utilizes next-generation sequencing (NGS), and they could be ligation-based or ligation-free (**Figure 2**). Each of these techniques has certain advantage over the other, and there is some limitation in using then individually in 3D genome analysis (**Table 1**). However, a combination of techniques provides better opportunity for improved specificity, sensitivity, and ultrahigh-resolution analysis.

The cytology/microscopy-based approaches can be further divided into either microscopy or labeling-based techniques. For single-cell genome organization study, microscopy is indispensable. Although fluorescence microscopy has enabled us to investigate larger chromosome organizations of micron length scale, the smaller structures remained invisible due to the limited spatial resolution of fluorescence microscopy (Lakadamyali and Cosma, 2015). However, the advent of super-resolution microscopy (SRM) enables us to investigate nano-scale chromosome organizations *in vivo*. Such SRM methods have the potential to enhance our knowledge of chromatin structure–function relationship. Microscopy-based techniques include confocal microscopy (CFM) (Carlsson et al., 1989), wide-field microscopy (WFM) (Wheeler and Tyler, 2011), chrom-electron microscopy tomography (ChromEMT) *in situ* visualization of chromatin using a fluorescent dye that stains DNA with an osmiophilic polymer with selectively enhances the contrast in electron microscopy (Ou et al., 2017), Hi-M (a multiplexed, sequential imaging approach) simultaneously reveals 3D chromatin organization and transcriptional activity; thus, enables detecting the spatial organization of cells and measurement of the changes in TAD organization during early embryogenesis and upon transcriptional activation (Cardozo Gizzi et al., 2019), SRM (Schubert, 2017), electron microscopy imaging (EMI) (Lobastov et al., 2005), and light-sheet fluorescence
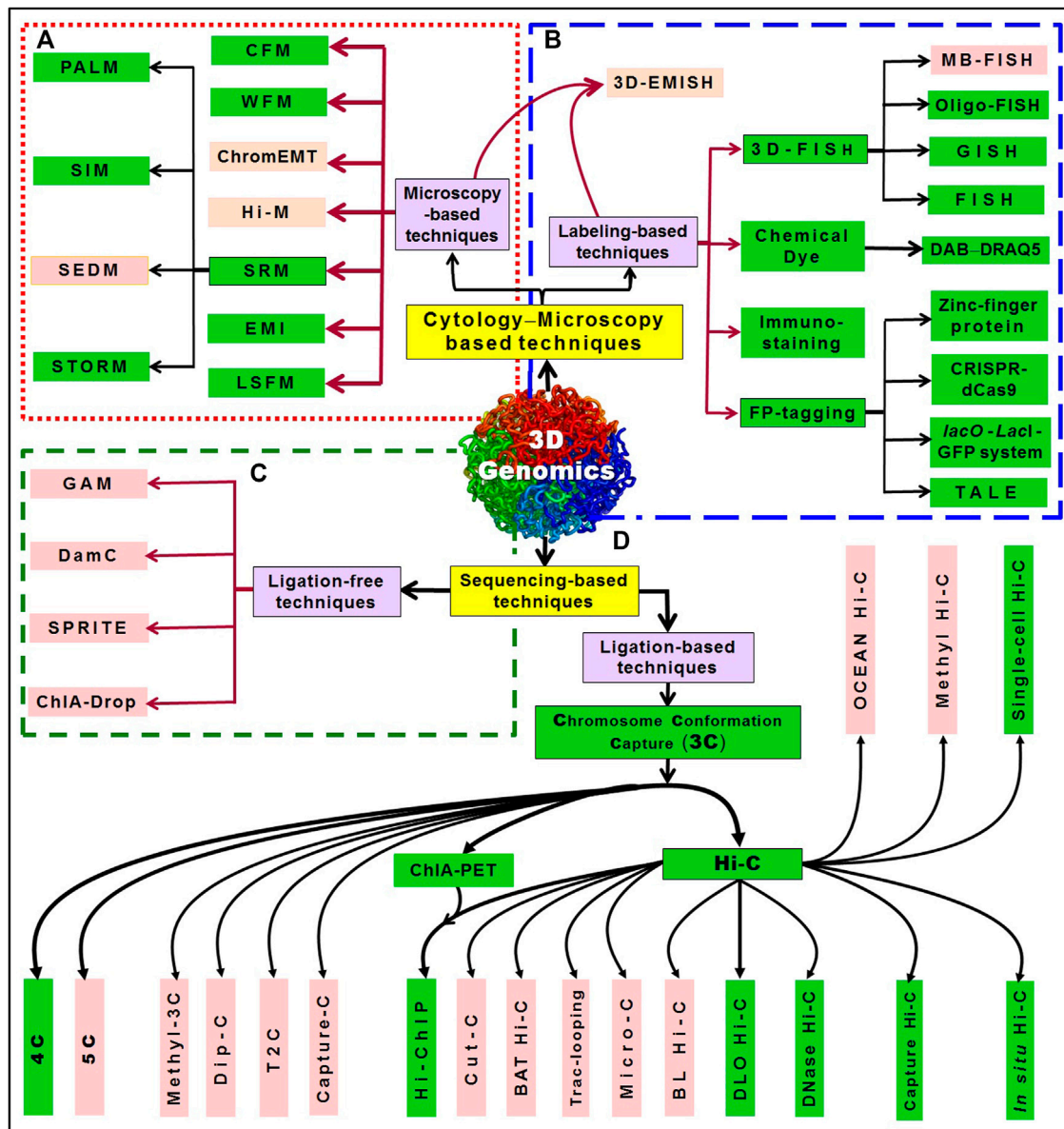
FIGURE 2 | An overview of 3D genomics techniques. The techniques can be broadly divided into two categories: one is based on cytological/microscopic examination/imaging; another is based on sequencing. **(A)** The microscopy-based techniques include WFM (wide-field microscopy), CFM (confocal fluorescence microscopy), ChromEMT (Chrom-electron microscopy tomography), Hi-M (multiplexed, sequential imaging approach) simultaneously reveals 3D chromatin organization and transcriptional activity, SRM (super-resolution microscopy), which include SIM (structured illumination microscopy), SEDM (stimulated emission depletion microscopy), PALM (photoactivated localization microscopy), STORM (stochastic optical reconstruction microscopy), EMI (electron microscopy imaging), and LSFM (light-sheet fluorescence microscopy). **(B)** While 3D-EMISH (three-dimensional electron microscopy with *in situ* hybridization) utilizes the advantages of both microscopy (electron microscopy) and labeling (*in situ* hybridization), the labeling-based techniques include 3D-FISH [such as fluorescence *in situ* hybridization, MB-FISH (molecular beacon-FISH), Oligo-FISH (oligonucleotides probe-based FISH), GISH (genomic *in situ* hybridization)], staining with chemical dyes like DAB–DRAQ5 system, immune-staining, and FP-tagging (fluorescent protein-tagging) including tagging with zinc-finger protein, *lacO*-LacI-GFP system, CRISPR-dCas9 (clustered regularly interspaced short palindromic repeats-nuclease-deficient Cas9), TALE (transcription activator-like effectors with a quantum dot labeling technique). The sequencing-based techniques can be ligation-free or those which require proximity ligation (3C, chromosome conformation capture). **(C)** While ligation-free techniques include GAM (genome architecture mapping) that combines micro-cutting and sequencing, SPRITE (split-pool recognition of interactions by tag extension), ChIA-Drop (chromatin interaction analysis *via* droplet-based and barcode-linked sequencing), DamC (DNA adenine methyltransferase-based chromosomal contacts), **(D)** ligation-based techniques include the advancements in 3C (chromosome conformation capture), like 4C, 5C (chromosome conformation capture carbon copy), methyl-3C (combination of DNA methylation detection and 3C technology), Dip-C (combination of single-cell 3C and transposon-based whole-genome amplification method), T2C (targeted Chromatin Capture), Capture 3C (combination of 3C with oligonucleotide capture technology. Further advancements like ChIA-PET (chromatin interaction analysis by paired-end tag sequencing) and Hi-C (high-throughput chromosome conformation capture), and their combination Hi-ChIP (chromatin conformation method that combines Hi-C with ChIA-PET technology) have advanced the 3D genome architectures. Combination of techniques like Cut-C (antibody-mediated cleavage by

(Continued)

microscopy (LSFM) (Santi, 2011). SRM depends on photoactivated localization microscopy (PALM) (Rust et al., 2006), structured illumination microscopy (SIM) (Fitzgibbon et al., 2010), stimulated emission depletion microscopy (SEDM) (Dyba et al., 2003), stochastic optical reconstruction microscopy (STORM) (Betzig et al., 2006) (**Figure 2A**). Intensive research on microscopic visualization of chromosome organization using labeling-based techniques like FISH has greatly improved the sensitivity, specificity and resolution (Cui et al., 2016). Combining FISH and super-resolution microscopy further boost the detailed characterization of structural chromatin domains (Boettiger et al., 2016). 3D-EMISH (combines serial block-face scanning electron microscopy with *in situ* hybridization) visualizes 3D chromatin folding at targeted genomic regions with ultrahigh-resolution (5 nm × 5 nm × 30 nm) (Trzaskoma et al., 2020). Depending on the type of label/dye used, the labeling techniques are divided into four categories: i) 3D-FISH, which include fluorescence *in situ* hybridization (FISH) (Solovei et al., 2002; Koornneef et al., 2003; Berr and Schubert, 2007; Cremer et al., 2008); Oligo-FISH, uses oligonucleotide probes (Beliveau et al., 2012; Beliveau et al., 2012); MB-FISH, uses molecular beacon probes (Wu et al., 2010; Ni et al., 2017); GISH, genomic *in situ* hybridization (Schubert et al., 2012), ii) staining with chemical dyes, e.g., DAB–DRAQ5 system, (Ou et al., 2017; Poulet et al., 2017), iii) immuno-staining (Fransz et al., 2002; She et al., 2013), and iv) fluorescent protein-tagging (FP-tagging) (Matzke et al., 2005; Lindhout et al., 2007; Deng et al., 2015; Ren et al., 2017; Nagaki and Yamaji, 2020) which include tagging with zinc-finger proteins, clustered regularly interspaced short palindromic repeats−nuclease-deficient Cas9 (CRISPR-dCas9) (Dreissig et al., 2017; Hong et al., 2018), *lacO*-LacI-GFP system (Ding and Hiraoka, 2017), and transcription activator-like effectors (TALE) coupled with quantum dot labelling technique (Ma et al., 2017). Chromatin domain can be labeled using dCRISPR-Cas9 reporter proteins (guided by sgRNA) or green fluorescent protein-tagged m6A-tracer protein, which allow tracking the location of chromatin domain in the nucleus (Qin et al., 2017; Ye et al., 2017; Hong et al., 2018; Wong et al., 2021). The *lacO*/LacI-GFP system provides a simple and useful method to visualize a chromosome locus by inserting *lacO* repeat arrays and expressing LacI–GFP fusion protein that binds to the *lacO* (Ding and Hiraoka, 2017). A novel bimolecular fluorescence complementation (BIFC) method, that combines the advantages of both dCas9-labeling and gRNA-labeling, enables live cell imaging with high signal-to-noise ratios without non-specific foci (Hong et al., 2018) (**Figure 2B**).

Some of the sequencing-based techniques that do not require proximity-ligation have also been devised for 3D genomics

studies, which include genome architecture mapping (GAM) (Beagrie et al., 2017), split-pool recognition of interactions by tag extension (SPRITE) (Quinodoz et al., 2018), DNA adenine methyltransferase identification of chromosomal interactions (DamC) (Redolfi et al., 2019), and chromatin interaction analysis *via* droplet-based and barcode-linked sequencing (ChIA-Drop) (Zheng et al., 2019) (**Figure 2C**). While GAM utilizes micro-sectioning and sequencing to decipher the relative location of genes and enhancers for studying the frequency of genomic interactions in nuclear sections (Beagrie et al., 2017), SPRITE detects pairwise interactions between two loci as well as DNA−RNA interactions (Quinodoz et al., 2018), DamC detects distal chromatin interaction along with the methylation status wherein DNA adenine methyltransferase and DNA-binding proteins are recruited to specific genomic locations (Redolfi et al., 2019). On the other hand, ChIA-Drop uses a specific antibody to capture the target protein and interacting DNA by ChIP for multiplex chromatin-interaction analysis adopting microfluidics to produce gel-bead-in-emulsion droplets (Zheng et al., 2019).

Another group of techniques that use proximity-ligation and NGS includes chromosome conformation capture (3C)-based approaches (Dekker et al., 2002) as well as its derivatives such as chromosome conformation capture-on-chip (4C) (Simonis et al., 2006), chromosome conformation capture carbon copy (5C) (Dostie et al., 2006) (**Figure 2D**). While 3C-based techniques rely on enzymatic digestion of DNA and proximity ligation to capture long-range chromatin interaction between two specific genomic loci (Dekker et al., 2002), 4C is used to visualize the interaction between a site of interest and other sites on the genome (Simonis et al., 2006). Moreover, 5C is used to analyze the chromatin interactions between multiple genomic loci (Dostie et al., 2006). A combination of DNA methylation detection and 3C technology (methyl-3C) (Lee et al., 2019), a combination of single-cell 3C and transposon-based whole-genome amplification (Dip-C) (Tan et al., 2018), targeted chromatin capture (T2C) studies chromatin organization for specific genomic regions (Kolovos et al., 2014), while capture-C combines 3C with oligonucleotide capture technology (Hughes et al., 2014). Further developments in 3C sequencing technologies resulted in Hi-C and ChIA-PET which have been quite helpful in 3D genomic studies (Fullwood et al., 2009; Lieberman-Aiden et al., 2009). A protein-centric chromatin conformation method that combines Hi-C with ChIA-PET technology (Hi-ChIP) (Mumbach et al., 2016) has improved our understanding of 3D genome architectures. Hi-ChIP improves the informative reads by over 10-fold and lowers input requirement by over 100-fold compared to ChIA-PET. Being an efficient and sensitive analysis of protein-directed genome architecture, Hi-ChIP for

TABLE 1 | Characteristic features of different techniques used for 3D genome organization analysis.

Technique	3D genomics approach	Advantage	Limitation	References
Microscopy-based techniques	Visualize chromatin conformation by cytological and microscopy, indispensable for single-cell genome organization studies	Cytological expertise can be exploited for more efficient analysis, may simultaneously analyze 3D chromatin organization and transcriptional activity	Limited resolution of the traditional microscopic technique needs to combine with other technique to improve the resolution	Lobastov et al. (2005); Wheeler and Tyler (2011); Cardozo Gizzi et al. (2019)
Labeling-based techniques	Label DNA/chromatin to visualize the spatial chromatin organization	Improve sensitivity, specificity and resolution; enable the possibility for live-cell imaging	Repetitive sequence required for easy visualization; stringent preparation/protocol	Schubert et al. (2001); Matzke et al. (2005); Wu et al. (2010); Ma et al. (2017); Nagaki and Yamaji (2020)
Ligation-free techniques	Do not require proximity-ligation but use sequencing technologies for in-depth chromosomal interaction analysis	Can detect distal chromatin interaction along with the methylation status, also detect DNA–RNA interactions	Pairwise interaction between two loci	Beagrie et al. (2017); Quinodoz et al. (2018); Redolfi et al. (2019); Zheng et al. (2019)
Chromosome Conformation Capture (3C)	Rely on enzymatic digestion of DNA followed by proximity ligation to capture long-range chromatin interaction between two specific genomic loci	Captures long-range chromatin interaction between two specific genomic loci	Low throughput coverage, provides chromatin configuration of population average, presents one-to-one interaction	Dekker et al. (2002)
Chromosome Conformation Capture-on-Chip (4C)	Circular chromosome conformation capture approach, detected by inverse-PCR using the primers for candidate gene	Studies the interaction between a chromatin site of interest and the other sites on whole genome	Less efficient to study the interactions of shorter distance (<50 Kb); reveals one-to-many interactions	Simonis et al. (2006); Zhao et al. (2006); Grob and Cavalli (2018)
Chromosome Conformation Capture Carbon Copy (5C)	Analyses the interactions with in a limited region like gene clusters, templates originating from the region of interest are PCR amplified and quantified using NGS approach	Used for chromatin interaction analysis between multiple genomic loci (many-to-many interactions); bioinformatics play important role in the analysis	Suitable for interaction studies on relatively smaller genomes only	Dostie et al. (2006); Sati and Cavalli (2017)
ChIA-PET	DNA–protein complex is cross-linked, fragmented by ultrasonication, and captured by the protein-specific antibody (ChIP) which is analysed by high-throughput sequencing	Efficient analysis of long-range chromatin contacts bound by a protein and provides a high-resolution map of chromatin interactions with considerably fewer sequencing reads	Captures the distal interactions where specific proteins are involved; hromatin configuration of population average; reveals many-to-many interactions	Fullwood et al., 2009; Li et al., 2010; Li et al., 2017
Hi-C	Relies on restriction enzyme to break the chromatin into smaller fragments, uses NGS approach to investigate both short- and long-range chromatin interactions at whole-genome level	Detects "all-to-all" interactions	May not be appropriate for the study of individual locus; generates abundant unenriched chromatin contact data	Lieberman-Aiden et al. (2009); Feng et al. (2014); Wang et al. (2017); Xie et al. (2019)
Hi-ChIP	A protein-centric chromatin conformation method, Hi-C is combined with ChIA-PET.	Ten-fold more informative reads with 100 times lesser input requirement; generates significantly better signal-to-noise ratio	The protein-specific antibody is required to capture the DNA–protein interactions; produces unenriched chromatin contact data	Mumbach et al. (2016); Ricci et al. (2019); Concia et al. (2020)
BAT-Hi-C	Combines AluI restriction with biotinylated linker-mediated proximity ligation analysis	Ideal for genome-wide in-depth analysis of long-range chromatin looping; economical and straightforward technique	Chromatin configuration of population average; need optimization for plant studies	Huang et al. (2020a)
Capture Hi-C	Combines Hi-C and hybridization-based capture of targeted genomic regions	Specific probes are used to capture the reads related to the target region, and chromatin interactions of the region are deciphered by NGS.	Chromatin configuration of population average	Mifsud et al. (2015)
In situ Hi-C	The intact nuclei are used, instead of free chromosomes, for ligation	Use of complete nuclei reduces wrong ligation of DNA fragments from different nuclei, effectively reduce the background noise, and improving the signal-to-noise ratio	Chromatin configuration of population average	Rao et al. (2014); Liu et al. (2017)
Methyl Hi-C	Combination of Hi-C and DNA methylation detection technology	Simultaneous captures the chromosome conformation and DNA methylation	Chromatin configuration of population average	Li et al. (2019)

(Continued on following page)

Technique	3D genomics approach	Advantage	Limitation	References
Single-cell Hi-C	Chromatin conformation of individual cell is captured and studied using Hi-C at single-cell level	Chromatin conformation of an individual cell is captured; avoids averaging of chromatin maps for a population of cells	Still in its infancy for plant studies	Nagano et al. (2013); Stevens et al. (2017); Zhou et al. (2019); Sun et al. (2021)

cohesin reveals multi-scale genome architecture with greater signal to the background than *in situ* Hi-C (Mumbach et al., 2016). Capture Hi-C combines Hi-C and hybridization-based capture of targeted genomic regions (Mifsud et al., 2015), *in situ* Hi-C is performed in the intact nuclei with DNA–DNA proximity-ligation (Rao et al., 2014), micro-C uses micrococcal nuclease for chromatin fragmentation (Hsieh et al., 2015), while DNase Hi-C uses DNase I enzyme to fragment the chromatin (Ma et al., 2015) for chromatin architecture analysis. Similarly, single-cell Hi-C (Hi-C analysis of nucleus from a single-cell, Nagano et al., 2013) bridges the gap between genomics and microscopy studies of chromosome structure, the *B*ridge *L*inker Hi-C (BL Hi-C) combines restriction enzyme (RE) targeting and two-step proximity ligation (Liang et al., 2017), while in *D*igestion-*L*igation-*O*nly Hi-C (DLO Hi-C) digestion and ligation are performed twice without biotin labeling and pulldown (Lin et al., 2018). The recently developed techniques like Trac-looping (transposase-mediated analysis of chromatin looping) for simultaneous detection of multiscale genome-wide chromatin interactions among regulatory elements and chromatin accessibility (Lai et al., 2018), and OCEAN Hi-C (open chromatin enrichment and network Hi-C) for antibody-independent mapping of global open chromatin interactions (Li T. et al., 2018) were used to decipher the chromatin architecture. More recently, Cut-C combined antibody-mediated cleavage by tethered nuclease with chromosome conformation capture to identify chromatin interactions mediated by a protein of interest (Shimbo et al., 2019). Cut-C identifies protein-centric chromatin conformations along with the genome-wide distribution of target proteins using a simple procedure. Applying Cut-C to a histone modification (H3K4me3) enriched at active gene promoters, Shimbo et al. (2019) could successfully identify the chromatin loops mediated by H3K4me3 along with the genome-wide distribution of H3K4me3. Further, methyl Hi-C (DNA methylation detection combined with Hi-C) for simultaneous capture of chromosome conformation and DNA methylome was used to delineate the DNA methylation profile and chromatin architecture of a cell (Li G. et al., 2019). A simple technique for economical but efficient analysis of chromatin conformational features in mouse embryonic stem cells, BAT Hi-C (*B*ridge linker-*A*luI-*T*n5 Hi-C), was developed by combining *A*luI restriction with biotinylated linker-mediated proximity ligation (Huang J. et al., 2020). With just one-third sequencing depth, BAT Hi-C could reveal the same spectrum of chromatin contacts as *in situ* Hi-C. Being an economical and straightforward technique, BAT Hi-C is ideal for genome-wide in-depth analysis of long-range chromatin looping (Huang J. et al., 2020). While many of these techniques have been successfully used in the animal system, some of them need to be optimized in plants as efficient/economical and simple techniques (**Figure 2**).

3.1 Chromosome Conformation Capture (3C) and Its Derivatives

The 3D genomic techniques have considerably advanced over the last decade. However, their efficiency in plant 3D genomic studies is comparatively less probably because of the cell wall. Some of the derivative techniques have successfully been used in plants for 3D genomic studies. The basic 3C technique allows "one-to-one" chromosomal interactions between two loci in the genome utilizing microarrays. Hence, the 4C technique was developed as a "one to all" strategy, which allows genome-wide screening for the interactions between one specific locus with all other loci in the genome (Zhao et al., 2006) using NGS to determine long-range chromatin interactions (Splinter et al., 2012). Since 4C is suitable for long-range interaction studies (Grob and Cavalli, 2018), the 5C technique was developed for the detection of "many to many" chromosomal interactions among thousands of selected genomic loci in a single run (Dostie et al., 2006; Simonis et al., 2007) (**Figure 3**).

Later, the Hi-C technique was devised for detecting "all-to-all" interactions between any locus with all other chromosomal loci with far-reaching impacts (Lieberman-Aiden et al., 2009). Hi-C uses high-throughput sequencing to investigate both short- and long-range chromatin interactions at the whole-genome level (Lieberman-Aiden et al., 2009). Hence, Hi-C has been extensively used for characterizing chromosomal architecture in plant species like Arabidopsis, rice, tomato, Brassica, cotton, foxtail millet, sorghum, and maize (Feng et al., 2014; Grob et al., 2014; Wang et al., 2015; Wang et al., 2017; Dong et al., 2018; Grob and Cavalli, 2018; Sotelo-Silveira et al., 2018; Wang M. et al., 2018; Ting et al., 2019; Xie et al., 2019) (**Figure 3**). Subsequently, a derivative of the Hi-C technique like *in situ* Hi-C was used in rice, sorghum, tomato, and Foxtail millet (Dong et al., 2017; Liu et al., 2017). Other modifications of Hi-C such as Capture Hi-C used in Arabidopsis, digestion-ligation-only Hi-C (DLO Hi-C) in maize (Sun Y. et al., 2020; Nutzmann et al., 2020), and single-cell Hi-C (without biotin purification and pull-down) was used in rice (Zhou et al., 2019).

3.1.1 Chromatin Interaction Analysis by Paired-End Tag (ChIA-PET) Sequencing

Chromatin interaction analysis by paired-end tag (ChIA-PET) sequencing combines chromatin immunoprecipitation (ChIP) with 3C-type analysis for comprehensive and efficient analysis of long-range chromatin contacts bound by a protein like

FIGURE 3 | An overview of the chromatin conformation capture (3C) and its derivative techniques used for 3D genomic studies. The DNA–protein interactions are fixed *in vivo* using formaldehyde and then chromatins are fragmented by restriction endonuclease treatment. The cross-linked chromatins are processed differentially for one vs one (3C), one vs many (4C), many vs many (5C, ChIA-PET), or all vs all (Hi-C) interaction analysis. The techniques successfully used in plants are presented in green boxes.

promoters at lower-kilobase resolution (Li et al., 2010; Li et al., 2014). In ChIA-PET, the DNA–protein complex is cross-linked, fragmented by ultrasonication, and captured by the protein-specific antibody. The captured chromatin is attached with a biotin-labeled oligonucleotide linker having a *Mme*I restriction site. The adjacent linkers are connected and *Mme*I restriction enzyme is used to digest the linker to obtain DNA fragments having paired-end tags (PETs). Then PETs are used for high-throughput sequencing (Fullwood et al., 2009). ChIA-PET includes chromatin immunoprecipitation (ChIP) for the enrichment of chromatin interactions, which provides functional specificity and efficiency along with a higher resolution for the detection of chromatin interactions. ChIA-PET generates enriched data for chromatin interaction utilizing the antibody specific to the protein that mediated interactions; hence, it provides a high-resolution map of chromatin interactions with considerably fewer sequencing reads. ChIA-PET also provides abundant unenriched chromatin contact data

(similar to that generated in Hi-C) which helps in the plotting of high-order neighborhood/topological proximity. Thus, ChIA-PET provides three different types of genomic datasets for 3D genome analysis: i) the protein binding sites, ii) the enriched chromatin interactions between the binding sites, and iii) unenriched chromatin interactions. A modification in ChIA-PET for long-read was reported with the help of longer paired-end-tags (up to 2 bp × 250 bp) (Li et al., 2017). The longer PET reads improve the mapping efficiency and increase the probability of covering phased single nucleotide polymorphism to enable the identification of haplotype-specific chromatin interactions. While Hi-C is used to identify the spatial/3D proximity (distal interactions) at the genome level (Dixon et al., 2012), ChIA-PET captures the distal interactions involving specific proteins in the genome (Fullwood et al., 2009).

However, depending on the scientific needs, several modifications in Hi-C and ChIA-PET have been adopted. Recent studies using ChIA-PET unraveled chromatin

interactions associated with gene expression in maize and rice (Li E. et al., 2019; Peng et al., 2019; Zhao et al., 2019). Maize ChIA-PET studies on chromatin domains with H3K4me3, H3K27ac, and RNA Pol-II identified the network of promoter-enhancer and promoter-promoter interactions in maize. Likewise, the ChIA-PET study on rice revealed the physical interactions between many expression quantitative trait loci (QTL) and target genes (Zhao et al., 2019). These studies present the benefits of identifying/annotating the functional/regulatory chromatin regions/architectures by combining one-dimensional (e.g., epigenetic marks) and 3D genomic (chromatin-chromatin interaction) features.

Hi-C combined with ChIA-PET (chromatin immunoprecipitation) known as Hi-ChIP (Mumbach et al., 2016) was successfully used in maize and wheat (Ricci et al., 2019; Concia et al., 2020). After using biotin to fill in the ends and ligation, the target protein-specific antibody is used to precipitate the DNA–protein complex. Once the specific fragment containing biotin is captured, a transposase-mediated library construction method is used to finally obtain the chromatin conformation bound by the protein of interest. Hi-ChIP requires a very small amount of tissue compared to that required for Hi-C, the signal-to-noise ratio is significantly better, and more informative reads are obtained compared to that obtained from ChIA-PET

4 SINGLE-CELL 3D GENOMICS

Chromatin configuration generally varies in different tissues and cells with changing environmental factors. The chromatin architecture and variations observed by Hi-C/ChIA-PET indicate the population average of cells. Therefore, the chromatin conformation of an individual cell should be captured and studied at the single-cell level. The difference in the 3D genome architecture of cells could be detected by the single-cell 3D genome mapping technique (single-cell Hi-C) (Nagano et al., 2017). Advances in ultra-high resolution microscopy, cytology, and Hi-C provide opportunities to study 3D genome structure at the single-cell level (Wang et al., 2016; Szabo et al., 2018; Sun et al., 2021).

Single-cell 3D genome mapping of mammalian cells demonstrated variation in TADs in different cells, whereas the chromatin compartments and lamina-associated domains remained stable (Stevens et al., 2017). Single-cell chromatin conformation was captured using the Dip-C method by Tan et al. (2018). They could demonstrate the 3D genome architecture of a diploid human lymphoblastoid and a primary blood cell at higher resolution. The cell-specific chromatin organizations like Rabl configuration in mouse embryonic stem cell and Rosette configuration in M/G1phase of the human lymphoblastoid cell line was discovered. Sun et al. (2021) used Hi-C to study 3D chromatin structures in Drosophila cells at different stages of embryogenesis. They observed TAD-like structures in >50% of pre-midblastula transition cells with boundaries at varying locations, while no detectable TAD structure could be observed in the corresponding population Hi-C maps.

Although the single-cell 3D genomic study in plants is still in its infancy, it has been performed successfully in rice wherein rice single cell was isolated manually for investigations on chromatin architecture and dynamics during fertilization (Zhou et al., 2019). The study also deciphered the characteristics of chromatin compartments and telomere/centromere at the single-cell level which are distinct from those of mammalian cells (Zhou et al., 2019). Hence, single-cell 3D genomic methods should be further developed and utilized to capture the modulation in chromatin conformation to understand the transcriptional regulation of gene expression at the single-cell level. Application of single-cell 3D genomic analysis in plants would enable a better understanding of the role of chromatin architecture in epigenetic regulation of growth and developmental processes at the cellular (egg, sperm, zygote or a mesophyll cell) level, avoiding the ensemble averaging of folded DNA/chromatin maps prepared for a population of cells.

5 MODULATION IN 3D GENOME ARCHITECTURE

Hierarchical 3D genome organization is observed in yeast, animals, and plants. Higher-order chromatin architectures like CTs and chromatin compartments are fairly conserved among the cell types, tissues, and species (Zheng and Xie, 2019). However, complex modulations in TADs have been observed under environmental changes. TADs were reported to get reorganized rapidly through relocalization of structural proteins from borders of TADs to the interiors in Drosophila in response to heat stress (Li et al., 2015). However, heat shock to human K562 and Drosophila S2 cells caused dramatic transcriptional alterations but no major change in global chromatin architecture was observed (Ray et al., 2019). Therefore, it is necessary to investigate the effects of alterations in chromatin structure on gene expression under varying environmental conditions.

Similar to the compartments and sub-compartments observed in animals (Rao et al., 2014; Rowley et al., 2017), the large global compartment in plants can also be divided into local sub-compartments like heterochromatin, euchromatin, and polycomb (Dong et al., 2017; Liu et al., 2017). The TAD-like domains identified in rice, sorghum, maize, foxtail millet, and tomato could be further divided into four sub-compartments depending on their epigenetic signatures, which include active domain (open chromatin), silenced domain (DNA methylation), Polycomb-repressive domain (H3K27me3 marks), and intermediate type (no specific feature) (Dong et al., 2017). The chromatin-interacting domains (CIDs) identified in rice through long-read ChIA-PET have also been divided into four groups including H3K9me2-associated heterochromatic interacting domains (HIDs), H3K4me3-related active interacting domains (AIDs), RNA polymerase II (RNA-Pol-II)-mediated transcriptional interacting domains (TIDs), and H3K4me3-H3K9me2 mixed interacting domains (Zhao et al., 2019). Similarly, the CIDs identified by the ChIA-PET study possessed distinct genomic features. The AID and TID showed

relatively higher expressed gene and active histone mark densities, lower DNA methylation levels, and higher transcription levels. On the contrary, the HID showed the opposite genomic properties. More than half of the TAD-like domains aligned with multiple CIDs, which suggest that the TAD-like domain is a comparatively larger structural unit containing various CIDs. The chromatin regions with similar epigenetic features tether together to form higher-ordered structural units having specific functional consequences (Ouyang et al., 2020).

Plants must perceive and respond to various environmental cues including light, temperature, nutrient status, abiotic and biotic stresses (Kaiserli et al., 2018; Kumar et al., 2021). In response to illumination, the light-inducible loci in Arabidopsis were reported to rapidly change their position from the interior to the periphery of the nucleus (Feng et al., 2014). Such light-induced reorganization of the genome was reported to be associated with transcriptional activation of gene expression. Effects of light on the size of the nucleus, chromatin accessibility, and chromatin organization were reported in Arabidopsis during seedling establishment (Bourbousse et al., 2015). Chromatin interaction maps prepared with *in situ* Hi-C reported stable genome architecture with chromosomal decondensation during cold stress in rice seedlings (Liu et al., 2017). Recently, transposon activation and modulation in the 3D genome of Arabidopsis under heat stress were reported (Sun L. et al., 2020). Increased nuclear size, decreased interactions among KEEs, switching (A→B and B→A) of A/B compartments, and weakening of chromatin compartmentalization under heat stress were demonstrated. However, there is a lack of consensus on modulation in chromatin conformation in response to environmental cues, which need to be built up to better understand the roles of 3D genome organization in gene regulation.

5.1 3D Genome Dynamics During Growth and Development

Dynamic changes in 3D genome organization during growth and development are being studied using genome mapping technologies. To better understand 3D genomics and its dynamics over time, a 4D nucleome project in mammals was conceived (Dekker et al., 2017). Several high-order structural reorganizations were observed through chromatin interaction analyses during the development of embryonic stem cells and fertilized eggs in humans (Dixon et al., 2015; Flyamer et al., 2017). However, only a fewer report on the dynamics of 3D genome organization during plant development is available. Changes in chromatin accessibility during plant cell differentiation imply that higher-order chromatin organization is a dynamic process (Wang et al., 2016; Sijacic et al., 2018; Sullivan et al., 2019). The dynamics of 3D genome folding at different developmental stages and growth conditions (4D genomics) in plants need to be explored.

5.1.1 Tissue-Specific Dynamics of Chromatin Architecture

Tissue-specific comparison of 3D chromatin architecture in rice, foxtail millet, and maize using Hi-C revealed stability in global A/B compartments across the tissues with tissue-specific

dynamism in local A/B compartments associated with differential gene expression (Dong et al., 2020a). Analysis of mesophyll and endosperm of rice, bundle sheath and mesophyll of foxtail millet, and bundle sheath, mesophyll, and endosperm tissues of maize indicated stable global A/B compartment partitions while dynamic local A/B compartments. Chen et al. (2020) revealed the features of chromatin architecture in sex differentiation in Jatropha, which provides regulatory mechanisms of sex determination in higher plants. Based on the high-quality reference genome assembly prepared with the help of Hi-C data, the differences in chromatin architecture between monoecious and gynoecious floral buds of Jatropha could be identified. The differentially expressed genes (DEGs) were observed to be significantly enriched in altered A/B compartments and TAD regions which occurred preferentially in the differential contact regions between monoecious and gynoecious buds (Chen et al., 2020). The DEGs associated with flower development/ hormone synthesis displayed different genomic interaction patterns, which demonstrate that chromatin organization plays important role in the regulation of gene expression during growth and development in plants.

5.1.2 Chromatin Dynamics During Cellular Processes

Chromatin dynamics is not only associated with transcriptional regulation of gene expression but also with other essential cellular processes like DNA replication. The process of DNA replication is essential for genomic content duplication before the cell enters mitosis. DNA replication throughout the genome is generally not a homogeneous process; rather, it is associated with the local histone marks and 3D chromosome architecture. Euchromatin (generally localized in the interior of the nucleus) is replicated earlier than the heterochromatin (localized in the perinuclear region) in animals (Rhind and Gilbert, 2013). Pope et al. (2014) reported that TADs are stable units of replication-timing regulation and replication domain boundaries share a near-perfect correlation with TAD boundaries in humans and mice. Similar studies in Arabidopsis suspension cells reported euchromatin to duplicate early compared to that of heterochromatin (Concia et al., 2018). Moreover, live imaging of replisomes in Arabidopsis revealed dynamics in DNA replication during the S phase of the cell cycle (Yokoyama et al., 2016). The same correlation was observed on comparing chromatin regions with different replication timing in the nuclei of maize root tip indicated open chromatin (euchromatin) to duplicate early compared to the densely packed heterochromatin domains during the S phase (Wear et al., 2017).

Hi-C analysis of tomato and maize genomes showed a large number of long-range chromatin loops to be formed, linking them with interstitial active chromatin regions and suggesting spatial clustering of the expressed genes (Dong et al., 2017). The interaction network of active chromatin by ChIA-PET in maize revealed the role of such physical interactions on gene expression (Li E. et al., 2019; Peng et al., 2019). The formation of chromatin loops connects with active genes, the genes forming long-range chromatin interactions show higher expression, and the gene pairs linked with chromatin loops show co-expression. A recent

ChIA-PET study in rice demonstrated coordinated expression of the active genes connected by the formation of chromatin loops (Zhao et al., 2019). These findings suggest that active chromatin domains in nuclei form extensive physical contacts, and associate with gene expression as well as certain essential cellular processes.

5.1.3 3D Genome Dynamics During Environmental Stresses

To ensure survival, proper growth, development, and reproduction, plants need to adapt to the prevailing environmental conditions, antedate potential changes, while maintaining the necessary flexibility to respond to other fluctuations. Light, temperature, water, etc. fluctuate on a seasonal basis throughout the year. Long-term adaptation and short-term reaction to environmental factors are underpinned by the changes in gene(s) expression (Franklin et al., 2014). The changes in gene expression and chromatin organization due to histone modifications and nuclear compartmentalization are vital for plant responses to environmental cues (Sun L. et al., 2020; Yung et al., 2021). This section focuses on how environmental factors affect histone modifications, chromatin architecture, nuclear localization, and their effects on regulation of gene expression, plant development, and stress tolerance. These might help to answer some of the questions like: does environmental stress influence positioning/accessibility of gene/chromatin in the nucleus, does such chromatin relocalization relate with the changes in gene expression? These may also help to decipher the structural determinants that energize chromatin localization and chromosomal interactions in cells, tissues, and species in response to the environmental stimuli (4D genomics).

There are two suggested mechanisms, among many other possible strategies, involving different enzymatic paths to accomplish chromatin reorganization. One operates through chromatin remodelers that modulate DNA–histone interactions *via* ATP hydrolysis, while the other utilizes specialized enzymes to (de)methylate DNA or post-translationally modify histone proteins. The SWI2/SNF2 family of chromatin remodeling complexes (part of a large superfamily of helicases–translocases) use ATP energy to gain access to the DNA sequences (Clapier and Cairns, 2009). DNA (de)methylases and histone (de)acetylases [histone acetyltransferases (HATs) and histone deacetylases (HDACs), e.g., BAF60] can regulate the accumulation of methylated DNA base(s) and H3K27me3/H3K9Ac histone marks to control chromatin architecture for regulation of gene expression during the developmental and/or under environmental stresses (4D genomics) (Jegu et al., 2014; Jegu et al., 2017). Despite the remarkable/continuous progress being made in decoding the linear genome, epigenome, and spatial genome architecture (3D genome), regulation of the functional changes in gene expression over time and environmental conditions (4D genome) remains unclear (Aboelnour and Bonev, 2021).

5.1.4 Drought-Induced Chromatin Dynamics

An SWI/SNF chromatin remodeler BAF60 was reported to have dual regulatory functions in epigenomic modification as well as on chromatin architecture. BAF60, having histone deacetylase

activity, regulates the level of H3K9Ac histone marks, and transcriptionally suppresses the downstream genes (Jegu et al., 2014; Jegu et al., 2017). The nuclear periphery has a proven role in the regulation of genome topology. Heterochromatic domains were reported to be enriched at the nuclear periphery and Crowded Nuclei 1 (CRWN1) interacts with the chromatin domains in modulating chromatin positioning at the nuclear periphery in Arabidopsis (Bi et al., 2017; Hu et al., 2019). SWI/SNF chromatin remodeler subunit OsSWI3C interacts with OsNMCP1 (a lamin-like protein), which regulates drought tolerance through modulation in chromatin accessibility in rice (Yang et al., 2020) (**Figure 4A**). Rice possesses a distinct 3D genome pattern of chromosomal compartment folding and spatial distribution which is different from the mammalian 3D genome (Zhao et al., 2019; Zhou et al., 2019). Integration of transcriptome, epigenome and other *omics* data might help better understand the effects of 3D genome dynamics on the regulation of gene expression affecting important agronomic traits, and could lay the foundation for crop improvement.

5.1.5 Salt-Induced Chromatin Dynamics

Chromatin accessibility was reported to be reduced under salt stress in Arabidopsis (Raxwal et al., 2020). Expressions of some of the chromatin remodeling complexes (e.g., SNF2 and SWR1 factors) have been reported to be responsive to salt stress (Li et al., 2011). Chromatin-remodeling complexes are involved in ATP-dependent repositioning of nucleosomes and changes in the core histone composition of a nucleosome, which regulates chromatin accessibility under stressful conditions (Clapier et al., 2017; Yung et al., 2021). Studies also suggest that Topless-like/Topless-like protein (TPL/TPR) interacts with HDAc to regulate stress responses (Tang et al., 2016; Cheng et al., 2018) (**Figure 4B**). TPL and Indeterminate Spikelet 1 (IDS1) interact with HDAc to form an IDS1-TPL-HDA1 transcriptional repression complex through histone deacetylation. Under salt stress, Pickle (PKL), a well-characterized CHD3-type chromatin-remodeling factor, mediates the accumulation of H3K27me3 at the target gene (Yang et al., 2019). *Photoperiod Independent Early* flowering 1 (PIE1) and *Actin-Related* Protein 6 (ARP6), the components of SWR1 chromatin-remodeling complex (Carter et al., 2018), mediate incorporation of H2A.Z into nucleosomes (Deal et al., 2007). PKL, PIE1, and APR6 were reported to be involved in salt stress tolerance in Arabidopsis (Sura et al., 2017; Yang et al., 2019). As the H2A.Z-enriched nucleosomes are also enriched with H3K27me3 at specific gene loci, PIE1 was suggested to be responsible for the incorporation of H2A.Z into nucleosomes.

Histone deacetylase 1 (OsHDAc1) was reported to repress *OsSOS1* in rice through interacting with a recruiter Indeterminate Spikelet 1 (OsIDS1) (Cheng et al., 2018). In soybean, Plant Homeodomain 5 (GmPHD5) protein (reader of H3K4me2) was reported to interact with HAT and Soybean Imitation Switch (GmISWI) protein (Wu et al., 2011). Acetylation of lysine residues in the tails of histone proteins neutralizes the positive charge and reduces electrostatic interaction between histones–DNA, and helps to loosen the DNA packing, allowing the access of transcription machinery to the gene(s)
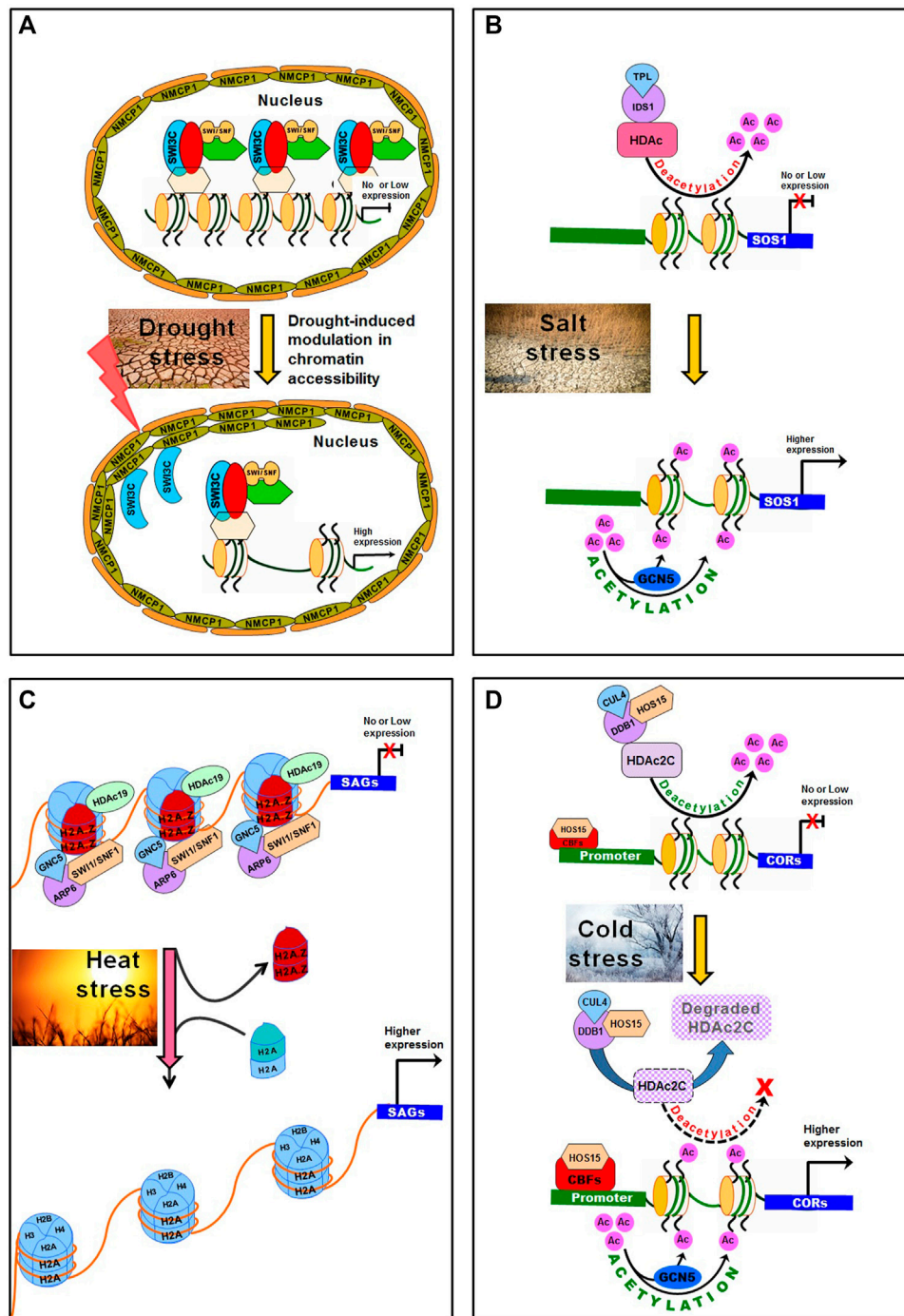
FIGURE 4 | Modulation in chromatin accessibility under abiotic stresses in plants. **(A)** Normally, the lamin-like proteins OsNMCP1 regulate drought tolerance through modulating chromatin accessibility *via* interaction with a chromatin remodeler OsSWI3C in rice. Switch/Sucrose Non-Fermenting (SWI/SNF) complexes interact with OsSWI3C to change the structure of nucleosome, resulting in gene silencing. Under drought stress, OsNMCP1 gets induced and interacts with OsSWI3C, which releases OsSWI3C from the gene-silencing SWI/SNF complexes, resulting in improved chromatin accessibility and higher expression of drought-responsive genes. **(B)** Topless-like/Topless-like protein (TPL/TPR) and Indeterminate Spikelet 1 (IDS1) interact with Histone Deacetylase (HDAc) to form an IDS1-TPL-HDA1 transcriptional repression complex through histone deacetylation. Under salt stress, acetylation of H3K9 and H3K14 by histone acetyltransferase (General Control Non-repressed Protein 5, GCN5), contributes to salt tolerance by activating salt stress-responsive genes (e.g. SOS1). **(C)** Under heat stress, SWI1/SNF1 complex interacts with GCN5 and ARP6 to dissociate H2A.Z (and insertion of H2A into the nucleosome), which causes no transcription of heat-responsive genes. On normal weather, the SWI1/SNF1—ARP6 complex plays important role in placing H2A.Z into the nucleosome. **(D)** Under cold stress, HOS15, in association with DNA Damaged Binding Protein1 (DDB1) and Cullin 4 (CUL4) acts as E3 ubiquitin ligase which degrades HDAc2C causing hyperacetylation of histone H3 on Cold Regulated (COR) chromatin. This makes binding of CBF proteins to COR promoter through High-expression of Osmotically Responsive Gene 15 (HOS15) leading to active expression of COR genes. Moreover, the GCN5 modulates histone acetylation of COR chromatin. At the normal temperature, HOS15 forms a complex with HDAc2C to repress COR expression *via* hypoacetylation of the COR chromatin.

(Bannister and Kouzarides, 2011). In addition, the chromatin-remodeling factor PKL was reported to modulate chromatin accessibility to other transcriptional regulators, leading to altered expression of salt stress-responsive genes (Yung et al., 2021). The accessibility of a gene was reported to be modulated by post-translational modifications of histone proteins as well as the chromatin-remodelling complexes that regulate nucleosome assembly and spacing (Lusser and Kadonaga, 2003; Hargreaves and Crabtree, 2011; Clapier et al., 2017).

5.1.6 Heat and Light-Induced Chromatin Dynamics

Varying temperature (low or high temperature) significantly affects plant growth and crop yield. Chromatin remodeling is one of the molecular mechanisms implicated in temperature sensing and regulating gene expression (Tasset et al., 2018). Repression of histone deacetylation was reported to prevent hypocotyl elongation under elevated temperatures (Tasset et al., 2018). Exclusion/integration of H2A.Z nucleosomes has been another chromatin remodeler that increases chromatin accessibility leading to changes in gene expression under elevated temperature (Quint et al., 2016; Cortijo et al., 2017; Dai et al., 2017) (**Figure 4C**). The increased temperature was reported to induce H3K9 deacetylation of nucleosome of PIF4 and YUCCA8 loci (Tasset et al., 2018; van der Woude et al., 2019) which are involved in temperature responses (Franklin et al., 2014; Quint et al., 2016). In many higher eukaryotes, heterochromatin comprises transposable elements (TEs) which are silenced by epigenetic modifications. Hi-C analysis for comparative genome-wide high-resolution chromatin packing under normal and heat stress conditions, the stress was reported to cause global rearrangement of 3D genome in Arabidopsis. Heat activation of TEs correlates with reduced chromosomal interactions engaging pericentromeric, KNOT, knob, and upstream and downstream flanking regions of activated TEs (Grob et al., 2014).

Temperature and light influence the developmental trajectory/morphology of the plant. The light-regulated modulations in chromatin architecture were initially reported based on the photomorphogenesis responses (Barneche et al., 2014). Studies show that shifting from dark to light results in increased nuclear size and the number of chromocenters in Arabidopsis. Changes in chromatin architecture and nuclear organization can modulate gene expression, which leads to short- and long-term plant acclimatization/adaptation to the environment. Hence, it becomes important to investigate the changes in chromatin architecture (composition, structure, and topology) that modulate the expression of genes in response to the variations in temperature and light (4D genomics).

5.1.7 Cold Stress-Induced Chromatin Dynamics

In plants, exposure to cold stress has been reported to alter chromatin configuration through the autonomous pathway and silencing of the MADS-box transcriptional repressor of Flowering Locus C (FLC) (Fornara et al., 2010; Wu et al., 2019). When the temperature returns to normal, FLC is silenced and activates the flowering genes (Fornara et al., 2010). Other proteins, like Curly Leaf (CLF) and Swinger (SWN) having H3K27me3 activity, mediate *FLC* repression during the vegetative stage of plant development (Bouyer et al., 2011; Lopez-Vernaza et al., 2012). *Clf* mutants were reported to show reduced H3K27me3 repressive mark causing up-regulated expression of *FLC* (Lopez-Vernaza et al., 2012). Some of the members of Polycomb repressive complex 2, which constitute CLF and SWN proteins, are also required for silencing of *FLC* (Berry et al., 2017; Portoso et al., 2017; Laugesen et al., 2019). When the *FLC* chromatin is active, it shows a low level of H3K27me3, and a high level of histone marks (H3K4me3, H3Ac, and H3K36me3) associated with transcriptionally active chromatin (Hyun et al., 2017; Wu et al., 2019). A recent study reported *H*igh-expression *O*smotically *R*esponsive gene 15 (HOS15) to work together with Histone Deacetylase 2C (HDAc2C) by directly binding to Cold Regulated (*COR*, e.g., *COR47* and *COR15A*) genes' promoter (Park et al., 2018). Histone acetylation/deacetylation (by HAT and HDAc) was reported to play role in cold responses (Kim et al., 2015). Arabidopsis histone deacetylase (HDAc6) was reported to be up-regulated by cold stress to positively regulate cold tolerance (To et al., 2011). Under normal temperature, HOS15 and HDAc2C make a complex that represses the expression of *COR* genes by hypoacetylation at the chromatin. Under cold stress, HOS15 acts as an E3 ubiquitin ligase in association with *D*NA *D*amaged *B*inding protein 1 (DDB1) and Cullin 4 (CUL4) to degrade HDAc2C, which leads to hyperacetylation of histone H3 at *COR* chromatin. This promotes CBF proteins binding at the *COR* promoters *via* HOS15 to activate *COR* genes expression. Moreover, the GCN5 promotes H3 acetylation at *COR* genes (Ding et al., 2019) (**Figure 4D**).

Although compartmentalization of genome into territories, compartments, TADs, and loops appears to arise largely independent of each other, the layers of genome folding is redundant, at least partially, which help maintain the gene expression pattern (Aboelnour and Bonev, 2021). Chromatin loops are highly context-dependent and rely on the *cis*-acting elements as well as on the local chromatin environment to coordinate gene expression in a time environment dependant manner. How the regulatory loops are established and remodeled during the developmental processes and environmental stresses, and what is the functional importance of physical proximity with the changes in linear epigenome are some of the critical questions in the field of 4D genomics. Thus, genome architecture is highly diverse across the cells, tissues, and species suggesting that the relationship between 3D genome organization and molecular events like transcription/gene expression is highly dynamic.

6 DIFFERENCE IN PLANT AND ANIMAL CHROMATIN ORGANIZATION

It is well-established now that spatial organization of chromatin plays important roles in several biological processes like DNA replication, repair, gene expression, repression of TE, etc. Therefore, investigations on the 3D organization of chromatin architecture would enable a better understanding of the transcriptional regulation of gene expression/biological process.

Recent advances in NGS-based 3C technologies have enabled us to examine the 3D organization of chromatin at unprecedented scale and resolution. 3D genome organizational studies indicate conserved but distinct chromatin structures between mammals and plants at different scales ranging from chromatin loops to chromosome territories (Dogan and Liu, 2018). Chromatin organization in mammals could be presented mainly at three hierarchical levels: compartments, domains, and loops that play important roles in the transcriptional regulation of genes. Though similar organizational levels have been reported in plants, these may not have the same functions as they have in their mammalian counterpart. Combinations of 3C and high-throughput sequencing techniques have considerably improved our understanding of the spatial organization of chromosomes. While Hi-C captures all the chromatin interactions at low resolution (Lieberman-Aiden et al., 2009), ChIA-PET (Fullwood et al., 2009) and Hi-ChIP or PLAC-Seq (Fang et al., 2016; Mumbach et al., 2016) generate high-resolution interaction maps of the loci occupied by proteins (modified histones, transcription factors, and RNA polymerase II) which can be pulled down by ChIP. These techniques provide extraordinary insights into 3D chromatin architecture and functions, but only a little is known about the functions of the chromatin structural organization in plants.

Studies show that active chromatin interacts with other active regions, and repressive chromatin interacts with other repressed regions. Thus, a genome is partitioned into two different compartments: active/euchromatin and repressive/heterochromatin, which are referred to as A and B compartments, respectively. Mammalian A compartment is actively transcribed, open chromatin, enriched with active histone marks like H3K4me3 and H3K27ac having high GC content. On the other hand, the B compartment is enriched with repressive histone marks like H3K9me3, associated with the nuclear lamina, and rich in AT (Lieberman-Aiden et al., 2009; Ryba et al., 2010). The compartment partitioning is dynamic and switches frequently in different tissues or at developmental stages. Dixon et al. (2015) reported that 36% of the human genome switches for the compartments and the loci that switched from A to B showed decreased expression, while those switched from B to A showed increased expression. In plants, the actively transcribed euchromatin arms form the A compartment and the pericentromeric heterochromatin forms the B compartment (Feng et al., 2014; Grob et al., 2014). This partitioning is largely stable across tissues (Dong et al., 2020b), and reduced compartment interaction has been reported in DNA methylation mutants of Arabidopsis (Feng et al., 2014), and in the endosperm tissues of rice/maize where DNA demethylation occurs naturally (Dong et al., 2020b). Based on Hi-C interaction analysis of Arabidopsis chromosome arms, the regions observed to interact with chromocenter were named *Compacted Structural Domains* (CSDs), while the other regions containing active/expressed genes are called *Loose Structural Domains* (LSDs) (Grob et al., 2014). Moreover, CSDs are associated with the nuclear periphery and require lamina-like proteins (CRWN1 and CRWN4), as well as DNA methylation at CHG and CHH contexts (Bi et al., 2017; Grob and Grossniklaus, 2019; Hu et al., 2019).

Chromatin domains are a prominent feature in the mammalian genome and are referred to as TADs. Interaction frequency within a TAD is higher than that between TADs, which reduce significantly at the domain boundaries (Dixon et al., 2012; Nora et al., 2012). Borders of some TADs were reported to have CTCF and cohesin to help the formation of chromatin loop, known as "loop domain" (**Figure 5A**) (Rao et al., 2014). Cohesin binding has been reported to be highly mobile, and its binding often occurs at the inner side of CTCF at the TAD border in the human genome (Tang et al., 2015). Flipping of CTCF binding site disrupts the TAD (Guo et al., 2015), and degradation of CTCF/cohesin subunit also disrupts the TAD structure (Nora et al., 2017; Rao et al., 2017). Based on high-resolution Hi-C analysis, human TADs could be further partitioned into subdomains (subTADs or contact domains) (Rao et al., 2014). Although TAD is not a prominent feature in Arabidopsis (Feng et al., 2014; Grob et al., 2014), TAD-like structures could be identified in Arabidopsis wherein boundaries are enriched with active genes/active epigenetic marks like open chromatin, H3K4me3, and H3K9ac (Wang et al., 2015). Compared to the mammalian TADs, Arabidopsis TAD-like structures are smaller and the interaction is weaker (**Figure 5B**). The occurrence of a few TAD-like structures was also reported in H3K27me3-rich and H3K9me2-rich chromocenter heterochromatic regions in Arabidopsis (Feng et al., 2014; Rowley et al., 2017). However, in plants having larger genomes like maize and tomato more frequent occurrence of TAD-like structures could be identified (**Figure 5C**) (Dong et al., 2017; Liu et al., 2017; Wang Q. et al., 2018; Dong et al., 2018; Dong et al., 2020a).

Distal regulatory elements can interact physically with genes through the formation of loops, which is well-studied in the β-globin gene revealing a causal relationship between looping and gene activation (Smallwood and Ren, 2013). The most prominent loops are observed between the loci bound by the CTCF and cohesion, which show higher interaction frequency and are relatively conserved (Dowen et al., 2014). Gene-to-gene and gene-to-distal active chromatin loops were recently identified in rice and maize (Ricci et al., 2019; Zhao et al., 2019). Such chromatin loops are often observed for the expression of QTL (Peng et al., 2019). Unlike the loops observed in animals, plant domains are not enriched with the loops 10 Kb range, which support the argument that plant domains do not confine enhancer-promoter interactions, in contrast to the mammalian TADs. The distance between two loci joined by a chromatin loop in maize was observed to be shorter for the syntenic gene pairs in the related species like rice, sorghum, and millet compared with those of the non-loop genes (Dong et al., 2020a). Despite a huge variation in genome size, most of the plant species have similar numbers of genes/open chromatin regions but due to the insertion of TEs and repeats between genes and distal regulatory elements the genome size increases (Dong et al., 2020b).

7 CONSTRAINTS OF 3D GENOMICS TECHNIQUES

The discoveries made with the use of 3D genomics techniques including hierarchical chromatin structures like chromatin loops,
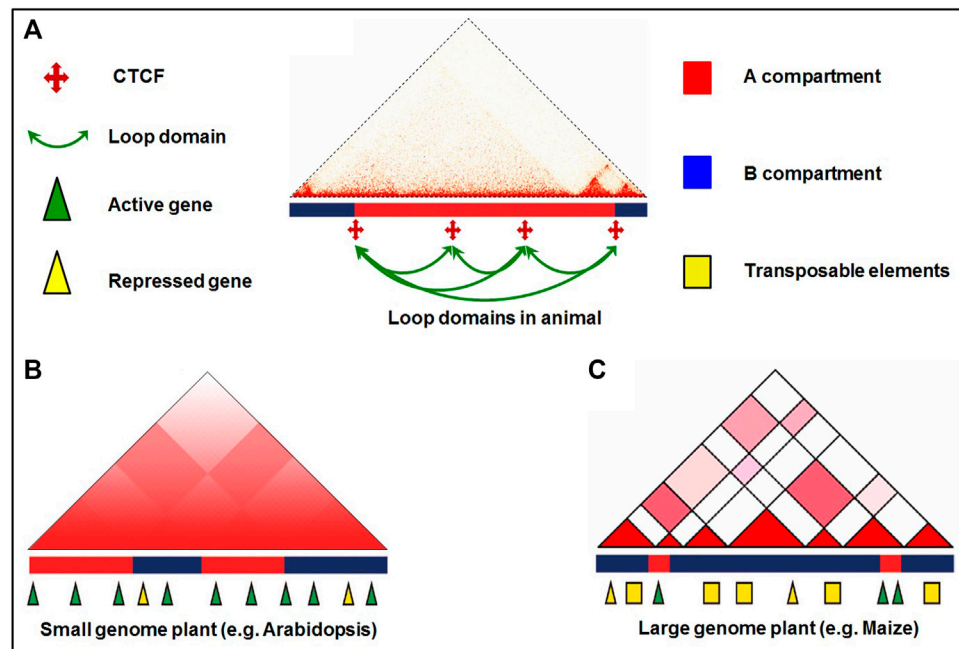
FIGURE 5 | Comparison of 3D chromatin organization of animals and plants. **(A)** Loop domain in the mammalian genome. CTCT loops are formed at the domain corner, and these domains are located within a compartment. Small adjacent loop domains form larger domains having nested structures. The dynamics of a loop domain is associated with changes in CTCF binding. **(B)** In Arabidopsis, chromosome arms are partitioned into loose structural domains (LSDs) and compacted structural domains (CSDs) which are comparable to the local A/B compartments rather than the mammalian TAD and the global compartment domain of large-genome plant. **(C)** Compartment domains in the large-genome plants often overlap with local compartments having active genes located inside the domain associated with the A compartment. Transposable elements and repressed genes are located in the domain with the B compartment.

TADs, A/B compartments and sub-compartments, chromatin territories have revolutionized structural and functional genomic analyses. However, there are certain limitations of these techniques. Most of the 3D genomics analyses indicate chromatin configuration of population average which varies in different tissues/cells with changing environmental conditions. Therefore, the need of the day is to study chromatin conformation at the single-cell level. Though FISH provides easy visualization of repetitive sequences and dynamics at an individual locus (Cui et al., 2016), its stringent preparation/ protocol affects chromatin organization. Similarly, low throughput coverage of 3C, limitation of 4C to one viewpoint, and limited coverage of 5C are some of the constraints being faced by the researchers. Moreover, 5C may not be suitable for interaction studies on relatively smaller genomes like that of yeast, Drosophila, and Arabidopsis (Zhang and Wang, 2021). Similarly, Hi-C can be better used for studying alterations in TAD/supra-TAD in chromatin organization, but it may not be appropriate for studies on the individual locus (Sati and Cavalli, 2017). As Hi-C relies on RE to break chromatin into smaller fragments, the restriction/recognition sites are heterogeneously distributed in the genome which limits the spatial resolution of the contact map (Ma et al., 2015). Incomplete digestion by RE, spurious ligation, and cross-molecular ligation (noise) might perplex the Hi-C findings (Hoshino et al., 2017). Different experimental methods result in the identification of different TAD sizes and numbers (Zufferey et al., 2018) probably because

of the low coverage of the 3C/derived techniques (Xu et al., 2020) and the different models that each algorithm employs (Boltsis et al., 2021). In single-cell Hi-C analysis, detection of TADs is generally not reproducible but reassembled on combining the maps for a population (Flyamer et al., 2017). This strengthens the view that a TAD is visible only when many cells are analyzed. However, further optimization and advances in the techniques with increased resolution and coverage are expected to make 3D genomics/Hi-C an exciting discovery.

8 FUTURE PERSPECTIVES

Chromatin conformation has considerable effects on gene expression and regulation, and regions with strong chromatin interaction generally show functional dependency (Mendes et al., 2013). Alterations in chromatin compactness affect the accessibility of chromatin to TFs, chromatin remodelers, and transcriptional machinery, which influence gene expression levels (Rutowicz et al., 2019). Single-gene resolution Hi-C map of Arabidopsis showed that local chromatin loops (between the 5′ and 3′ ends of the genes) were associated with highly expressed genes (Liu C. et al., 2016). ChIA-PET and DLO Hi-C based high-resolution chromatin interaction maps of maize demonstrated chromatin loops to be formed between the regulatory elements and the genes (Li E. et al., 2019; Peng et al., 2019; Sun Y. et al., 2020). Promoter–promoter

interaction map associated with H3K4me3 and RNA polymerase II in rice reported co-transcription of the genes (Zhao et al., 2019). Based on the DNA methylation, histone modification, and chromatin accessibility data, enhancers are being identified, which are noncoding DNA elements that function independently of transcriptional direction, relative position with the promoter, and participate in gene regulation through long-distance chromatin interaction through chromatin loop formation (Zhao et al., 2019).

Until recently, studies on 3D genome organization have been challenging tasks due to technical difficulties; however, technological advances have enabled us to take up such studies with unprecedented resolution and accessibility. Technological developments in the 3D genomics techniques like ChIA-PET (Fullwood et al., 2009), Capture-Hi-C (Mifsud et al., 2015), and Hi-ChIP (Mumbach et al., 2016) have enabled investigating short- and long-range chromatin interactions with better resolution for their regulatory roles (Li E. et al., 2019; Ricci et al., 2019; Concia et al., 2020). With the integration of robotics and microfluidics, 3D chromatin topology can be analyzed at a single-cell level for cell-type-specific studies (Boettiger and Murphy, 2020). Moreover, the resolution limit imposed by traditional microscopy can be surmounted by next-generation, super-resolution (optical resolution ≥50 nm) techniques like structured illumination microscopy (SIM), photoactivated localization microscopy (PALM), stochastic optical reconstruction microscopy (STORM), and stimulated emission depletion microscopy (STED) (Schubert, 2017; Xu and Dixon, 2020). Such next-generation microscopy for visualizing chromatin architecture has been used in mammals Xu and Dixon, 2020); however, these have rarely been used for plant studies. While 3C-based techniques provide a high-resolution map of chromatin state/genomic region of interest, the next-generation super-resolution microscopy complements the techniques by providing nano-scale imaging. Imaging resolution can be further improved by using two/multiphoton microscopy, which allows fast and dynamic imaging of nuclei (Komis et al., 2018).

Fine structures of the 3D genome are also being investigated by combining improved CRISPR technologies with ultra-high resolution microscopy in mammals. Locations of transcriptionally active and inactive regions in the nucleus were determined using sgRNAs (targeting 16 MS2 binding motifs) and catalytically inactive Cas9 (dCas9) protein (Qin et al., 2017). Moreover, CRISPR technology was also used for the functional validation of 3D genome folding by knock-out/knock-in of TAD boundary/structural proteins (CTCF and cohesins) involved in chromatin loop formation in animals (Guo et al., 2015; Lupianez et al., 2015; Fei et al., 2019). Knocking-out of 3D structural elements using CRISPR technology, TAD, and loop structures could be altered in plants which affected gene expression (Pei et al., 2021). TAD boundaries in rice and maize exhibited enrichment of plant-specific transcriptional factor binding sites, which indicates the possibility of TFs being involved in the formation of TADs (Liu et al., 2017), as observed in mammals (Stadhouders et al., 2019). Enhancers can also be knocked out to explore their effects on gene expression, which may prove to be an efficient technique for functional validation of 3D genomic findings. The lack of CTCF homologs, but the presence of cohesin homolog subunits, and TADs being not as distinct in plants as in animals (Liu et al., 2002; Zhang et al., 2020) would require further investigations on plant 3D genomics.

Certainly, there is still a lot to examine and learn, which would require more contact maps at higher resolution particularly for the plant genomes differing in size and gene density. To gain more knowledge, a comparison of the contact maps of the same genome under different environmental conditions and/or developmental stages (4D genomics) of a single cell or single cell type would be desirable. Moreover, the use of synchronized cells would help to understand the changes in chromatin architecture during the cell cycle. Furthermore, the participation of RNAs in the formation/maintenance of chromatin structures, if any, would also need to be studied. Finally, several outstanding questions will need to be answered including, but not limited to: i) do different chromatin structures exist in a cell type under changing environmental conditions and/or developmental stages? ii) Do TADs/TAD-like structures exist only in the nuclei of plant species with larger genome sizes? iii) Can the changes in gene expression modify chromatin configuration? As soon as we would get answers to these questions, several other new questions will be required to be answered. Indeed, the experiments designed to answer some of these questions are on the go in laboratories worldwide, and we believe that the next 5 years of research on 3D genomics would be more exciting than they had been in the past.

9 CONCLUSION

Sequencing and assembly of genomes for model animal and plant species were some of the ground-breaking biological research findings of the second half of the 20th century. After preparing the draft genome for the model organisms, the scientific attention moved to annotate and decipher the biological functions of protein-coding genes to get the answer to many relevant biological questions. After understanding the biological function of specific gene/protein and protein complexes, which has provided better understanding in all the fields of biology, now it has become clear that the DNA/genome sequence itself is not the absolute determinant of phenotypic traits. Subsequently, researchers around the world started investigating the so-called 'junk DNA' (which in some cases embodies the vast majority of the eukaryotic genome) that might play regulatory roles in gene expression. Hence, during the last 3 decades, efforts were made to explore the non-genetic/epigenetic/3D genomic mechanisms/features responsible for phenotypic plasticity observed in living beings.

With the advances in 3D genomics technologies, chromatin loops are being detected with the help of Hi-C/ChIA-PET which identify enhancer–promoter interactions affecting gene expression (Liu C. et al., 2016; Wang M. et al., 2018; Li E. et al., 2019; Peng et al., 2019; Sun Y. et al., 2020). Moreover, the effects of genetic structural variation (SV) on chromatin

organization in rice were analyzed which revealed alterations in chromatin topology and the rate of transcription (Zhao et al., 2019). Sequence variation and meiotic recombination rate were reported to correlate with 3D genome structures. TADs showed more single nucleotide polymorphism, SVs, and higher recombination compared to that in the inter-TAD regions, which could be associated with the epigenetic landscape of TAD, TE composition, and increased incidence of meiotic crossovers (Liao et al., 2021). Implementation of state-of-the-art techniques like CRISPR/dCas9 for editing the interacting regions/regulatory elements and chromatin interactions with the help of RNA molecules can be of particular interest to better understand the regulatory functions of chromatin architecture. Understanding the spatial organization of the genome in the nucleus and their functional implications have become a fundamental pursuit in the post-genomic era (Kong and Zhang, 2019), as this allows integration of the knowledge of linear genome with epigenomic/3D genomic regulatory networks and phenotypic data. Future 3D genomic studies will greatly benefit from the investigations at the single-cell level with the help of advancing long-read sequencing techniques and live-cell imaging which would be the key to deciphering the importance of 4D genomics for manipulation of gene regulation/expression towards the development of climate-smart crops.

REFERENCES

Aboelnour, E., and Bonev, B. (2021). Decoding the Organization, Dynamics, and Function of the 4D Genome. *Developmental Cell* 56, 1562–1573. doi:10.1016/j.devcel.2021.04.023

Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H., and Shiroishi, T. (2009). Chromosomal Dynamics at the *Shh* Locus: Limb Bud-specific Differential Regulation of Competence and Active Transcription. *Developmental Cell* 16, 47–57. doi:10.1016/j.devcel.2008.11.011

Bannister, A. J., and Kouzarides, T. (2011). Regulation of Chromatin by Histone Modifications. *Cell Res* 21, 381–395. doi:10.1038/cr.2011.22

Barneche, F., Malapeira, J., and Mas, P. (2014). The Impact of Chromatin Dynamics on Plant Light Responses and Circadian Clock Function. *J. Exp. Bot.* 65, 2895–2913. doi:10.1093/jxb/eru011

Barragán-Rosillo, A. C., Peralta-Alvarez, C. A., Ojeda-Rivera, J. O., Arzate-Mejía, R. G., Recillas-Targa, F., and Herrera-Estrella, L. (2021). Genome Accessibility Dynamics in Response to Phosphate Limitation Is Controlled by the PHR1 Family of Transcription Factors in *Arabidopsis. Proc. Natl. Acad. Sci. USA* 118, e2107558118.

Beagrie, R. A., Scialdone, A., Schueler, M., Kraemer, D. C. A., Chotalia, M., Xie, S. Q., et al. (2017). Complex Multi-Enhancer Contacts Captured by Genome Architecture Mapping. *Nature* 543, 519–524. doi:10.1038/nature21411

Beliveau, B. J., Joyce, E. F., Apostolopoulos, N., Yilmaz, F., Fonseka, C. Y., McCole, R. B., et al. (2012). Versatile Design and Synthesis Platform for Visualizing Genomes with Oligopaint FISH Probes. *Proc. Natl. Acad. Sci.* 109, 21301–21306. doi:10.1073/pnas.1213818110

Berr, A., and Schubert, I. (2007). Interphase Chromosome Arrangement in *Arabidopsis thaliana* Is Similar in Differentiated and Meristematic Tissues and Shows a Transient Mirror Symmetry after Nuclear Division. *Genetics* 176, 853–863. doi:10.1534/genetics.107.073270

Berry, S., Dean, C., Howard, M., Dean, C., and Howard, M. (2017). Slow Chromatin Dynamics Allow Polycomb Target Genes to Filter Fluctuations in Transcription Factor Activity. *Cell Syst.* 4, 445–457. e8. doi:10.1016/j.cels.2017.02.013

Bersaglieri, C., and Santoro, R. (2019). Genome Organization in and Around the Nucleolus. *Cells* 8 (6), 579. doi:10.3390/cells8060579

Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., et al. (2006). Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science* 313, 1642–1645. doi:10.1126/science.1127344

Bhadouriya, S. L., Mehrotra, S., Basantani, M. K., Loake, G. J., and Mehrotra, R. (2021). Role of Chromatin Architecture in Plant Stress Responses: An Update. *Front. Plant Sci.* 11, 603380. doi:10.3389/fpls.2020.603380

Bi, X., Cheng, Y.-J., Hu, B., Ma, X., Wu, R., Wang, J.-W., et al. (2017). Nonrandom Domain Organization of the Arabidopsis Genome at the Nuclear Periphery. *Genome Res.* 27, 1162–1173. doi:10.1101/gr.215186.116

Boettiger, A., and Murphy, S. (2020). Advances in Chromatin Imaging at Kilobase-Scale Resolution. *Trends Genet.* 36 (4), 273–287. doi:10.1016/j.tig.2019.12.010

Boettiger, A. N., Bintu, B., Moffitt, J. R., Wang, S., Beliveau, B. J., Fudenberg, G., et al. (2016). Super-resolution Imaging Reveals Distinct Chromatin Folding for Different Epigenetic States. *Nature* 529, 418–422. doi:10.1038/nature16496

Boltsis, I., Grosveld, F., Giraud, G., and Kolovos, P. (2021). Chromatin Conformation in Development and Disease. *Front. Cell Dev. Biol.* 9, 723859. doi:10.3389/fcell.2021.723859

Bonev, B., and Cavalli, G. (2016). Organization and Function of the 3D Genome. *Nat. Rev. Genet.* 17, 661–678. doi:10.1038/nrg.2016.112

Bourbousse, C., Mestiri, I., Zabulon, G., Bourge, M., Formiggini, F., Koini, M. A., et al. (2015). Light Signaling Controls Nuclear Architecture Reorganization during Seedling Establishment. *Proc. Natl. Acad. Sci. USA* 112, E2836–E2844. doi:10.1073/pnas.1503512112

Bouyer, D., Roudier, F., Heese, M., Andersen, E. D., Gey, D., Nowack, M. K., et al. (2011). Polycomb Repressive Complex 2 Controls the Embryo-To-Seedling Phase Transition. *Plos Genet.* 7, e1002014. doi:10.1371/journal.pgen.1002014

Cardozo Gizzi, A. M., Cattoni, D. I., Fiche, J.-B., Espinola, S. M., Gurgo, J., Messina, O., et al. (2019). Microscopy-based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Mol. Cell* 74, 212–222. doi:10.1016/j.molcel.2019.01.011

Carlsson, K., Wallén, P., and Brodin, L. (1989). Three-dimensional Imaging of Neurons by Confocal Fluorescence Microscopy. *J. Microsc.* 155, 15–26. doi:10.1111/j.1365-2818.1989.tb04296.x

Carter, B., Bishop, B., Ho, K. K., Huang, R., Jia, W., Zhang, H., et al. (2018). The Chromatin Remodelers PKL and PIE1 Act in an Epigenetic Pathway that

AUTHOR CONTRIBUTIONS

SuK and TM conceived the review. SuK, SaK, and SiK prepared the initial draft. SuK, KS, and SiK revised the manuscript. SuK and TM finalized the manuscript. All the authors approved the final draft for submission.

FUNDING

ACKNOWLEDGMENTS

Determines H3K27me3 Homeostasis in Arabidopsis. *Plant Cell* 30, 1337–1352. doi:10.1105/tpc.17.00867

Chen, M. S., Niu, L., Zhao, M. L., Xu, C., Pan, B. Z., Fu, Q., et al. (2020). *De Novo* genome Assembly and Hi-C Analysis Reveal an Association between Chromatin Architecture Alterations and Sex Differentiation in the Woody Plant *Jatropha Curcas*. *Gigascience* 9, 1–12. doi:10.1093/gigascience/giaa009

Cheng, X., Zhang, S., Tao, W., Zhang, X., Liu, J., Sun, J., et al. (2018). INDETERMINATE SPIKELET1 Recruits Histone Deacetylase and a Transcriptional Repression Complex to Regulate rice Salt Tolerance. *Plant Physiol.* 178, 824–837. doi:10.1104/pp.18.00324

Clapier, C. R., and Cairns, B. R. (2009). The Biology of Chromatin Remodeling Complexes. *Annu. Rev. Biochem.* 78, 273–304. doi:10.1146/annurev.biochem.77.062706.153223

Clapier, C. R., Iwasa, J., Cairns, B. R., and Peterson, C. L. (2017). Mechanisms of Action and Regulation of ATP-dependent Chromatin-Remodelling Complexes. *Nat. Rev. Mol. Cell Biol.* 18, 407–422. doi:10.1038/nrm.2017.26

Concia, L., Brooks, A. M., Wheeler, E., Zynda, G. J., Wear, E. E., LeBlanc, C., et al. (2018). Genome-wide Analysis of the Arabidopsis Replication Timing Program. *Plant Physiol.* 176, 2166–2185. doi:10.1104/pp.17.01537

Concia, L., Veluchamy, A., Ramirez-Prado, J. S., Martin-Ramirez, A., Huang, Y., Perez, M., et al. (2020). Wheat Chromatin Architecture Is Organized in Genome Territories and Transcription Factories. *Genome Biol.* 21, 104. doi:10.1186/s13059-020-01998-1

Cortijo, S., Charoensawan, V., Brestovitsky, A., Buning, R., Ravarani, C., Rhodes, D., et al. (2017). Transcriptional Regulation of the Ambient Temperature Response by H2A.Z Nucleosomes and HSF1 Transcription Factors in Arabidopsis. *Mol. Plant* 10, 1258–1273. doi:10.1016/j.molp.2017.08.014

Cremer, M., Grasser, F., Lanctôt, C., M¨uller, S., Neusser, M., Zinner, R., et al. (2008). "Multicolor 3-D Fluorescence *In Situ* Hybridization for Imaging Interphase Chromosomes," in *The Nucleus: Volume 1: Nuclei and Subnuclear Components*. Editor R. Hancock (Totowa, NJ: Humana Press), 205–239.

Crevillén, P., Sonmez, C., Wu, Z., and Dean, C. (2013). A Gene Loop Containing the floral Repressor FLC Is Disrupted in the Early Phase of Vernalization. *EMBO J.* 32, 140–148. doi:10.1038/emboj.2012.324

Cui, C., Shu, W., and Li, P. (2016). Fluorescence *In Situ* Hybridization: Cell-Based Genetic Diagnostic and Research Applications. *Front. Cell Dev. Biol.* 4, 89. doi:10.3389/fcell.2016.00089

Dai, X., Bai, Y., Zhao, L., Dou, X., Liu, Y., Wang, L., et al. (2017). H2A.Z Represses Gene Expression by Modulating Promoter Nucleosome Structure and Enhancer Histone Modifications in Arabidopsis. *Mol. Plant* 10, 1274–1292. doi:10.1016/j.molp.2017.09.007

Deal, R. B., Topp, C. N., McKinney, E. C., and Meagher, R. B. (2007). Repression of Flowering inArabidopsisRequires Activation of FLOWERING LOCUS CExpression by the Histone Variant H2A.Z. *Plant Cell* 19, 74–83. doi:10.1105/tpc.106.048447

Dekker, J., Belmont, A. S., Belmont, A. S., Guttman, M., Leshyk, V. O., Lis, J. T., et al. (2017). The 4D Nucleome Project. *Nature* 549, 219–226. doi:10.1038/nature23884

Dekker, J., and Heard, E. (2015). Structural and Functional Diversity of Topologically Associating Domains. *FEBS Lett.* 589, 2877–2884. doi:10.1016/j.febslet.2015.08.044

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing Chromosome Conformation. *Science* 295, 1306–1311. doi:10.1126/science.1067799

Deng, W., Shi, X., Tjian, R., Lionnet, T., and Singer, R. H. (2015). CASFISH: CRISPR/Cas9-mediated *In Situ* Labeling of Genomic Loci in Fixed Cells. *Proc. Natl. Acad. Sci. USA* 112, 11870–11875. doi:10.1073/pnas.1515692112

Ding, D.-Q., and Hiraoka, Y. (2017). Visualization of a Specific Genome Locus by the lacO/LacI-GFP System. *Cold Spring Harb. Protoc.* 2017, 091934. doi:10.1101/pdb.prot091934

Ding, Y., Shi, Y., and Yang, S. (20192019). Advances and Challenges in Uncovering Cold Tolerance Regulatory Mechanisms in Plants. *New Phytol.* 222, 1690–1704. doi:10.1111/nph.15696

Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., et al. (2015). Chromatin Architecture Reorganization during Stem Cell Differentiation. *Nature* 518, 331–336. doi:10.1038/nature14222

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012). Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. *Nature* 485, 376–380. doi:10.1038/nature11082

Dogan, E. S., and Liu, C. (2018). Three-dimensional Chromatin Packing and Positioning of Plant Genomes. *Nat. Plants* 4, 521–529.

Dong, P., Tu, X., Chu, P.-Y., Lü, P., Zhu, N., Grierson, D., et al. (2017). 3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments. *Mol. Plant* 10, 1497–1509. doi:10.1016/j.molp.2017.11.005

Dong, P., Tu, X., Li, H., Zhang, J., Grierson, D., Li, P., et al. (2020a). Tissue-specific Hi-C Analyses of rice, Foxtail Millet and maize Suggest Non-canonical Function of Plant Chromatin Domains. *J. Integr. Plant Biol.* 62, 201–217. doi:10.1111/jipb.12809

Dong, P., Tu, X., Liang, Z., Kang, B.-H., and Zhong, S. (2020b). Plant and Animal Chromatin Three-Dimensional Organization: Similar Structures but Different Functions. *J. Exp. Bot.* 71, 5119–5128. doi:10.1093/jxb/eraa220

Dong, Q., Li, N., Li, X., Yuan, Z., Xie, D., Wang, X., et al. (2018). Genome-wide Hi-C Analysis Reveals Extensive Hierarchical Chromatin Interactions in rice. *Plant J.* 94, 1141–1156. doi:10.1111/tpj.13925

Dostie, J., Richmond, T. A., Arnaout, R. A., Selzer, R. R., Lee, W. L., Honan, T. A., et al. (2006). Chromosome Conformation Capture Carbon Copy (5C): a Massively Parallel Solution for Mapping Interactions between Genomic Elements. *Genome Res.* 16, 1299–1309. doi:10.1101/gr.5571506

Dowen, J. M., Fan, Z. P., Hnisz, D., Ren, G., Abraham, B. J., Zhang, L. N., et al. (2014). Control of Cell Identity Genes Occurs in Insulated Neighborhoods in Mammalian Chromosomes. *Cell* 159, 374–387. doi:10.1016/j.cell.2014.09.030

Dreissig, S., Schiml, S., Schindele, P., Weiss, O., Rutten, T., Schubert, V., et al. (2017). Live-cell CRISPR Imaging in Plants Reveals Dynamic Telomere Movements. *Plant J.* 91, 565–573. doi:10.1111/tpj.13601

Dyba, M., Jakobs, S., and Hell, S. W. (2003). Immunofluorescence Stimulated Emission Depletion Microscopy. *Nat. Biotechnol.* 21, 1303–1304. doi:10.1038/nbt897

Eberharter, A., and Becker, P. B. (2002). Histone Acetylation: a Switch between Repressive and Permissive Chromatin. *EMBO Rep.* 3, 224–229. doi:10.1093/embo-reports/kvf053

Ebert, A., Hill, L., and Busslinger, M. (2015). Spatial Regulation of V-(D)J Recombination at Antigen Receptor Loci. *Adv. Immun.* 128, 93–121. doi:10.1016/bs.ai.2015.07.006

Eser, U., Chandler-Brown, D., Ay, F., Straight, A. F., Duan, Z., Noble, W. S., et al. (2017). Form and Function of Topologically Associating Genomic Domains in Budding Yeast. *Proc. Natl. Acad. Sci. USA* 114, E3061–E3070. doi:10.1073/pnas.1612256114

Fang, R., Yu, M., Li, G., Chee, S., Liu, T., Schmitt, A. D., et al. (2016). Mapping of Long-Range Chromatin Interactions by Proximity Ligation-Assisted ChIP-Seq. *Cell Res* 26, 1345–1348. doi:10.1038/cr.2016.137

Fei, T., Li, W., Peng, J., Xiao, T., Chen, C.-H., Wu, A., et al. (2019). Deciphering Essential Cistromes Using Genome-wide CRISPR Screens. *Proc. Natl. Acad. Sci. USA* 116, 25186–25195. doi:10.1073/pnas.1908155116

Feng, S., Cokus, S. J., Schubert, V., Zhai, J., Pellegrini, M., and Jacobsen, S. E. (2014). Genome-wide Hi-C Analyses in Wild-type and Mutants Reveal High-Resolution Chromatin Interactions in Arabidopsis. *Mol. Cell* 55, 694–707. doi:10.1016/j.molcel.2014.07.008

Fitzgibbon, J., Bell, K., King, E., and Oparka, K. (2010). Super-Resolution Imaging of Plasmodesmata Using Three-Dimensional Structured Illumination Microscopy. *Plant Physiol.* 153, 1453–1463. doi:10.1104/pp.110.157941

Flyamer, I. M., Gassler, J., Imakaev, M., Brandão, H. B., Ulianov, S. V., Abdennur, N., et al. (2017). Single-nucleus Hi-C Reveals Unique Chromatin Reorganization at Oocyte-To-Zygote Transition. *Nature* 544, 110–114. doi:10.1038/nature21711

Fornara, F., de Montaigu, A., and Coupland, G. (2010). SnapShot: Control of Flowering in Arabidopsis. *Cell* 141, 550–552. doi:10.1016/j.cell.2010.04.024

Franklin, K. A., Toledo-Ortiz, G., Pyott, D. E., and Halliday, K. J. (2014). Interaction of Light and Temperature Signalling. *J. Exp. Bot.* 65, 2859–2871. doi:10.1093/jxb/eru059

Fransz, P., and de Jong, H. (2011). From Nucleosome to Chromosome: a Dynamic Organization of Genetic Information. *Plant J.* 66, 4–17. doi:10.1111/j.1365-313x.2011.04526.x

Fransz, P., De Jong, J. H., Lysak, M., Castiglione, M. R., and Schubert, I. (2002). Interphase Chromosomes in Arabidopsis Are Organized as Well Defined Chromocenters from Which Euchromatin Loops Emanate. *Proc. Natl. Acad. Sci.* 99, 14584–14589. doi:10.1073/pnas.212325299

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L. A. (2016). Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* 15, 2038–2049. doi:10.1016/j.celrep.2016.04.085

Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. B., et al. (2009). An Oestrogen-Receptor-α-Bound Human Chromatin Interactome. *Nature* 462, 58–64. doi:10.1038/nature08497

Gallusci, P., Dai, Z., Génard, M., Gauffretau, A., Leblanc-Fournier, N., Richard-Molard, C., et al. (2017). Epigenetics for Plant Improvement: Current Knowledge and Modeling Avenues. *Trends Plant Sci.* 22, 610–623. doi:10.1016/j.tplants.2017.04.009

Golicz, A. A., Bhalla, P. L., Edwards, D., and Singh, M. B. (2020). Rice 3D Chromatin Structure Correlates with Sequence Variation and Meiotic Recombination Rate. *Commun. Biol.* 3, 235. doi:10.1038/s42003-020-0932-2

Gong, C., Li, T., Li, Q., Yan, L., and Wang, T. (2011). Rice OsRAD21-2 Is Expressed in Actively Dividing Tissues and its Ectopic Expression in Yeast Results in Aberrant Cell Division and Growth. *J. Integr. Plant Biol.* 53, 14–24. doi:10.1111/j.1744-7909.2010.01009.x

Gonzalez-Sandoval, A., and Gasser, S. M. (2016). On TADs and LADs: Spatial Control over Gene Expression. *Trends Genet.* 32, 485–495. doi:10.1016/j.tig.2016.05.004

Grob, S., and Cavalli, G. (2018). "Technical Review: A Hitchhiker's Guide to Chromosome Conformation Capture," in *Plant Chromatin Dynamics: Methods and Protocols*. Editors M. Bemer and C. Baroux (New York, NY: Springer), 233–246. doi:10.1007/978-1-4939-7318-7_14

Grob, S., and Grossniklaus, U. (2017). Chromosome Conformation Capture-Based Studies Reveal Novel Features of Plant Nuclear Architecture. *Curr. Opin. Plant Biol.* 36, 149–157. doi:10.1016/j.pbi.2017.03.004

Grob, S., and Grossniklaus, U. (2019). Invasive DNA Elements Modify the Nuclear Architecture of Their Insertion Site by KNOT-Linked Silencing in *Arabidopsis thaliana*. *Genome Biol.* 20, 120. doi:10.1186/s13059-019-1722-3

Grob, S., Schmid, M. W., and Grossniklaus, U. (2014). Hi-C Analysis in Arabidopsis Identifies the KNOT, a Structure with Similarities to the Flamenco Locus of Drosophila. *Mol. Cell* 55, 678–693. doi:10.1016/j.molcel.2014.07.009

Grob, S. (2020). Three-dimensional Chromosome Organization in Flowering Plants. *Brief. Funct. Genomics* 19, 83–91. doi:10.1093/bfgp/elz024

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M. B., Talhout, W., et al. (2008). Domain Organization of Human Chromosomes Revealed by Mapping of Nuclear Lamina Interactions. *Nature* 453, 948–951. doi:10.1038/nature06947

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D. U., et al. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/promoter Function. *Cell* 162, 900–910. doi:10.1016/j.cell.2015.07.038

Hargreaves, D. C., and Crabtree, G. R. (2011). ATP-dependent Chromatin Remodeling: Genetics, Genomics and Mechanisms. *Cell Res* 21, 396–420. doi:10.1038/cr.2011.32

Harper, L., Golubovskaya, I., and Cande, W. Z. (2004). A Bouquet of Chromosomes. *J. Cell Sci.* 117, 4025–4032. doi:10.1242/jcs.01363

Hong, Y., Lu, G., Duan, J., Liu, W., and Zhang, Y. (2018). Comparison and Optimization of CRISPR/dCas9/gRNA Genome-Labeling Systems for Live Cell Imaging. *Genome Biol.* 19, 39. doi:10.1186/s13059-018-1413-5

Hoshino, A., Matsunaga, T. M., Sakamoto, T., and Matsunaga, S. (2017). Hi-C Revolution: From a Snapshot of DNA-DNA Interaction in a Single Cell to Chromosome-Scale De Novo Genome Assembly. *Cytologia* 82, 223–226. doi:10.1508/cytologia.82.223

Hsieh, T.-H. S., Weiner, A., Lajoie, B., Dekker, J., Friedman, N., and Rando, O. J. (2015). Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* 162, 108–119. doi:10.1016/j.cell.2015.05.048

Hu, B., Wang, N., Bi, X., Karaaslan, E. S., Weber, A.-L., Zhu, W., et al. (2019). Plant Lamin-like Proteins Mediate Chromatin Tethering at the Nuclear Periphery. *Genome Biol.* 20, 87. doi:10.1186/s13059-019-1694-3

Huang, J., Jiang, Y., Zheng, H., and Ji, X. (2020a). BAT Hi-C Maps Global Chromatin Interactions in an Efficient and Economical Way. *Methods* 170, 38–47. doi:10.1016/j.ymeth.2019.08.004

Huang, Y., Rodriguez-Granados, N. Y., Latrasse, D., Raynaud, C., Benhamed, M., and Ramirez-Prado, J. S. (2020b). The Matrix Revolutions: towards the Decoding of the Plant Chromatin Three-Dimensional Reality. *J. Exp. Bot.* 71, 5129–5147. doi:10.1093/jxb/eraa322

Hughes, J. R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., et al. (2014). Analysis of Hundreds of Cis-Regulatory Landscapes at High Resolution in a Single, High-Throughput experiment. *Nat. Genet.* 46, 205–212. doi:10.1038/ng.2871

Hurel, A., Phillips, D., Vrielynck, N., Mézard, C., Grelon, M., and Christophorou, N. (2018). A Cytological Approach to Studying Meiotic Recombination and Chromosome Dynamics in *Arabidopsis thaliana* Male Meiocytes in Three Dimensions. *Plant J.* 95, 385–396. doi:10.1111/tpj.13942

Hyun, K.-g., Noh, Y.-S., and Song, J.-J. (2017). Arabidopsis FRIGIDA Stimulates EFS Histone H3 Lys36 Methyltransferase Activity. *Plant Cell Rep* 36, 1183–1185. doi:10.1007/s00299-017-2161-9

Jégu, T., Latrasse, D., Delarue, M., Hirt, H., Domenichini, S., Ariel, F., et al. (2014). The BAF60 Subunit of the SWI/SNF Chromatin-Remodeling Complex Directly Controls the Formation of a Gene Loop at FLOWERING LOCUS C in Arabidopsis. *Plant Cell* 26, 538–551. doi:10.1105/tpc.113.114454

Jégu, T., Veluchamy, A., Ramirez-Prado, J. S., Rizzi-Paillet, C., Perez, M., Lhomme, A., et al. (2017). The Arabidopsis SWI/SNF Protein BAF60 Mediates Seedling Growth Control by Modulating DNA Accessibility. *Genome Biol.* 18, 114. doi:10.1186/s13059-017-1246-7

Jin, F., Li, Y., Dixon, J. R., Selvaraj, S., Ye, Z., Lee, A. Y., et al. (2013). A High-Resolution Map of the Three-Dimensional Chromatin Interactome in Human Cells. *Nature* 503, 290–294. doi:10.1038/nature12644

Kaiserli, E., Perrella, G., and Davidson, M. L. (2018). Light and Temperature Shape Nuclear Architecture and Gene Expression. *Curr. Opin. Plant Biol.* 45, 103–111. doi:10.1016/j.pbi.2018.05.018

Kantidze, O. L., and Razin, S. V. (2020). Weak Interactions in Higher-Order Chromatin Organization. *Nucleic Acids Res.* 48, 4614–4626. doi:10.1093/nar/gkaa261

Kim, J.-M., Sasaki, T., Ueda, M., Sako, K., and Seki, M. (2015). Chromatin Changes in Response to Drought, Salinity, Heat, and Cold Stresses in Plants. *Front. Plant Sci.* 6, 114. doi:10.3389/fpls.2015.00114

Kolovos, P., van de Werken, H. J., Kepper, N., Zuin, J., Brouwer, R. W., Kockx, C. E., et al. (2014). Targeted Chromatin Capture (T2C): a Novel High Resolution High Throughput Method to Detect Genomic Interactions and Regulatory Elements. *Epigenetics & Chromatin* 7, 10. doi:10.1186/1756-8935-7-10

Komis, G., Novák, D., Ovečka, M., Šamajová, O., and Šamaj, J. (2018). Advances in Imaging Plant Cell Dynamics. *Plant Physiol.* 176, 80–93. doi:10.1104/pp.17.00962

Kong, S., and Zhang, Y. (2019). Deciphering Hi-C: from 3D Genome to Function. *Cell Biol. Toxicol.* 35, 15–32. doi:10.1007/s10565-018-09456-2

Koornneef, M., Fransz, P., and Jong, H. d. (2003). Cytogenetic Tools for *Arabidopsis thaliana*. *Chromosome Res.* 11, 183–194. doi:10.1023/a:1022827624082

Kumar, S., Chinnusamy, V., and Mohapatra, T. (2018). Epigenetics of Modified DNA Bases: 5-Methylcytosine and beyond. *Front. Genet.* 9, 640. doi:10.3389/fgene.2018.00640

Kumar, S., and Mohapatra, T. (2021). Dynamics of DNA Methylation and its Functions in Plant Growth and Development. *Front. Plant Sci.* 12, 596236. doi:10.3389/fpls.2021.596236

Kumar, S., PallaviChugh, C., Chugh, C., Seem, K., Kumar, S., Vinod, K. K., et al. (2021). Characterization of Contrasting rice (Oryza Sativa L.) Genotypes Reveals the Pi-Efficient Schema for Phosphate Starvation Tolerance. *BMC Plant Biol.* 21, 282. doi:10.1186/s12870-021-03015-4

Lai, B., Tang, Q., Jin, W., Hu, G., Wangsa, D., Cui, K., et al. (2018). Trac-looping Measures Genome Structure and Chromatin Accessibility. *Nat. Methods* 15, 741–747. doi:10.1038/s41592-018-0107-y

Lai, F., Orom, U. A., Cesaroni, M., Beringer, M., Taatjes, D. J., Blobel, G. A., et al. (2013). Activating RNAs Associate with Mediator to Enhance Chromatin Architecture and Transcription. *Nature* 494, 497–501. doi:10.1038/nature11884

Lakadamyali, M., and Cosma, M. P. (2015). Advanced Microscopy Methods for Visualizing Chromatin Structure. *FEBS Lett.* 589, 3023–3030. doi:10.1016/j.febslet.2015.04.012

Lanctôt, C., Cheutin, T., Cremer, M., Cavalli, G., and Cremer, T. (2007). Dynamic Genome Architecture in the Nuclear Space: Regulation of Gene Expression in Three Dimensions. *Nat. Rev. Genet.* 8, 104–115. doi:10.1038/nrg2041

Laugesen, A., Højfeldt, J. W., and Helin, K. (2019). Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. *Mol. Cell* 74, 8–18. doi:10.1016/j.molcel.2019.03.011

Lee, D.-S., Luo, C., Zhou, J., Chandran, S., Rivkin, A., Bartlett, A., et al. (2019). Simultaneous Profiling of 3D Genome Structure and DNA Methylation in Single Human Cells. *Nat. Methods* 16, 999–1006. doi:10.1038/s41592-019-0547-z

Li, E., Liu, H., Huang, L., Zhang, X., Dong, X., Song, W., et al. (2019a). Long-range Interactions between Proximal and Distal Regulatory Regions in maize. *Nat. Commun.* 10, 2633. doi:10.1038/s41467-019-10603-4

Li, G., Cai, L., Chang, H., Hong, P., Zhou, Q., Kulakova, E. V., et al. (2014). Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) Sequencing Technology and Application. *BMC Genomics* 15, S11. doi:10.1186/1471-2164-15-s12-s11

Li, G., Fullwood, M. J., Xu, H., Mulawadi, F. H., Velkov, S., Vega, V., et al. (2010). ChIA-PET Tool for Comprehensive Chromatin Interaction Analysis with Paired-End Tag Sequencing. *Genome Biol.* 11, R22. doi:10.1186/gb-2010-11-2-r22

Li, G., Liu, Y., Zhang, Y., Kubo, N., Yu, M., Fang, R., et al. (2019b). Joint Profiling of DNA Methylation and Chromatin Architecture in Single Cells. *Nat. Methods* 16, 991–993. doi:10.1038/s41592-019-0502-z

Li, G., Ruan, X., Auerbach, R. K., Sandhu, K. S., Zheng, M., Wang, P., et al. (2012). Extensive Promoter-Centered Chromatin Interactions Provide a Topological Basis for Transcription Regulation. *Cell* 148, 84–98. doi:10.1016/j.cell.2011.12.014

Li, L., Lyu, X., Hou, C., Takenaka, N., Nguyen, H. Q., Ong, C.-T., et al. (2015). Widespread Rearrangement of 3D Chromatin Organization Underlies Polycomb-Mediated Stress-Induced Silencing. *Mol. Cell* 58, 216–231. doi:10.1016/j.molcel.2015.02.023

Li, T., Jia, L., Cao, Y., Chen, Q., and Li, C. (2018a). OCEAN-C: Mapping Hubs of Open Chromatin Interactions across the Genome Reveals Gene Regulatory Networks. *Genome Biol.* 19, 54. doi:10.1186/s13059-018-1430-4

Li, X., Luo, O. J., Wang, P., Zheng, M., Wang, D., Piecuch, E., et al. (2017). Long-read ChIA-PET for Base-Pair-Resolution Mapping of Haplotype-specific Chromatin Interactions. *Nat. Protoc.* 12, 899–915. doi:10.1038/nprot.2017.012

Li, X., Xu, B., Li, X., Wang, D., An, Z., Jia, Y., et al. (2019c). Widespread Transcriptional Responses to the thermal Stresses Are Prewired in Human 3D Genome. *BioRxiv.* doi:10.1101/728220

Li, X. Y., Wang, C., Nie, P. P., Lu, X. W., Wang, M., Liu, W., et al. (2011). Characterisation and Expression Analysis of the SNF2 Family Genes in Response to Phytohormones and Abiotic Stresses in rice. *Biol. Plant* 55, 625–633. doi:10.1007/s10535-011-0160-1

Li, Z., Jiang, D., and He, Y. (2018b). FRIGIDA Establishes a Local Chromosomal Environment for FLOWERING LOCUS C mRNA Production. *Nat. Plants* 4, 836–846. doi:10.1038/s41477-018-0250-6

Liang, Z., Li, G., Wang, Z., Djekidel, M. N., Li, Y., Qian, M.-P., et al. (2017). BL-Hi-C Is an Efficient and Sensitive Approach for Capturing Structural and Regulatory Chromatin Interactions. *Nat. Commun.* 8, 1622. doi:10.1038/s41467-017-01754-3

Liao, Y., Zhang, X., Chakraborty, M., and Emerson, J. J. (2021). Topologically Associating Domains and Their Role in the Evolution of Genome Structure and Function in *Drosophila*. *Genome Res.* 31, 397–410. doi:10.1101/gr.266130.120

Lichter, P., Cremer, T., Tang, C. J., Watkins, P. C., Manuelidis, L., and Ward, D. C. (1988). Rapid Detection of Human Chromosome 21 Aberrations by *In Situ* Hybridization. *Proc. Natl. Acad. Sci.* 85, 9664–9668. doi:10.1073/pnas.85.24.9664

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* 326, 289–293. doi:10.1126/science.1181369

Lin, D., Hong, P., Zhang, S., Xu, W., Jamal, M., Yan, K., et al. (2018). Digestion-ligation-only Hi-C Is an Efficient and Cost-Effective Method for Chromosome Conformation Capture. *Nat. Genet.* 50, 754–763. doi:10.1038/s41588-018-0111-2

Lindhout, B. I., Fransz, P., Tessadori, F., Meckel, T., Hooykaas, P. J. J., and van der Zaal, B. J. (2007). Live Cell Imaging of Repetitive DNA Sequences via GFP-Tagged Polydactyl Zinc finger Proteins. *Nucleic Acids Res.* 35, e107. doi:10.1093/nar/gkm618

Liu, C.-m., McElver, J., Tzafrir, I., Joosen, R., Wittich, P., Patton, D., et al. (2002). Condensin and Cohesin Knockouts in Arabidopsis Exhibit a Titan Seed Phenotype. *Plant J.* 29, 405–415. doi:10.1046/j.1365-313x.2002.01224.x

Liu, C., Cheng, Y.-J., Wang, J.-W., and Weigel, D. (2017). Prominent Topologically Associated Domains Differentiate Global Chromatin Packing in rice from Arabidopsis. *Nat. Plants* 3, 742–748. doi:10.1038/s41477-017-0005-9

Liu, C., Wang, C., Wang, G., Becker, C., Zaidem, M., and Weigel, D. (2016a). Genome-wide Analysis of Chromatin Packing in *Arabidopsis thaliana* at Single-Gene Resolution. *Genome Res.* 26, 1057–1068. doi:10.1101/gr.204032.116

Liu, Z., Chen, O., Zheng, M., Wang, L., Zhou, Y., Yin, C., et al. (2016b). Re-patterning of H3K27me3, H3K4me3 and DNA Methylation during Fibroblast Conversion into Induced Cardiomyocytes. *Stem Cell Res.* 16, 507–518. doi:10.1016/j.scr.2016.02.037

Lobastov, V. A., Srinivasan, R., and Zewail, A. H. (2005). Four-dimensional Ultrafast Electron Microscopy. *Proc. Natl. Acad. Sci.* 102, 7069–7073. doi:10.1073/pnas.0502607102

Lopez-Vernaza, M., Yang, S., Müller, R., Thorpe, F., de Leau, E., and Goodrich, J. (2012). Antagonistic Roles of SEPALLATA3, FT and FLC Genes as Targets of the Polycomb Group Gene CURLY LEAF. *PLoS One* 7, e30715. doi:10.1371/journal.pone.0030715

Louwers, M., Bader, R., Haring, M., van Driel, R., de Laat, W., and Stam, M. (2009). Tissue- and Expression Level-specific Chromatin Looping at Maize*b1*Epialleles. *Plant Cell* 21, 832–842. doi:10.1105/tpc.108.064329

Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., et al. (2015). Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* 161, 1012–1025. doi:10.1016/j.cell.2015.04.004

Lusser, A., and Kadonaga, J. T. (2003). Chromatin Remodeling by ATP-dependent Molecular Machines. *BioEssays* 25, 1192–1200. doi:10.1002/bies.10359

Ma, W., Ay, F., Lee, C., Gulsoy, G., Deng, X., Cook, S., et al. (2015). Fine-scale Chromatin Interaction Maps Reveal the Cis-Regulatory Landscape of Human lincRNA Genes. *Nat. Methods* 12, 71–78. doi:10.1038/nmeth.3205

Ma, Y., Wang, M., Li, W., Zhang, Z., Zhang, X., Tan, T., et al. (2017). Live Cell Imaging of Single Genomic Loci with Quantum Dot-Labeled TALEs. *Nat. Commun.* 8, 15318. doi:10.1038/ncomms15318

Mao, Y. S., Zhang, B., and Spector, D. L. (2011). Biogenesis and Function of Nuclear Bodies. *Trends Genet.* 27, 295–306. doi:10.1016/j.tig.2011.05.006

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S. O., Wicker, T., et al. (2017). A Chromosome Conformation Capture Ordered Sequence of the Barley Genome. *Nature* 544, 427–433. doi:10.1038/nature22043

Matzke, A. J. M., Huettel, B., van der Winden, J., and Matzke, M. (2005). Use of Two-Color Fluorescence-Tagged Transgenes to Study Interphase Chromosomes in Living Plants. *Plant Physiol.* 139, 1586–1596. doi:10.1104/pp.105.071068

Mendes, M. A., Guerra, R. F., Berns, M. C., Manzo, C., Masiero, S., Finzi, L., et al. (2013). MADS Domain Transcription Factors Mediate Short-Range DNA Looping that Is Essential for Target Gene Expression in Arabidopsis. *The Plant Cell* 25, 2560–2572. doi:10.1105/tpc.112.108688

Mifsud, B., Tavares-Cadete, F., Young, A. N., Sugar, R., Schoenfelder, S., Ferreira, L., et al. (2015). Mapping Long-Range Promoter Contacts in Human Cells with High-Resolution Capture Hi-C. *Nat. Genet.* 47, 598–606. doi:10.1038/ng.3286

Muller, H., Gil, J., Jr., and Drinnenberg, I. A. (2019). The Impact of Centromeres on Spatial Genome Architecture. *Trends Genet.* 35, 565–578. doi:10.1016/j.tig.2019.05.003

Mumbach, M. R., Rubin, A. J., Flynn, R. A., Dai, C., Khavari, P. A., Greenleaf, W. J., et al. (2016). HiChIP: Efficient and Sensitive Analysis of Protein-Directed Genome Architecture. *Nat. Methods* 13, 919–922. doi:10.1038/nmeth.3999

Nagaki, K., and Yamaji, N. (2020). Decrosslinking Enables Visualization of RNA-Guided Endonuclease-In Situ Labeling Signals for DNA Sequences in Plant Tissues. *J. Exp. Bot.* 71, 1792–1800. doi:10.1093/jxb/erz534

Nagano, T., Lubling, Y., Stevens, T. J., Schoenfelder, S., Yaffe, E., Dean, W., et al. (2013). Single-cell Hi-C Reveals Cell-To-Cell Variability in Chromosome Structure. *Nature* 502, 59–64. doi:10.1038/nature12593

Nagano, T., Lubling, Y., Várnai, C., Dudley, C., Leung, W., Baran, Y., et al. (2017). Cell-cycle Dynamics of Chromosomal Organization at Single-Cell Resolution. *Nature* 547, 61–67. doi:10.1038/nature23001

Németh, A., Conesa, A., Santoyo-Lopez, J., Medina, I., Montaner, D., Péterfia, B., et al. (2010). Initial Genomics of the Human Nucleolus. *Plos Genet.* 6, e1000889. doi:10.1371/journal.pgen.1000889

Ni, Y., Cao, B., Ma, T., Niu, G., Huo, Y., Huang, J., et al. (2017). Super-resolution Imaging of a 2.5 Kb Non-repetitive DNA *In Situ* in the Nuclear Genome Using Molecular beacon Probes. *eLife* 6, e21660. doi:10.7554/eLife.21660

Nora, E. P., Goloborodko, A., Valton, A.-L., Gibcus, J. H., Uebersohn, A., Abdennur, N., et al. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* 169, 930–944. e922. doi:10.1016/j.cell.2017.05.004

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., et al. (2012). Spatial Partitioning of the Regulatory Landscape of the X-Inactivation centre. *Nature* 485, 381–385. doi:10.1038/nature11049

Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N., and Mirny, L. A. (2018). Chromatin Organization by an Interplay of Loop Extrusion and Compartmental Segregation. *Proc. Natl. Acad. Sci. USA* 115, E6697–E6706. doi:10.1073/pnas.1717730115

Nützmann, H.-W., Doerr, D., Ramírez-Colmenero, A., Sotelo-Fonseca, J. E., Wegel, E., Di Stefano, M., et al. (2020). Active and Repressed Biosynthetic Gene Clusters Have Spatially Distinct Chromosome States. *Proc. Natl. Acad. Sci. USA* 117, 13800–13809. doi:10.1073/pnas.1920474117

Ou, H. D., Phan, S., Deerinck, T. J., Thor, A., Ellisman, M. H., and O'Shea, C. C. (2017). ChromEMT: Visualizing 3D Chromatin Structure and Compaction in Interphase and Mitotic Cells. *Science* 357, eaag0025. doi:10.1126/science.aag0025

Ouyang, W., Xiong, D., Li, G., and Li, X. (2020). Unraveling the 3D Genome Architecture in Plants: Present and Future. *Mol. Plant* 13, 1676–1693. doi:10.1016/j.molp.2020.10.002

Parada, L. A., and Misteli, T. (2002). Chromosome Positioning in the Interphase Nucleus. *Trends Cell Biol.* 12, 425–432. doi:10.1016/s0962-8924(02)02351-6

Park, J., Lim, C. J., Shen, M., Park, H. J., Cha, J.-Y., Iniesto, E., et al. (2018). Epigenetic Switch from Repressive to Permissive Chromatin in Response to Cold Stress. *Proc. Natl. Acad. Sci. USA* 115, E5400–E5409. doi:10.1073/pnas.1721241115

Pecinka, A., Schubert, V., Meister, A., Kreth, G., Klatte, M., Lysak, M. A., et al. (2004). Chromosome Territory Arrangement and Homologous Pairing in Nuclei of *Arabidopsis thaliana* Are Predominantly Random except for NOR-Bearing Chromosomes. *Chromosoma* 113, 258–269. doi:10.1007/s00412-004-0316-2

Pefanis, E., Wang, J., Rothschild, G., Lim, J., Kazadi, D., Sun, J., et al. (2015). RNA Exosome-Regulated Long Non-coding RNA Transcription Controls Super-enhancer Activity. *Cell* 161, 774–789. doi:10.1016/j.cell.2015.04.034

Pei, L., Li, G., Lindsey, K., Zhang, X., and Wang, M. (2021). Plant 3D Genomics: the Exploration and Application of Chromatin Organization. *New Phytol.* 230, 1772–1786. doi:10.1111/nph.17262

Peng, Y., Xiong, D., Zhao, L., Ouyang, W., Wang, S., Sun, J., et al. (2019). Chromatin Interaction Maps Reveal Genetic Regulation for Quantitative Traits in maize. *Nat. Commun.* 10, 2632. doi:10.1038/s41467-019-10602-5

Perrella, G., Zioutopoulou, A., Headland, L. R., and Kaiserli, E. (2020). The Impact of Light and Temperature on Chromatin Organization and Plant Adaptation. *J. Exp. Bot.* 71, 5247–5255. doi:10.1093/jxb/eraa154

Peterson, C. L., and Workman, J. L. (2000). Promoter Targeting and Chromatin Remodeling by the SWI/SNF Complex. *Curr. Opin. Genet. Development* 10, 187–192. doi:10.1016/S0959-437X(00)00068-X

Pinkel, D., Landegent, J., Collins, C., Fuscoe, J., Segraves, R., Lucas, J., et al. (1988). Fluorescence *In Situ* Hybridization with Human Chromosome-specific Libraries: Detection of Trisomy 21 and Translocations of Chromosome 4. *Proc. Natl. Acad. Sci.* 85, 9138–9142. doi:10.1073/pnas.85.23.9138

Pontvianne, F., and Grob, S. (2020). Three-dimensional Nuclear Organization in *Arabidopsis thaliana*. *J. Plant Res.* 133, 479–488. doi:10.1007/s10265-020-01185-0

Pontvianne, F., and Liu, C. (2020). Chromatin Domains in Space and Their Functional Implications. *Curr. Opin. Plant Biol.* 54, 1–10. doi:10.1016/j.pbi.2019.11.005

Pope, B. D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., et al. (2014). Topologically Associating Domains Are Stable Units of Replication-Timing Regulation. *Nature* 515, 402–405. doi:10.1038/nature13986

Portoso, M., Ragazzini, R., Brenčič, Ž., Moiani, A., Michaud, A., Vassilev, I., et al. (2017). PRC 2 Is Dispensable for HOTAIR -mediated Transcriptional Repression. *EMBO J.* 36, 981–994. doi:10.15252/embj.201695335

Poulet, A., Duc, C., Voisin, M., Desset, S., Tutois, S., Vanrobays, E., et al. (2017). The LINC Complex Contributes to Heterochromatin Organisation and Transcriptional Gene Silencing in Plants. *J. Cell Sci.* 130, 590–601. doi:10.1242/jcs.194712

Probst, A. V. (2018). A Compendium of Methods to Analyze the Spatial Organization of Plant Chromatin. *Methods Mol. Bio.* 1675, 397–418. doi:10.1007/978-1-4939-7318-7_23

Qin, P., Parlak, M., Kuscu, C., Bandaria, J., Mir, M., Szlachta, K., et al. (2017). Live Cell Imaging of Low- and Non-repetitive Chromosome Loci Using CRISPR-Cas9. *Nat. Commun.* 8, 14725. doi:10.1038/ncomms14725

Quinodoz, S. A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J. M., Detmar, E., et al. (2018). Higher-order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* 174, 744–757. doi:10.1016/j.cell.2018.05.024

Quint, M., Delker, C., Franklin, K. A., Wigge, P. A., Halliday, K. J., and Van Zanten, M. (2016). Molecular and Genetic Control of Plant Thermomorphogenesis. *Nat. Plants* 2, 15190–15199. doi:10.1038/nplants.2015.190

Rao, S. S. P., Huang, S.-C., Glenn St Hilaire, B., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K.-R., et al. (2017). Cohesin Loss Eliminates All Loop Domains. *Cell* 171, 305–320. doi:10.1016/j.cell.2017.09.026

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* 159, 1665–1680. doi:10.1016/j.cell.2014.11.021

Raxwal, V. K., Ghosh, S., Singh, S., Katiyar-Agarwal, S., Goel, S., Jagannath, A., et al. (2020). Abiotic Stress-Mediated Modulation of the Chromatin Landscape in *Arabidopsis thaliana*. *J. Exp. Bot.* 71, 5280–5293. doi:10.1093/jxb/eraa286

Ray, J., Munn, P. R., Vihervaara, A., Lewis, J. J., Ozer, A., Danko, C. G., et al. (2019). Chromatin Conformation Remains Stable upon Extensive Transcriptional Changes Driven by Heat Shock. *Proc. Natl. Acad. Sci. USA* 116, 19431–19439. doi:10.1073/pnas.1901244116

Redolfi, J., Zhan, Y., Valdes-Quezada, C., Kryzhanovska, M., Guerreiro, I., Iesmantavicius, V., et al. (2019). DamC Reveals Principles of Chromatin Folding *In Vivo* without Crosslinking and Ligation. *Nat. Struct. Mol. Biol.* 26, 471–480. doi:10.1038/s41594-019-0231-0

Ren, R., Deng, L., Xue, Y., Suzuki, K., Zhang, W., Yu, Y., et al. (2017). Visualization of Aging-Associated Chromatin Alterations with an Engineered TALE System. *Cell Res* 27, 483–504. doi:10.1038/cr.2017.18

Rhind, N., and Gilbert, D. M. (2013). DNA Replication Timing. *Cold Spring Harbor Perspect. Biol.* 5, a010132. doi:10.1101/cshperspect.a010132

Ricci, W. A., Lu, Z., Ji, L., Marand, A. P., Ethridge, C. L., Murphy, N. G., et al. (2019). Widespread Long-Range Cis-Regulatory Elements in the maize Genome. *Nat. Plants* 5, 1237–1249. doi:10.1038/s41477-019-0547-0

Robson, M. I., Ringel, A. R., and Mundlos, S. (2019). Regulatory Landscaping: How Enhancer-Promoter Communication Is Sculpted in 3D. *Mol. Cell* 74, 1110–1122. doi:10.1016/j.molcel.2019.05.032

Rodriguez-Granados, N. Y., Ramirez-Prado, J. S., Veluchamy, A., Latrasse, D., Raynaud, C., Crespi, M., et al. (2016). Put Your 3D Glasses on: Plant Chromatin Is on Show. *Exbotj* 67, 3205–3221. doi:10.1093/jxb/erw168

Rowley, M. J., Nichols, M. H., Lyu, X., Ando-Kuri, M., Rivera, I. S. M., Hermetz, K., et al. (2017). Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol. Cell* 67, 837–852. doi:10.1016/j.molcel.2017.07.022

Rust, M. J., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit Imaging by Stochastic Optical Reconstruction Microscopy (STORM). *Nat. Methods* 3, 793–796. doi:10.1038/nmeth929

Rutowicz, K., Lirski, M., Mermaz, B., Teano, G., Schubert, J., Mestiri, I., et al. (2019). Linker Histones Are fine-scale Chromatin Architects Modulating Developmental Decisions in Arabidopsis. *Genome Biol.* 20, 157. doi:10.1186/s13059-019-1767-3

Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., et al. (2010). Evolutionarily Conserved Replication Timing Profiles Predict Long-Range Chromatin Interactions and Distinguish Closely Related Cell Types. *Genome Res.* 20, 761–770. doi:10.1101/gr.099655.109

Sakamoto, Y., Sato, M., Sato, Y., Harada, A., Suzuki, T., Goto, C., et al. (2020). Subnuclear Gene Positioning through Lamina Association Affects Copper Tolerance. *Nat. Commun.* 11, 5914. doi:10.1038/s41467-020-19621-z

Santi, P. A. (2011). Light Sheet Fluorescence Microscopy. *J. Histochem. Cytochem.* 59, 129–138. doi:10.1369/0022155410394857

Sati, S., and Cavalli, G. (2017). Chromosome Conformation Capture Technologies and Their Impact in Understanding Genome Function. *Chromosoma* 126, 33–44. doi:10.1007/s00412-016-0593-6

Schubert, I., Fransz, P. F., Fuchs, J., and Hans de Jong, J. (2001). Chromosome Painting in Plants. *Methods Cell Sci* 23, 57–69. doi:10.1007/978-94-010-0330-8_7

Schubert, V., Berr, A., and Meister, A. (2012). Interphase Chromatin Organisation in Arabidopsis Nuclei: Constraints versus Randomness. *Chromosoma* 121, 369–387. doi:10.1007/s00412-012-0367-8

Schubert, V., Rudnik, R., and Schubert, I. (2014). Chromatin Associations in Arabidopsis Interphase Nuclei. *Front. Genet.* 5, 389. doi:10.3389/fgene.2014.00389

Schubert, V. (2017). Super-resolution Microscopy - Applications in Plant Cell Research. *Front. Plant Sci.* 8, 531. doi:10.3389/fpls.2017.00531

Sexton, T., and Cavalli, G. (2015). The Role of Chromosome Domains in Shaping the Functional Genome. *Cell* 160, 1049–1059. doi:10.1016/j.cell.2015.02.040

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., et al. (2012). Three-dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell* 148, 458–472. doi:10.1016/j.cell.2012.01.010

She, W., Grimanelli, D., Rutowicz, K., Whitehead, M. W. J., Puzio, M., Kotliński, M., et al. (2013). Chromatin Reprogramming during the Somatic-To-Reproductive Cell Fate Transition in Plants. *Development* 140, 4008–4019. doi:10.1242/dev.095034

Shimbo, T., Kawamura, M., Wijaya, E., Takaki, E., Kaneda, Y., and Tamai, K. (2019). Cut-C: Cleavage under Tethered Nuclease for Conformational Capture. *BMC Genomics* 20, 614. doi:10.1186/s12864-019-5989-2

Sijacic, P., Bajic, M., McKinney, E. C., Meagher, R. B., and Deal, R. B. (2018). Changes in Chromatin Accessibility between Arabidopsis Stem Cells and Mesophyll Cells Illuminate Cell Type-specific Transcription Factor Networks. *Plant J.* 94, 215–231. doi:10.1111/tpj.13882

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., De Wit, E., et al. (2006). Nuclear Organization of Active and Inactive Chromatin Domains Uncovered by Chromosome Conformation Capture-On-Chip (4C). *Nat. Genet.* 38, 1348–1354. doi:10.1038/ng1896

Simonis, M., Kooren, J., and de Laat, W. (2007). An Evaluation of 3C-Based Methods to Capture DNA Interactions. *Nat. Methods* 4, 895–901. doi:10.1038/nmeth1114

Smallwood, A., and Ren, B. (2013). Genome Organization and Long-Range Regulation of Gene Expression by Enhancers. *Curr. Opin. Cell Biol.* 25, 387–394. doi:10.1016/j.ceb.2013.02.005

Solovei, I., Cavallo, A., Schermelleh, L., Jaunin, F., Scasselati, C., Cmarko, D., et al. (2002). Spatial Preservation of Nuclear Chromatin Architecture during Three-Dimensional Fluorescence *In Situ* Hybridization (3D-FISH). *Exp. Cell Res.* 276, 10–23. doi:10.1006/excr.2002.5513

Sotelo-Silveira, M., Chávez Montes, R. A., Sotelo-Silveira, J. R., Marsch-Martínez, N., and de Folter, S. (2018). Entering the Next Dimension: Plant Genomes in 3D. *Trends Plant Sci.* 23, 598–612. doi:10.1016/j.tplants.2018.03.014

Splinter, E., de Wit, E., van de Werken, H. J. G., Klous, P., and de Laat, W. (2012). Determining Long-Range Chromatin Interactions for Selected Genomic Sites Using 4C-Seq Technology: from Fixation to Computation. *Methods* 58, 221–230. doi:10.1016/j.ymeth.2012.04.009

Stadhouders, R., Filion, G. J., and Graf, T. (2019). Transcription Factors and 3D Genome Conformation in Cell-Fate Decisions. *Nature* 569, 345–354. doi:10.1038/s41586-019-1182-7

Stam, M., Tark-Dame, M., and Fransz, P. (2019). 3D Genome Organization: a Role for Phase Separation and Loop Extrusion? *Curr. Opin. Plant Biol.* 48, 36–46. doi:10.1016/j.pbi.2019.03.008

Stevens, T. J., Lando, D., Basu, S., Atkinson, L. P., Cao, Y., Lee, S. F., et al. (2017). 3D Structures of Individual Mammalian Genomes Studied by Single-Cell Hi-C. *Nature* 544, 59–64. doi:10.1038/nature21429

Sullivan, A. M., Arsovski, A. A., Thompson, A., Sandstrom, R., Thurman, R. E., Neph, S., et al. (2019). Mapping and Dynamics of Regulatory DNA in Maturing *Arabidopsis thaliana* Siliques. *Front. Plant Sci.* 10, 1434. doi:10.3389/fpls.2019.01434

Sun, L., Jing, Y., Liu, X., Li, Q., Xue, Z., Cheng, Z., et al. (2020a). Heat Stress-Induced Transposon Activation Correlates with 3D Chromatin Organization Rearrangement in Arabidopsis. *Nat. Commun.* 11, 1886. doi:10.1038/s41467-020-15809-5

Sun, Q., Perez-Rathke, A., Czajkowsky, D. M., Shao, Z., and Liang, J. (2021). High-resolution Single-Cell 3D-Models of Chromatin Ensembles during Drosophila Embryogenesis. *Nat. Commun.* 12, 205. doi:10.1038/s41467-020-20490-9

Sun, Y., Dong, L., Zhang, Y., Lin, D., Xu, W., Ke, C., et al. (2020b). 3D Genome Architecture Coordinates Trans and Cis Regulation of Differentially Expressed Ear and Tassel Genes in maize. *Genome Biol.* 21, 143. doi:10.1186/s13059-020-02063-7

Sura, W., Kabza, M., Karlowski, W. M., Bieluszewski, T., Kus-Slowinska, M., Pawełoszek, Ł., et al. (2017). Dual Role of the Histone Variant H2A.Z in Transcriptional Regulation of Stress-Response Genes. *Plant Cell* 29, 791–807. doi:10.1105/tpc.16.00573

Szabo, Q., Bantignies, F., and Cavalli, G. (2019). Principles of Genome Folding into Topologically Associating Domains. *Sci. Adv.* 5, eaaw1668. doi:10.1126/sciadv.aaw1668

Szabo, Q., Jost, D., Chang, J. M., Cattoni, D. I., Papadopoulos, G. L., Bonev, B., et al. (2018). TADs Are 3D Structural Units of Higher-Order Chromosome Organization in Drosophila. *Sci. Adv.* 4, eaar8082. doi:10.1126/sciadv.aar8082

Tan, L., Xing, D., Chang, C.-H., Li, H., and Xie, X. S. (2018). Three-dimensional Genome Structures of Single Diploid Human Cells. *Science* 361, 924–928. doi:10.1126/science.aat5641

Tang, N., Ma, S., Zong, W., Yang, N., Lv, Y., Yan, C., et al. (2016). MODD Mediates Deactivation and Degradation of OsbZIP46 to Negatively Regulate ABA Signaling and Drought Resistance in rice. *Plant Cell* 28, 2161–2177. doi:10.1105/tpc.16.00171

Tang, Z., Luo, O. J., Li, X., Zheng, M., Zhu, J. J., Szalaj, P., et al. (2015). CTCF-mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* 163, 1611–1627. doi:10.1016/j.cell.2015.11.024

Tannenbaum, M., Sarusi-Portuguez, A., Krispil, R., Schwartz, M., Loza, O., Benichou, J. I. C., et al. (2018). Regulatory Chromatin Landscape in *Arabidopsis thaliana* Roots Uncovered by Coupling INTACT and ATAC-Seq. *Plant Methods* 14, 113. doi:10.1186/s13007-018-0381-9

Tao, J., Zhang, L., Chong, K., and Wang, T. (2007). OsRAD21-3, an Orthologue of Yeast RAD21, Is Required for Pollen Development in Oryza Sativa. *Oryza Sativa. Plant J.* 51, 919–930. doi:10.1111/j.1365-313x.2007.03190.x

Tasset, C., Singh Yadav, A., Sureshkumar, S., Singh, R., van der Woude, L., Nekrasov, M., et al. (2018). POWERDRESS-mediated Histone Deacetylation Is Essential for Thermomorphogenesis in *Arabidopsis thaliana*. *Plos Genet.* 14, e1007280–21. doi:10.1371/journal.pgen.1007280

Tiang, C.-L., He, Y., and Pawlowski, W. P. (2012). Chromosome Organization and Dynamics during Interphase, Mitosis, and Meiosis in Plants. *Plant Physiol.* 158, 26–34. doi:10.1104/pp.111.187161

To, T. K., Nakaminami, K., Kim, J.-M., Morosawa, T., Ishida, J., Tanaka, M., et al. (2011). Arabidopsis HDA6 Is Required for Freezing Tolerance. *Biochem. Biophysical Res. Commun.* 406, 414–419. doi:10.1016/j.bbrc.2011.02.058

Trzaskoma, P., Ruszczycki, B., Lee, B., Pels, K. K., Krawczyk, K., Bokota, G., et al. (2020). Ultrastructural Visualization of 3D Chromatin Folding Using Volume Electron Microscopy and DNA *In Situ* Hybridization. *Nat. Commun.* 11, 2120. doi:10.1038/s41467-020-15987-2

van der Woude, L. C., Perrella, G., Snoek, B. L., van Hoogdalem, M., Novák, O., van Verk, M. C., et al. (2019). HISTONE DEACETYLASE 9 Stimulates Auxin-dependent Thermomorphogenesis inArabidopsis Thalianaby Mediating H2A.Z Depletion. *Proc. Natl. Acad. Sci. USA* 116, 25343–25354. doi:10.1073/pnas.1911694116

van Koningsbruggen, S., Gierliński, M., Schofield, P., Martin, D., Barton, G. J., Ariyurek, Y., et al. (2010). High-resolution Whole-Genome Sequencing Reveals that Specific Chromatin Domains from Most Human Chromosomes Associate with Nucleoli. *MBoC* 21, 3735–3748. doi:10.1091/mbc.e10-06-0508

van Steensel, B., and Belmont, A. S. (2017). Lamina-associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* 169, 780–791. doi:10.1016/j.cell.2017.04.022

Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D. T., Tanay, A., et al. (2015). Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. *Cell Rep* 10, 1297–1309. doi:10.1016/j.celrep.2015.02.004

Wang, C., Liu, C., Roqueiro, D., Grimm, D., Schwab, R., Becker, C., et al. (2015). Genome-wide Analysis of Local Chromatin Packing in *Arabidopsis thaliana*. *Genome Res.* 25, 246–256. doi:10.1101/gr.170332.113

Wang, H., Li, S., Li, Y. a., Xu, Y., Wang, Y., Zhang, R., et al. (2019). MED25 Connects Enhancer-Promoter Looping and MYC2-dependent Activation of Jasmonate Signalling. *Nat. Plants* 5, 616–625. doi:10.1038/s41477-019-0441-9

Wang, M., Tu, L., Lin, M., Lin, Z., Wang, P., Yang, Q., et al. (2017). Asymmetric Subgenome Selection and Cis-Regulatory Divergence during Cotton Domestication. *Nat. Genet.* 49, 579–587. doi:10.1038/ng.3807

Wang, M., Wang, P., Lin, M., Ye, Z., Li, G., Tu, L., et al. (2018a). Evolutionary Dynamics of 3D Genome Architecture Following Polyploidization in Cotton. *Nat. Plants* 4, 90–97. doi:10.1038/s41477-017-0096-3

Wang, Q., Sun, Q., Czajkowsky, D. M., and Shao, Z. (2018b). Sub-kb Hi-C in *D. melanogaster* Reveals Conserved Characteristics of TADs between Insect and Mammalian Cells. *Nat. Commun.* 9, 188–8. doi:10.1038/s41467-017-02526-9

Wang, S., Su, J.-H., Beliveau, B. J., Bintu, B., Moffitt, J. R., Wu, C.-t., et al. (2016). Spatial Organization of Chromatin Domains and Compartments in Single Chromosomes. *Science* 353, 598–602. doi:10.1126/science.aaf8084

Wear, E. E., Song, J., Zynda, G. J., LeBlanc, C., Lee, T.-J., Mickelson-Young, L., et al. (2017). Genomic Analysis of the DNA Replication Timing Program during Mitotic S Phase in maize (*Zea mays*) Root Tips. *Plant Cell* 29, 2126–2149. doi:10.1105/tpc.17.00037

Wheeler, G., and Tyler, K. M. (20111808). Widefield Microscopy for Live Imaging of Lipid Domains and Membrane Dynamics. *Biochim. Biophys. Acta (Bba) - Biomembranes* 1808, 634–641. doi:10.1016/j.bbamem.2010.11.017

Wijchers, P. J., Krijger, P. H. L., Geeven, G., Zhu, Y., Denker, A., Verstegen, M. J. A. M., et al. (2016). Cause and Consequence of Tethering a SubTAD to Different Nuclear Compartments. *Mol. Cell* 61, 461–473. doi:10.1016/j.molcel.2016.01.001

Wong, X., Cutler, J. A., Hoskins, V. E., Gordon, M., Madugundu, A. K., Pandey, A., et al. (2021). Mapping the Micro-proteome of the Nuclear Lamina and Lamina-Associated Domains. *Life Sci. Alliance* 4, e202000774. doi:10.26508/lsa.202000774

Wu, S.-M., Tian, Z.-Q., Zhang, Z.-L., Huang, B.-H., Jiang, P., Xie, Z.-X., et al. (2010). Direct Fluorescence *In Situ* Hybridization (FISH) in *Escherichia coli* with a Target-specific Quantum Dot-Based Molecular beacon. *Biosens. Bioelectron.* 26, 491–496. doi:10.1016/j.bios.2010.07.067

Wu, T., Pi, E.-X., Tsai, S.-N., Lam, H.-M., Sun, S.-M., Kwan, Y. W., et al. (2011). GmPHD5 Acts as an Important Regulator for Crosstalk between Histone H3K4 Di-methylation and H3K14 Acetylation in Response to Salinity Stress in Soybean. *BMC Plant Biol.* 11, 178. doi:10.1186/1471-2229-11-178

Wu, Z., Fang, X., Zhu, D., and Dean, C. (2019). Autonomous Pathway: FLOWERING LOCUS C Repression through an Antisense-Mediated Chromatin-Silencing Mechanism. *Plant Physiol.* 182, 27–37. doi:10.1104/pp.19.01009

Xie, T., Zhang, F. G., Zhang, H. Y., Wang, X. T., Hu, J. H., and Wu, X. M. (2019). Biased Gene Retention during Diploidization in Brassica Linked to Three-Dimensional Genome Organization. *Nat. Plants* 5, 822–832. doi:10.1038/s41477-019-0479-8

Xie, T., Zhang, F.-G., Zhang, H.-Y., Wang, X.-T., Hu, J.-H., and Wu, X.-M. (2019). Biased Gene Retention during Diploidization in Brassica Linked to Three-Dimensional Genome Organization. *Nat. Plants* 5, 822–832. doi:10.1038/s41477-019-0479-8

Xu, H., Zhang, S., Yi, X., Plewczynski, D., and Li, M. J. (2020). Exploring 3D Chromatin Contacts in Gene Regulation: the Evolution of Approaches for the Identification of Functional Enhancer-Promoter Interaction. *Comput. Struct. Biotechnol. J.* 18, 558–570. doi:10.1016/j.csbj.2020.02.013

Xu, Z., and Dixon, J. R. (2020). Genome Reconstruction and Haplotype Phasing Using Chromosome Conformation Capture Methodologies. *Brief. Funct. Genomics* 19, 139–150. doi:10.1093/bfgp/elz026

Yang, J., Chang, Y., Qin, Y., Chen, D., Zhu, T., Peng, K., et al. (2020). A Lamin-like Protein OsNMCP1 Regulates Drought Resistance and Root Growth through Chromatin Accessibility Modulation by Interacting with a Chromatin Remodeller OsSWI3C in rice. *New Phytol.* 227, 65–83. doi:10.1111/nph.16518

Yang, R., Hong, Y., Ren, Z., Tang, K., Zhang, H., Zhu, J.-K., et al. (2019). A Role for PICKLE in the Regulation of Cold and Salt Stress Tolerance in Arabidopsis. *Front. Plant Sci.* 10, 900. doi:10.3389/fpls.2019.00900

Ye, H., Rong, Z., and Lin, Y. (2017). Live Cell Imaging of Genomic Loci Using dCas9-SunTag System and a Bright Fluorescent Protein. *Protein Cell* 8, 853–855. doi:10.1007/s13238-017-0460-0

Yokoyama, R., Hirakawa, T., Hayashi, S., Sakamoto, T., and Matsunaga, S. (2016). Dynamics of Plant DNA Replication Based on PCNA Visualization. *Sci. Rep.* 6, 29657. doi:10.1038/srep29657

Yung, W. S., Li, M. W., Sze, C. C., Wang, Q., and Lam, H. M. (2021). Histone Modifications and Chromatin Remodelling in Plants in Response to Salt Stress. *Physiologia Plantarum* 2021, 1–19. doi:10.1111/ppl.13467

Zhang, F., Tang, D., Shen, Y., Xue, Z., Shi, W., Ren, L., et al. (2017). The F-Box Protein ZYGO1 Mediates Bouquet Formation to Promote Homologous Pairing, Synapsis, and Recombination in rice Meiosis. *Plant Cell* 29, 2597–2609. doi:10.1105/tpc.17.00287

Zhang, J., Feng, C., Su, H., Liu, Y., Liu, Y., and Han, F. (2020). The Cohesin Complex Subunit ZmSMC3 Participates in Meiotic Centromere Pairing in maize. *Plant Cell* 32, 1323–1336. doi:10.1105/tpc.19.00834

Zhang, L.-R., Tao, J. Y., and Wang, T. (2004). Molecular Characterization of OsRAD21-1, a rice Homologue of Yeast RAD21 Essential for Mitotic Chromosome Cohesion. *J. Exp. Bot.* 55, 1149–1152. doi:10.1093/jxb/erh123

Zhang, L., Tao, J., Wang, S., Chong, K., and Wang, T. (2006). The rice OsRad21-4, an Orthologue of Yeast Rec8 Protein, Is Required for Efficient Meiosis. *Plant Mol. Biol.* 60, 533–554. doi:10.1007/s11103-005-4922-z

Zhang, X., and Wang, T. (2021). Plant 3D Chromatin Organization: Important Insights from Chromosome Conformation Capture Analyses of the Last 10 Years. *Plant Cell Physiol*, pcab134. doi:10.1093/pcp/pcab134

Zhao, L., Wang, S., Cao, Z., Ouyang, W., Zhang, Q., Xie, L., et al. (2019). Chromatin Loops Associated with Active Genes and Heterochromatin Shape rice Genome Architecture for Transcriptional Regulation. *Nat. Commun.* 10, 3640. doi:10.1038/s41467-019-11535-9

Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., et al. (2006). Circular Chromosome Conformation Capture (4C) Uncovers Extensive Networks of Epigenetically Regulated Intra- and Interchromosomal Interactions. *Nat. Genet.* 38, 1341–1347. doi:10.1038/ng1891

Zheng, H., and Xie, W. (2019). The Role of 3D Genome Organization in Development and Cell Differentiation. *Nat. Rev. Mol. Cell Biol.* 20, 535–550. doi:10.1038/s41580-019-0132-4

Zheng, M., Tian, S. Z., Capurso, D., Kim, M., Maurya, R., Lee, B., et al. (2019). Multiplex Chromatin Interactions with Single-Molecule Precision. *Nature* 566, 558–562. doi:10.1038/s41586-019-0949-1

Zhou, S., Jiang, W., Zhao, Y., and Zhou, D.-X. (2019). Single-cell Three-Dimensional Genome Structures of rice Gametes and Unicellular Zygotes. *Nat. Plants* 5, 795–800. doi:10.1038/s41477-019-0471-3

Zufferey, M., Tavernari, D., Oricchio, E., and Ciriello, G. (2018). Comparison of Computational Methods for the Identification of Topologically Associating Domains. *Genome Biol.* 19, 217. doi:10.1186/s13059-018-1596-9

Deciphering the Role of 3D Genome Organization in Breast Cancer Susceptibility

Brittany Baur[1†], Da-Inn Lee[1†], Jill Haag[2], Deborah Chasman[1], Michael Gould[2‡] and Sushmita Roy[1,3]*

[1]Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI, United States, [2]McArdle Laboratory for Cancer Research, University of Wisconsin-Madison, Madison, WI, United States, [3]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, United States

Cancer risk by environmental exposure is modulated by an individual's genetics and age at exposure. This age-specific period of susceptibility is referred to as the "Window of Susceptibility" (WOS). Rats have a similar WOS for developing breast cancer. A previous study in rat identified an age-specific long-range regulatory interaction for the cancer gene, *Pappa*, that is associated with breast cancer susceptibility. However, the global role of three-dimensional genome organization and downstream gene expression programs in the WOS is not known. Therefore, we generated Hi-C and RNA-seq data in rat mammary epithelial cells within and outside the WOS. To systematically identify higher-order changes in 3D genome organization, we developed NE-MVNMF that combines network enhancement followed by multitask non-negative matrix factorization. We examined three-dimensional genome organization dynamics at the level of individual loops as well as higher-order domains. Differential chromatin interactions tend to be associated with differentially up-regulated genes with the WOS and recapitulate several human SNP-gene interactions associated with breast cancer susceptibility. Our approach identified genomic blocks of regions with greater overall differences in contact count between the two time points when the cluster assignments change and identified genes and pathways implicated in early carcinogenesis and cancer treatment. Our results suggest that WOS-specific changes in 3D genome organization are linked to transcriptional changes that may influence susceptibility to breast cancer.

Keywords: window of susceptibility, breast cancer, 3D genome organization, gene regulation, matrix factorization

INTRODUCTION

A major goal of breast cancer research is to prevent cancer. Breast cancer susceptibility by environmental exposure is modulated by an individual's genetics and age at exposure. For example, environmental or diagnostic radiation exposure poses a high risk to women in early childhood to young adult stage and is significantly reduced starting in the mid-30s (Terry et al., 2019). This age-specific period of high susceptibility is referred to as the window of susceptibility (WOS). Large scale consortia efforts in breast cancer research have significantly advanced our ability to identify genomic loci and molecular pathways that contribute to breast cancer susceptibility (Koboldt et al., 2012; Welter et al., 2014). However, the gene regulatory mechanisms in the WOS remain poorly characterized.

Three-dimensional (3D) organization of the genome, which defines how the DNA is packaged inside the nucleus has emerged as a major component of the gene regulation machinery in mammalian genomes (Bonev and Cavalli, 2016). Three-dimensional genome organization enables long-range interactions between distal regulatory sequences, such as enhancers, and target gene(s) through chromosome looping that brings the regulatory element in close spatial proximity to the target gene. In addition to looping patterns, the chromatin is organized into high-order structural units such as topologically associating domains (TADs) within the cell (Dixon et al., 2012; Hou et al., 2012; Sexton et al., 2012). TADs refer to groups or clusters of genomic regions that preferentially interact among themselves (Bonev and Cavalli, 2016; Rowley and Corces, 2018).

Changes in 3D genome organization, both at the loop and the TAD levels, have been associated with developmental and disease processes (Chakraborty and Ay, 2018; Zheng and Xie, 2019). In particular, genome-wide chromatin looping has been shown to occur in a stage-related manner in the developing limb (Andrey et al., 2017) and in blood cell differentiation (Javierre et al., 2016). TAD-level changes have been associated with timepoint-specific regulatory interactions during differentiation and development (Hug and Vaquerizas, 2018; Paulsen et al., 2019; Zheng and Xie, 2019). Disruptions in TADs have also been associated with numerous diseases including cancer. Delayed replication of large genes near TAD boundaries underlies common fragile sites, hotspots of chromosome instability in cancer (Sarni et al., 2020). Furthermore, disruptions to the TAD-level interaction patterns have been implicated in oncogenesis (Hnisz et al., 2016; Taberlay et al., 2016; Rhie et al., 2019).

Genome architecture has been implicated in cancer susceptibility due to environmental factors (Henning et al., 2016; García-Nieto et al., 2017). For example, lamina-associated heterochromatin at the nuclear periphery is more susceptible to ultraviolet radiation, an environmental carcinogen that causes skin cancer, compared to accessible euchromatin (García-Nieto et al., 2017). At the individual loop level, a 8.5 kb regulatory element, called the temporal control element (TCE) was shown to interact with the *Pappa* gene via long-range chromatin looping of 517 kb (Henning et al., 2016) in both breast cancer resistant rats and susceptible rats. This element lies within the 170 kb *mammary cancer susceptibility (Mcs5c)* locus, a gene desert on rat chromosome 5 which is conserved in the human genome. Furthermore, this interaction was dependent upon the age of the rat, being stronger in young rats (in WOS) versus old rats. Correspondingly, *Pappa* expression was increased in susceptible rats compared to resistant rats within WOS and there was no difference between the two alleles in the adult phase. The *Pappa* gene is a breast cancer-associated gene, which positively regulates the IGF signaling pathway and is important for normal mammary gland development. This is the first validated example of WOS-specific chromatin looping. However, the contribution of the loops and TADs on a genome-wide scale to breast cancer risk from enviromental factors in the window of susceptibility is poorly understood.

To gain mechanistic insight into age-dependent, WOS-specific chromatin looping on a genome-wide scale, we generated Hi-C (Lieberman-Aiden et al., 2009) and RNA-seq data for rats within WOS and outside WOS. We compared the temporal changes in looping to those in expression and found that genes up-regulated within the WOS are associated with interactions that are higher within WOS, and a similar trend exists for genes outside the WOS. We developed a computational approach that combined network enhancement and non-negative matrix factorization (NMF) to identify "dynamic" blocks representing larger scale topological changes between the two time points. We found that network enhancement was important for reliable detection of dynamic blocks, many of which harbored genes implicated in cancer-related pathways and processes. Finally, we mapped human breast cancer GWAS SNPs to loci in rat and found conserved interactions with genes between human and rat. Taken together, these results identified individual loop level and larger-scale topological differences between within-WOS and outside-WOS, many of which are related to transcriptional differences.

METHODS

Tissue Collection and Hi-C Assay

Fresh mammary glands from the abdominal/inguinal regions of 6-week-old and 12-week-old female mammary cancer susceptible Wistar-Furth rats were individually collected, scissor minced and digested for 2 h at 37°C in 10 ml of GIBCO Dulbecco's modified Eagle's medium/F12 (DMEM/F12; ThermoFisher) containing 0.005 g/ml of type 3 collagenase (Worthington-Biochem). Centrifugation was used to remove fat and collect the cell pellets. Individual cell pellets were washed and resuspended in DMEM/F12 media. Each cell suspension was filtered using 40 μm nylon to enrich the mammary ductal fragments and remove stromal cells. The filter was inverted and rinsed to collect the fragments, and the resulting cell pellet containing mammary epithelial cells (MECs) was diluted in PBS and treated for 10 min with 1.5% formaldehyde for DNA/chromatin fixation. After a series of washes, the final cell pellets were collected using centrifugation and stored at −80°C. A total of 6 samples were sent to Arima Genomic, Inc. ($n = 3$ for 6-week-old and $n = 3$ for 12-week-old) for Hi-C analysis, consisting of complete sample processing for Hi-C and library preparation and Illumina Next-Generation sequencing.

Hi-C Data Processing and Differential Interactions

Hi-C data was generated with ~430 M reads per replicate. Hi-C reads were processed using HiC-Pro version 2.7 (Servant et al., 2015) with the default BowTie2 parameters (--very-sensitive -L 30 --score-min L,-0.6,-0.2 --end-to-end–reorder) and aligned to the rn6 genome and 10 kb contact maps were generated. The 6 and 12 weeks samples were aggregated to one 6 weeks and one 12 weeks contact count matrix, respectively. HiC-Pro's implementation of ICE normalization (Imakaev et al., 2012) with default parameters was performed on the two resulting Hi-C matrices.

In order to determine a set of differential chromatin interactions (DCIs), we used both Selfish (Ardakany et al., 2019) and Fit-Hi-C (Ay et al., 2014). Selfish uncovers DCIs between two contact maps directly using a novel self-similarity measure (Ardakany et al., 2019). We obtained 453,513 differential interactions from Selfish (p-value cutoff 10^{-4}). We also used Fit-Hi-C with one pass and a mappability threshold of 1 to determine significant interactions (q-value < 0.05) within WOS and outside WOS. We took differential interactions as those that were significant in one but not the other, resulting in 1,306,601 interactions. We took the union of the resulting FitHiC and Selfish interactions to generate a total of 1,447,082 interactions. We filtered these interactions by computing the mean and standard deviation for all within WOS DCI and all outside WOS DCIs separately for each distance bin (bins at 50 kb intervals from 0 to 1 Mb). Only differential interactions with a z-score greater than one and a distance equal to or less than 1 Mb were considered, for a final set of 1,072,652 interactions. The Fit-Hi-C approach tended to yield pairs with greater differences at longer distances while Selfish tended to yield pairs with greater differences at shorter differences (**Supplementary Figure S1**).

Tissue Collection, RNA Extraction and RNA-seq Experiments

To examine the transcriptional differences associated with WOS on a genome-wide scale, we measured gene expression levels in the MEC from 6-week-old (entering WOS) and 10-week-old (exiting WOS) susceptible rats ($n = 6$ for 6-week-old, $n = 7$ for 10-week-old) using RNA-seq following a similar protocol as described in Henning et al. (2016). Briefly, mammary glands were removed, minced, digested with collagenase, followed by differential centrifugation to collect mammary ductal organoids, which are mainly composed of epithelial cells along with stromal fibroblasts and immune cells. To isolate RNA, cells were homogenized in TRI Reagent (Ambion), followed by RNA extraction using the MagMAX-96 for Microarrays Total RNA kit (Ambion). RNA was extracted using the RNeasy Mini Kit (Qiagen). The fastq files were processed by the UW Biotech center. Counts were obtained using RSEM v1.2.22 (Li and Dewey, 2011).

For all samples, we calculated Transcripts per Million (TPM) for 14,792 genes in the rat genome using RSEM v1.2.22 (Li and Dewey, 2011). We applied several algorithms to determine differential expression: DESeq (Love et al., 2014), EBSeq (Leng et al., 2013) and EdgeR (Robinson et al., 2010). EBSeq was the most conservative with 461 genes. DESeq (3401) and EdgeR (2547) had an intersection of 2071 genes (**Supplementary Figure S2**). We therefore took EBSeq plus the intersection of DESeq and EdgeR as the total set of 2533 differentially expressed (DE) genes.

Network Enhancement and Multiview Non-Negative Matrix Factorization

We developed the NE-MVNMF approach to analyze multiple Hi-C datasets. NE-MVNMF applies Network Enhancement (NE) followed by Multiview Non-negative Matrix Factorization (MVNMF) (Liu et al., 2013) to our Hi-C datasets.

Network Enhancement (NE) is a method for denoising a biological network (Wang et al., 2018). We consider a Hi-C dataset as a weighted network of genomic regions, where each node in the network corresponds to each genomic region and the weighted edges connecting the nodes represent the interaction frequency between genomic regions. NE takes a noisy Hi-C matrix as input and applies iterative graph diffusion process to strengthen edge weights that are well-supported by strong neighboring edges and weaken poorly supported edges. The output of NE is a denoised, enhanced, symmetric matrix which can be used as input to the next step in our pipeline, MVNMF.

Multiview Non-negative Matrix Factorization (MVNMF) is a multi-task non-negative factorization (NMF) method which allows us to find a common underlying structure in multiple matrices (Liu et al., 2013), each task corresponding to a matrix. MVNMF does this by finding low-dimensional factors of multiple matrices such that the factors are regularized towards a common consensus. These factors can then be used as latent features for clustering to reveal the underlying shared or divergent structure in the data. Formally, given $t \in \{1, \dots, T\}$ tasks, each with input matrix $X^{(t)} \in \mathbb{R}_{\geq 0}^{n_t \times m}$, the objective is to find task-specific factors $U^{(t)} \in \mathbb{R}_{\geq 0}^{n_t \times k}$, $V^{(t)} \in \mathbb{R}_{\geq 0}^{m \times k}$ and the consensus factor $V^{(c)} \in \mathbb{R}_{\geq 0}^{m \times k}$ such that:

$$min_{U^{(t)},V^{(t)},V^{(c)}} \sum_{t=1}^{T} \left\| X^{(t)} - U^{(t)}V^{(t)T} \right\|_{\mathrm{F}}^2 + \alpha \left\| V^{(t)} - V^{(c)} \right\|_{\mathrm{F}}^2$$

Here k, is the number of factors or reduced dimensions and is much smaller than n_t or m. The regularization term, α, will constrain factor $V^{(t)}$ of task t to be similar to the consensus $V^{(c)}$. Liu et al. originally proposed an iterative multiplicative update algorithm for MVNMF. However, multiplicative updates algorithm is often slow to converge. Therefore, we implemented an algorithm that optimizes this objective using hierarchical alternating least squares (HALS) with convergence guarantee to a local minimum (Kim et al., 2014).

In our application of MVNMF, we have two tasks, each corresponding to an input Hi-C matrix at 10 kb resolution, for each chromosome: one matrix from week 6 and another one from week 12. The rows and columns of this matrix correspond to a 10 kb bin. Since intra-chromosomal Hi-C matrices (as well as their network-enhanced versions) are symmetric, $X^{(t)}$, $U^{(t)}$, $V^{(t)}$, $V^{(c)}$ take on the dimensions of $n \times n$, $n \times k$, $n \times k$, and $n \times k$, respectively.

We use a simple heuristic to pick k, the number of the factors, which also is the number of clusters. Based on our previous work on single-task NMF to Hi-C data, we set k such that the expected size of each cluster is about 1 Mb in length, which corresponds to the average size of TADs (Lee and Roy, 2021). For example, for an input matrix that corresponds to a chromosome of size 10 Mb, we set $k = 10$. Here, we used 56-282 factors to capture TAD like structures, corresponding to the size of the rat chromosomes.

We verified that network enhancement (NE) and downstream NMF does not overcorrect the underlying structure of the input matrix by comparing NMF results on Hi-C data of different depths. Briefly, we first downsampled high-depth Hi-C matrices from the GM12878 cell line (Rao et al., 2014) to four different lower depth levels (equivalent to the read depth of cell lines HMEC, HUVEC, NEHK, and K562 from the same study). We then applied NE to the downsampled matrices, and then applied NMF to the original high-depth matrices, downsampled matrices, and downsampled + NE matrices. When we compared to original Hi-C data, downsampled Hi-C data + NE does not lead to significant differences in the number of regions in each cluster **Supplementary Figure S3A**). However, downsampled + NE does lead to significantly larger number of regions in each cluster when compared to downsampled without NE (t-test p-value < 0.05 for all downsampled depths). Additionally, when compared to the original data, downsampled + NE does not lead to significant differences in the length of contiguous regions with the same cluster assignment (**Supplementary Figure S3B**). Furthermore, we measured the similarity of the clustering results from the high-depth matrices to those from the downsampled matrices, as well as between the high-depth matrices and the downsampled + NE matrices. The cluster similarity was measured with Rand Index, which measures the concordance between a pair of clustering results. Rand Index ranges from 0 to 1; Rand Index value of 1 means all data points found in one cluster in one result are also in one cluster in the other result, and those in distinct clusters in one result are also kept separate in the other result. We find that the cluster similarity between the original high-depth matrices versus the downsampled matrices is comparable to the cluster similarity between the original and NE matrices (**Supplementary Figure S3C**), suggesting that NE does not overcorrect the underlying structure of the data.

MVNMF, like NMF can converge to different local optima. Therefore, we verified the stability of our results to different random initializations. Briefly, we applied MVNMF to chromosomes 7, 11, 15 and 19 on the within-WOS and outside-WOS matrices with 5 different random initialization seeds. We evaluated the stability of the clusters from different random initializations using Rand Index. We measured the Rand Index between every pair of clustering results from different random initialization seeds (**Supplementary Figure S4**). We find that the mean Rand Index across these comparisons is around 0.9 suggesting that the clustering results are stable to the random initialization seeds.

Identification of Static *Versus* Dynamic Blocks

Once we have the factors, we use them to identify genomic regions dynamically changing their interaction profile across tasks, which we refer to as "dynamic blocks." First, we assign all regions to a cluster based on the factors from MVNMF, then find regions whose cluster assignment changes. We take advantage of the fact that column j in $V^{(t)}$ of task t corresponds to the latent feature or column j in $V^{(s)}$ of task s. Since $X^{(t)}$ is symmetric in Hi-C data, either $U^{(t)}$ or $V^{(t)}$ can be

used to define the clusters of regions. Assuming we use $U^{(t)}$, we assign each row i (corresponding to genomic region i) to its most dominant latent feature, $c_i^{(t)} = argmax_{j \in \{1,\ldots,k\}} U^{(t)}[i, j]$, where $U^{(t)}[i, j]$ represents the entry in the i th row and the j th column/latent feature of $U^{(t)}$. We repeat this procedure across all tasks. A dynamic block between task t and s is a contiguous stretch of 10 kb regions, at least 50 kb in length, whose cluster assignment changed between them, i.e., $c_i^{(t)} \neq c_i^{(s)}$. Furthermore, all regions within the block have to have the same cluster ID within a task. Conversely, a static block is one where $c_i^{(t)} = c_i^{(s)}$. To further assess if a dynamic block is indicative of a changing interaction frequencies, we compared the count differences within the block across time points. We expect a dynamic block to exhibit significantly greater count differences compared to static blocks. We further verified these trends using a t-test to assess the difference in counts between time points among regions inside a static block as well as among regions inside a dynamic block.

Human GWAS Study Integration

We downloaded supplementary table S15 from Zhang et al. which contains gene-SNP interactions from the INQUISIT software (Zhang et al., 2020) and supplementary table S5 from Baxter et al. which contains Capture Hi-C SNP-gene associations (Baxter et al., 2018). We used liftOver (Kuhn et al., 2013) to map these SNPs from the human hg19 assembly to a locus in the rn6 rat assembly, with a minimum base overlap ratio of 0.1. Since the position of the SNP in human may not be a SNP in the corresponding lifted over position in rat, we refer to the position as a 'locus' in rat. We intersected the rat locus with differential chromatin interactions (DCIs) by checking if the SNP was within the boundary of either 10 kb bin in the interaction. We mapped the other 10 kb end to a gene if it overlapped any 10 kb bin within the genomic coordinates of the gene provided by Ensembl release 96 (Yates et al., 2020). We referred to resulting gene locus pairs as locus-gene DCIs. We used the common names to match genes from human to rat.

RESULTS

Differential Looping is Associated With Differential Expression of Within WOS Versus Outside WOS Rats

To globally characterize chromatin looping and examine its role in establishing WOS and associated gene expression programs, we generated Hi-C and RNA-seq datasets for rat mammary epithelial cells within WOS (6-weeks) and outside WOS (10-weeks RNA-seq, and 12-weeks Hi-C, **Figure 1A**). Hi-C data was generated with ~430 M reads per replicate. We aggregated reads to 10 kb resolution and used Iterative Correction and Eigen vector decomposition (ICE) (Servant et al., 2015) to normalize the Hi-C matrices from the two time points (Methods). We used ICE for normalization because it is recommended for Fit-Hi-C (Ay et al., 2014), however, our approach is applicable to data from other normalization methods as well (Hu et al., 2012).

FIGURE 1 | Characterizing WOS-specific chromatin interactions and gene expression. **(A)** Hi-C and RNA-seq data was generated to globally characterize the three-dimensional genome organization and transcriptome within and outside the window of susceptibility (WOS). We first characterized these changes at the level of individual interactions or loops. Within WOS, we identified 538,199 differential chromatin interactions (DCIs) across all chromosomes and 534,454 DCIs outside the WOS. Here CIs refer to those pairs with significantly high counts within and/or outside WOS, whereas DCIs refer to interactions exclusively higher in one of within-WOS or outside-WOS context. **(B)** Contact counts for DCIs (red) compared to non-DCI (gray) for within WOS (above *x*-axis) and outside WOS (below *x*-axis) in rat

(Continued)

TABLE 1 | Number of interactions associated with genes differentially up-regulated within WOS and outside WOS.

	# Of within WOS genes (total = 1,358)	# Of interactions assoc. With within WOS genes	# Of outside WOS genes (total = 942)	# Of interactions assoc. With outside WOS genes
FitHiC interactions	1,331	118,169	925	151,845
DCI interactions (total)	1,310	45,385	919	59,158
DCI interactions (within WOS)	1,245	23,591	886	27,257
DCI interactions (outside WOS)	1,262	21,794	907	31,901

We first identified differential chromatin interactions (DCIs) between the WOS and outside the WOS by taking the union of results from two approaches: Selfish (Ardakany et al., 2019) and the difference in significant chromatin interactions (CIs) identified by applying Fit-Hi-C individually to each sample (Ay et al., 2014). We then filtered these DCIs based on a distance-stratified absolute value of z-score of 1.0 (Methods) to focus on the differential interactions with the greatest magnitude of change in and out of the WOS (**Figure 1B**). In total we identified 538,199 DCIs with counts higher in the WOS (within WOS DCIs) and 534,454 DCIs with counts lower in the WOS compared to outside the WOS (outside WOS DCIs). Among the DCIs, we recapitulated several TCE-*Pappa* gene interactions that are higher in the WOS compared to outside the WOS (**Figure 1C**), which is consistent with previous observations that the TCE interacts with *Pappa* in a WOS-dependent manner (Henning et al., 2016). In parallel, we applied DESeq2 (Love et al., 2014), EBSeq (Leng et al., 2013) and EdgeR (Robinson et al., 2010) to identify differentially expressed (DE) genes between the WOS and outside the WOS (FDR corrected p-val < 0.05, Methods). In total we identified 2300 DE genes, 1,358 of which were up-regulated within the WOS and 942 were down-regulated within the WOS compared to outside.

To examine the relationship between differential expression and chromatin organization we linked DE genes, regardless of direction of expression change, to Fit-Hi-C CIs either in and/or out of the WOS (**Figure 1D**; **Table 1**). We computed the average difference in contact count, stratified by distance, for genes upregulated in WOS compared to outside. We found that at all distance bins compared, genes upregulated in WOS have a higher average contact count in WOS compared to outside WOS for significant interactions (t-test p-value = 3.8e-107, **Figure 1D**). Likewise, genes upregulated outside WOS (or downregulated in the WOS) have a higher average contact count outside WOS compared to within WOS. We performed a similar analysis for DCIs that are significant within or outside WOS and found a similar result (**Supplementary Figure S5**). These results show that changes in gene expression between the two time points is in part due to differences in 3D genome organization.

We next examined the biological processes associated with WOS-specific changes. Genes that are up-regulated within the WOS (1,358 genes) are enriched for cell cycle and DNA replication (Hypergeometric test with FDR<0.05, **Figure 1E**). We also examined genes that are up-regulated in the WOS and associated with 5 or more within WOS DCIs (355 genes). We chose this threshold since many DE genes are associated with at least one DCI in WOS (**Table 1**). Genes that are up-regulated within the WOS and are additionally associated with 5 or more DCIs in the WOS are enriched for these processes as well as DNA packaging and conformation. Genes that are up-regulated outside the WOS (942 genes) are enriched for general transcriptional regulation, and RNA metabolism processes. Genes up-regulated outside WOS and interacting with 5 or more outside WOS DCIs (316 genes) were enriched for similar terms. We also found that within WOS upregulated genes with long-range interactions were enriched for similar terms as all upregulated genes within the WOS, while genes upregulated within WOS and interacting with DCIs were enriched with markedly different terms compared to all upregulated genes in the WOS (**Figure 1E**). Taken together, these results suggest that DCIs are likely involved with the regulation of cell cycle and DNA packaging in younger rats, while in mature rats, DCIs may be more involved with maintaining transcriptional control.

Matrix Factorization-Based Approach to Examine Higher Order Organization Dynamics

In addition to changes at the level of individual interactions, higher-order structural changes in chromatin organization within and outside the WOS could be important for molecular changes associated with breast cancer susceptibility. However, identification of higher-order structural changes, such as in TADs, poses two challenges: (1) handling the noise and sparsity in the input Hi-C matrices and (2) the difficulty in matching TADs across datasets or conditions so that changes

FIGURE 2 | Overview of NE-MVNMF. **(A)** The goal of NE-MVNMF is to find higher-order, e.g., TAD-level changes between the two timepoints denoted as red-dashed diamond. **(B)** Steps of NE-MVNMF. First, network enhancement (NE) smooths out the within-WOS and outside-WOS Hi-C matrices. Then MVNMF is applied to jointly factor the two matrices. By clustering the factor matrices, regions that are switching cluster assignments between within and outside WOS can be identified. These contiguous blocks of such regions represent domain-level changes which we call "dynamic blocks." On the dynamic blocks, we do downstream analysis, such as check for association with DCIs or DE genes.

between them can be pinpointed. To address these two challenges, we developed and applied an approach, Network Enhancement with Multi-view Non-negative Matrix Factorization (NE-MVNMF), which first applies network enhancement (NE) to smooth the noisy and sparse Hi-C matrices (Wang et al., 2018) and then performs multi-view non-negative matrix factorization (MVNMF) (Liu et al., 2013) on these matrices to identify large-scale conserved and differential structural units (**Figure 2**, Methods).

The first component of our pipeline, network enhancement (NE) was developed originally for denoising biological networks (Wang et al., 2018) (**Figure 2B**, Step 1). In our application, we treat a Hi-C dataset as a weighted network of chromatin regions: each node in the network corresponds to a region and the edge weights between nodes represent the interaction counts between a pair of regions. NE iteratively enhances edge weights that are well-supported by strong neighboring edges and weakens those that are poorly supported, then outputs a denoised matrix which is then used as input to the next step in our pipeline, MVNMF. We verified that NE does not overcorrect the underlying structure of the input matrix by comparing results on Hi-C data of different depths before and after smoothing (**Supplementary Figure S3**, Methods).

MVNMF combines Non-negative Matrix Factorization (NMF) (Lee and Seung, 2001) with multi-task learning (Zhang and Yang, 2021). NMF is a powerful dimensionality reduction method that can be used to recover interpretable, lower-dimensional patterns from large, high-dimensional data in imaging, text, and biomedical domains (Lee and Seung, 2001). Applying NMF to a Hi-C matrix yields low-dimensional factors or latent features which can be used to cluster the row or the column entities, i.e., genomic regions. These clusters of genomic regions correspond to densely interacting regions of the genome such as topologically associating domains (TADs) (Lee and Roy, 2021). MVNMF extends NMF to a multi-task setting with multiple input matrices, each corresponding to a task or a time point (e.g. within WOS). It jointly factorizes the input matrices such that their lower-dimensional factor representations are similar to a single common factor, and clusters derived from these factors can be matched across tasks (Methods). MVNMF identifies clusters that are matched across tasks. This cluster correspondence across tasks allows us to
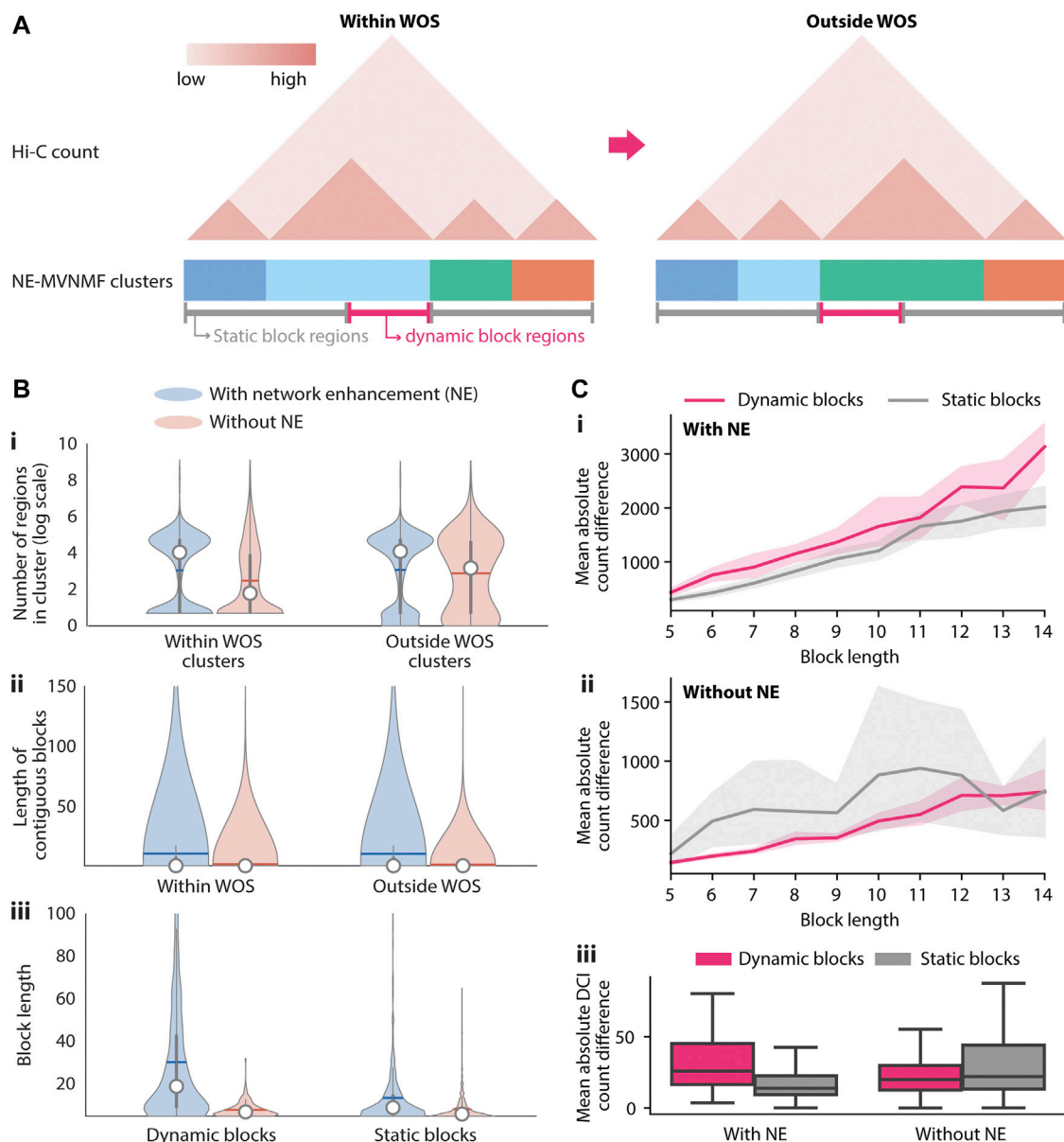
FIGURE 3 | Identification and characterization of dynamic 3D genome blocks with NE-MVNMF. **(A)** Schematic of how dynamic blocks of regions involved in large-scale topological changes are identified from the NE-MVNMF clusters. The NE-MVNMF clusters are depicted at the bottom of the exemplar Hi-C count matrix, with regions in the same cluster to have the same color. Regions in dynamic blocks (magenta line) are regions whose cluster assignment switched between within WOS and outside WOS. Conversely regions in static blocks (gray) are those whose cluster assignment stayed the same. **(B)** Distribution of the number of regions in each cluster within and outside WOS (i), length of contiguous blocks within and outside WOS (ii), and length of contiguous dynamic or static blocks (iii), with and without NE. **(C)** Difference in interaction counts among regions within dynamic blocks and static blocks. Top, for each dynamic or static block, we summed up the absolute value difference between interactions from within WOS and those form outside WOS. We plot the mean of the absolute value difference by block length with NE (i) and without NE (ii). The shaded area represents the 95% confidence interval. iii, Mean absolute value difference for DCIs only is plotted for dynamic and static blocks, with and without NE (iii).

easily identify genomic regions whose cluster assignment has changed in different contexts (**Figure 2B**, Step 2). From these matched clusters, we define a "dynamic block" as a stretch of 5 or more contiguous 10 kb bins (50 kb region) that have a different cluster assignment between a pair of conditions (**Figure 3A**). A "static block" is similarly defined but for contiguous 10 kb bins

that have the same cluster assignment between the conditions compared. Regions that do not have contiguous cluster assignments for 5 or more regions are considered noisy.

We first compared the effect of network enhancement on the ability to detect higher-order topological units and the quality of the dynamic versus static blocks based on different metrics. We

applied both MVNMF and NE-MVNMF at different regularization values ($\alpha \in \{1e5, 1e6, 1e7, 1e8\}$ Methods) to within and outside the WOS Hi-C matrices at 10 kb resolution. Higher regularization values will constrain the factors from each task to be more similar to the consensus factor (Methods). We examined the distribution of the cluster sizes with and without NE (**Figure 3Bi**), the extent to which they were contiguous (**Figure 3Bii**) and the distribution of sizes of static and dynamic blocks (**Figure 3Biii**). When comparing the overall size of the clusters, without NE tends to obtain on average "smaller" units compared to with NE (**Figure 3Bi**). Furthermore, when we consider the number of contiguous regions with the same cluster assignment before reaching a different cluster assignment, the lengths of such contiguous regions are larger for NE versus no NE in both within and outside WOS (**Figure 3Bii**). When comparing the dynamic versus static blocks, we see a similar trend in the size distribution (**Figure 3Biii**): network enhancement and higher levels of regularization tended to increase the number of static blocks and decrease the number of dynamic blocks overall (**Supplementary Table S1**). These results show that network enhancement led to more contiguous larger blocks, which are indicative of less noisy clustering assignments. Finally, we hypothesized that there would be a larger difference in overall contact count in dynamic blocks compared to static blocks. We took the overall difference for each block as the sum of the absolute value of the difference for all contact counts within WOS and outside WOS. We compared these count differences for blocks of different sizes (**Figures 3Ci,ii**) as well as across all blocks (**Figure 3Ciii**). We find that for the same α regularization value, dynamic blocks have a greater overall difference than static blocks when using network enhancement (**Figure 3C** for $\alpha = 1e8$, **Supplementary Figure S6** for $\alpha \in \{1e5, 1e6, 1e7\}$). For blocks with 3 or more regions, the overall difference for dynamic blocks was significantly greater than static blocks for 6 out of the 10 bins (t-test p-value <0.05). When comparing across blocks of all sizes, the count difference of DCIs between static and dynamic blocks was much more dramatic compared to without NE (**Figure 3Ciii**). Taken together, NE-MVNMF allows us to reliably identify regions involved in higher-order topological changes across multiple biological contexts.

NE-MVNMF Reveals Large-Scale WOS-Specific Changes

We next examined the dynamic blocks obtained from NE-MVNMF to gain insight into the 3D genome organizational properties within and outside the WOS. There were 168 dynamic blocks in total across all chromosomes, which ranged in size from 50 to 320 kb. We prioritized blocks for interpretation based on the difference in counts of blocks between the two conditions as well as based on visual inspection. Of the total 168 blocks, we identified 35 blocks that had a significant change in count between the two conditions when considering all pairs of regions in these blocks (T-test and Rank sum p-value <0.05). These blocks ranged from 50 to 180 kb in size and included blocks

spanning genic regions (28) and those spanning non-coding regions (7, **Figure 4** and **Table 2**).

For ease of interpretation, we focused on blocks that harbored genes, regardless of their differential expression status. For example, one block (#31,110 kb, **Figure 4A**), included the genes *Pdlim1* and *Sorbs1*, of which *Sorbs1* exhibited a significantly lower expression within WOS (6 weeks), while *Pdlim1* exhibited a relatively higher, expression in the WOS. *Pdlim1* is expressed in fibroblasts and involved in cell polarity and migration (Stelzer et al., 2016) and has been shown to be associated with breast cancer progression (Liu et al., 2015). *Sorbs1* is involved in signaling pathways and low expression of *Sorbs1* is associated with poor prognosis of breast cancer (Song et al., 2017). Another block (#50, 60 kb, **Figure 4B**) spanned two genes, *Kif18a* and *Mettl15* of which *Kif18a* has a significantly high expression within WOS. *Kif18a* is a kinesin protein involved in chromosomal stability, low expression of kinesin proteins has been associated with cell proliferation of chromosomally unstable genes (Marquis et al., 2021) and is a candidate target for cancer treatment (Sabnis, 2020). Another block harboring a down-regulated gene within the WOS was #38 (50 kb, **Figure 4C**), containing *Dnajc21* a heat shock protein and *Brix1* involved in ribosome biogenesis. Over-expression of heat shock proteins has been associated with a large number of cancers (Calderwood and Gong, 2016). Another block (#78, 90 kb, **Figure 4D**) was associated a number of zinc finger proteins, including *Zfp871*, which was shown to be part of the P53 pathway (Mohibi et al., 2021) and cytochrome P450, an enzyme that metabolizes several pre-carcinogens and is broadly involved in both cancer formation and treatment (Rodriguez-Antona and Ingelman-Sundberg, 2006). Finally, blocks 143 (160 kb, **Figure 4E**) and 162 (80 kb, **Figure 4F**) had several genes that either encoded chromatin remodeling factors (*Supt16h* and *Chd8*), genes representing families often mutated in cancers (*Tox4*, Yu and Li, 2015), genes that have been implicated as oncogenes as well as tumor suppressors (*Sall2*, Hermosilla et al., 2017; *Mettl3*, Zeng et al., 2020) and involved in glycosylation (*Large*1, block162) which is used as a marker and offers novel therapeutic targets (Costa et al., 2020). Overall, our analysis identified dynamic blocks that harbored genes implicated in cancer and related pathways including chromosomal stability, ribosome biogenesis and stress response.

WOS-Specific Looping can be Leveraged to Examine Regulatory Variation

Many studies have identified disease associated variants inside distal regulatory elements that loop to genes, for example, in autoimmune disorders (Javierre et al., 2016) and cancer (Zhang et al., 2019) including breast cancer susceptibility (Baxter et al., 2018). What is less explored is the context-specificity and timing of these long-range interactions, which can impact when a variant modulates a target gene's expression. The Temporal Control Element (TCE) interaction with the *Pappa* gene is an example of a time window-specific interaction and is present in young rats (within WOS), but not older rats (Henning et al., 2016). In the susceptible genotype, *Pappa* expression levels are increased
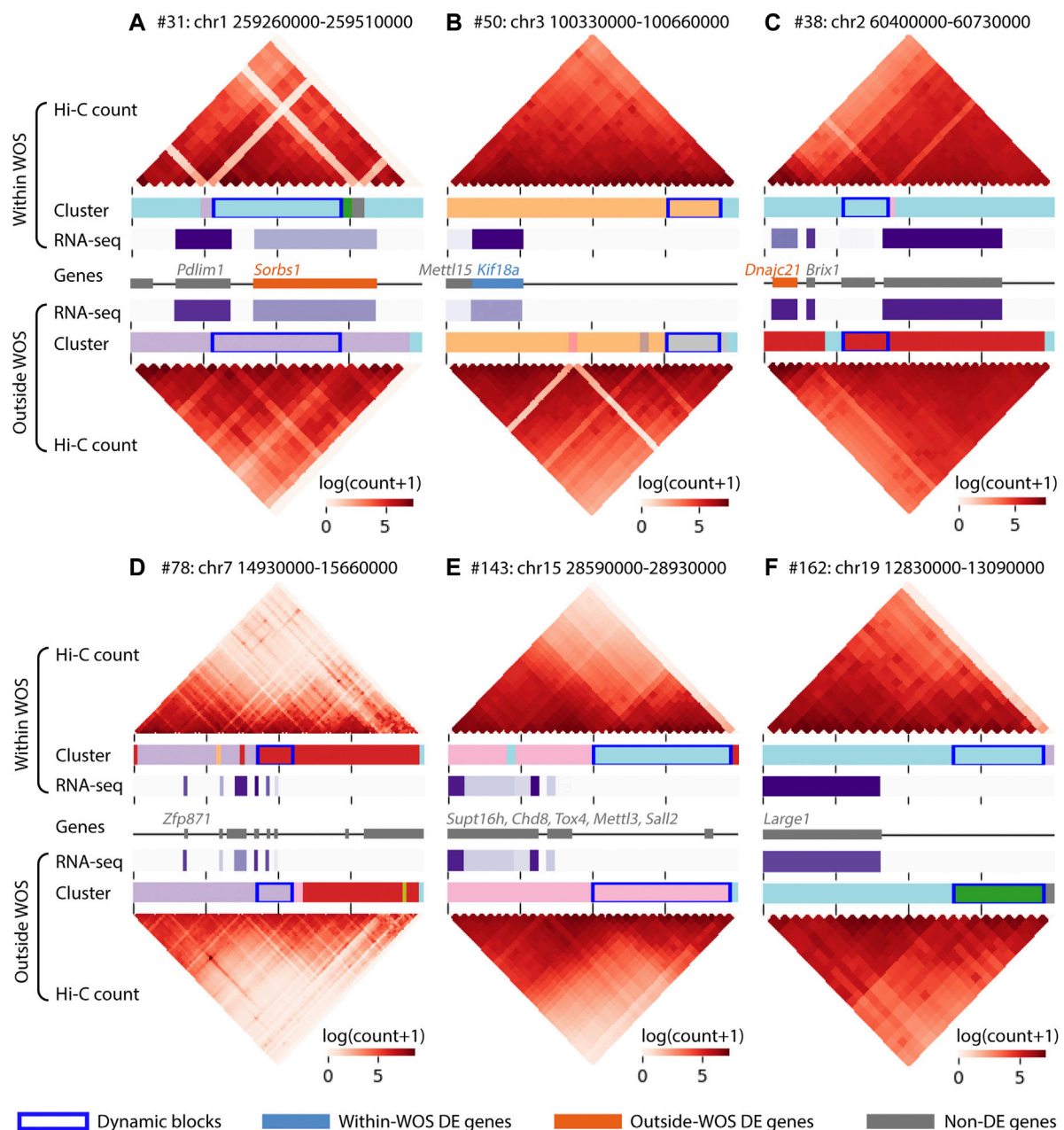
FIGURE 4 | Visualization of regions surrounding 6 dynamic blocks of interest in the window of susceptibility. Within each panel, the top and the bottom heatmaps visualize the interaction counts from within and outside WOS, respectively. The horizontal bars associated with Cluster below each heatmap are colored by cluster ID; dynamic blocks are highlighted with a dark blue box outline. Gene expression (RNA-seq in TPM) is visualized in a horizontal purple heatmap, with darker purple representing higher expression. Finally, the gene track in the middle denotes gene locations; within-WOS and outside-WOS DE genes are colored with blue and orange, respectively. Gray indicates the gene is not DE. The dynamic blocks of interest are found within **(A)** #31, chr1 259260000-259510000, **(B)** #50, chr3 100330000-100660000, **(C)** #38, chr2 60400000-60730000, **(D)** #78, chr7 14930000-15660000, **(E)** #143, chr15 28590000-28930000, **(F).** #162, chr19 12830000-13090000.

relative to the resistant genotype leading to increased breast cancer susceptibility, indicating a genotype-specific effect on gene expression. A similar model could underlie other SNPs associated with breast cancer susceptibility, where the SNP occurs in an enhancer region that loops to regulate a gene's expression in a condition-specific manner (e.g. in the WOS but not outside the

WOS). The SNP may disrupt the binding site of a transcription factor which may result in aberrant expression of the target gene, but the loop itself is operational only in a particular condition. It is also possible that the SNP effects the loop strength, for example by perturbing the binding site of an architectural protein, e.g. CTCF, which may affect the regulation of a gene (de Wit et al.,

	All blocks	Significant (t-test and rank sum)
Total	168	35
Genic	95	28
Non-genic	73	7

2015; Tang et al., 2015; Guo et al., 2018). The TCE-*Pappa* interaction was conserved in human and rat. Therefore, we asked if we could examine additional variants associated with breast cancer for their participation in condition-specific long-range interactions (**Figure 5**).

We considered two studies for this problem, one that mapped SNPs to genes using a computational tool, INQUISIT (Zhang et al., 2020), and two, that used Capture-Hi-C (Baxter et al., 2018). We obtained 26 SNPs from a recent human breast cancer GWAS study that were linked to potential target genes in 201

interactions using INQUISIT (Zhang et al., 2020). We mapped these SNPs to loci in rn6 using liftOver and then identified target genes with the DCIs (Methods), (Kuhn et al., 2013). Since the lifted over position in rat likely does not correspond to a SNP in rat, we refer to the interactions as SNP-gene interactions in human and locus-gene interactions in rat. A total of 11 SNPs mapped to a locus in rat corresponding to a total of 101 human SNP-gene associations. Of these 11 SNPs, we identified 15 locus-gene DCIs, connecting 6 SNPs and 9 genes in total across these interactions (Zhang et al., 2020) (**Table 3**). Of these interactions, 7 locus-gene DCIs were within WOS (5 SNPs, 5 genes) and 8 locus-gene DCIs were outside WOS (4 SNPs, 6 genes). Of the 5 genes connected to within WOS locus-gene interactions, *Jag1* is differentially up-regulated within WOS (**Figure 5A**). *Jag1* is part of the notch signaling pathway involved in the renewal of stem and progenitor cells in mammary glands and has been associated with poor overall survival in breast cancer (Reedijk et al., 2005). Of the 6 genes connected to outside WOS locus-gene interactions, *Trps1* and *Phf1* are differentially up-regulated outside the WOS. Knockdown of *Trps1* results in reduced tumor growth in shRNA
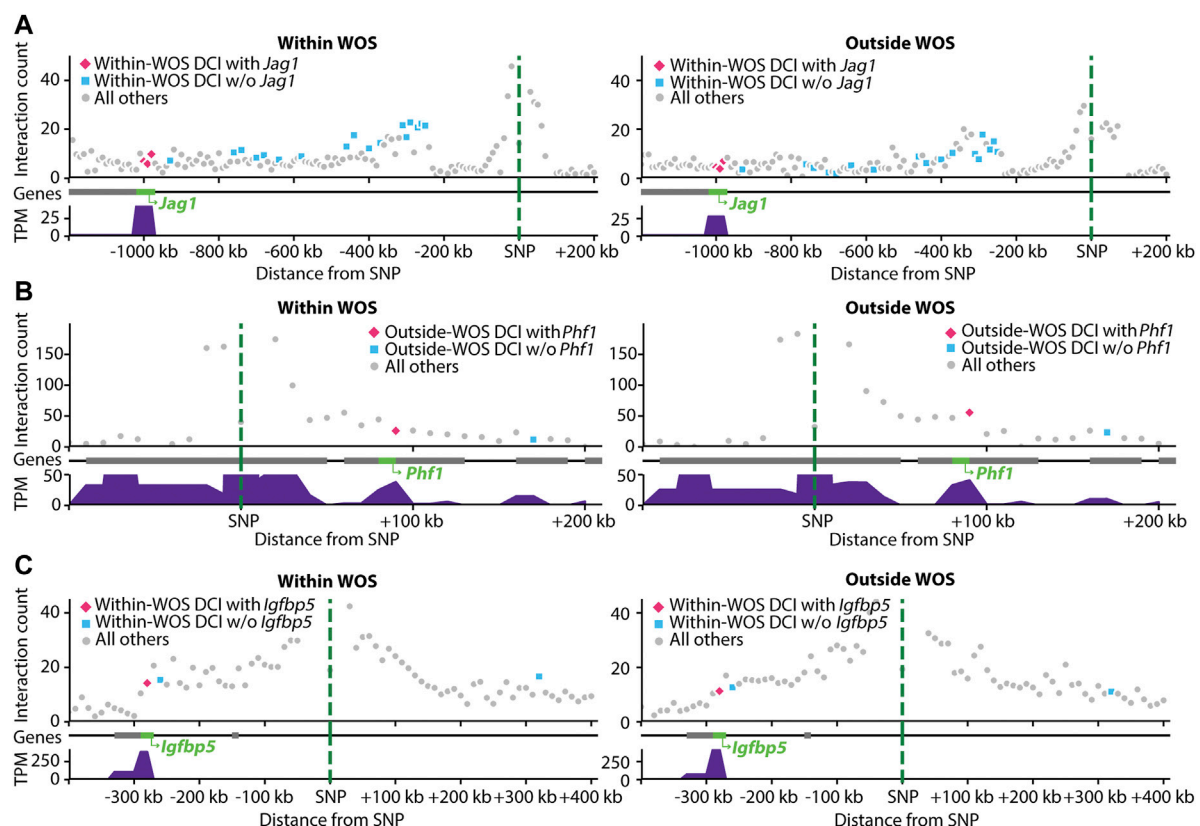
FIGURE 5 | Visualization of interactions involving SNPs associated with human breast cancer mapped to rat genome. Within each panel, shown are the interaction counts from within WOS **(left)** and outside WOS **(right)** between the mapped-SNP region (green dotted line) and neighboring regions by distance. Pink dots are DCIs associated with the gene of interest; blue dots are other DCIs; gray dots are all others. Gene tracks denote gene locations; the gene of interest is highlighted in green. Gene expression (in TPM) is plotted below the gene tracks. **(A)** Interactions involving human breast cancer SNP (chr20 11502618 A- > AAC) mapped to a region in rat. The region is involved in within-WOS DCI with the gene *Jag1*. **(B)** Interactions involving human breast cancer SNP (chr6 33239869 C- > T) mapped to a region in rat. This region is involved in outside-WOS DCI with gene *Phf1*. **(C)** Interactions involving human breast cancer SNP (chr3 156535958 AT- > A) mapped to a region in rat. This region is involved in within-WOS DCI with gene *Igfbp5*.

TABLE 3 | Rat loci-gene interactions that recapitulate human breast cancer GWAS SNP-gene interaction from Zhang et al. (2020). Shown is the name of the human SNP, the rat 10 kb bin that has the gene and the variant of the conserved loop. The human SNP is named to show the chromosome ID and the genomic coordinate.

Gene name	Human SNP	Chr	GeneContainingBin	SNPContainingBin	Within or outside WOS
Hnf1a	rs17215231	12	47430000	47420000	Within
Trps1	rs13277568	7	90150000	90310000	Within
Jag1	rs141526427	3	130090000	131090000	Within
Jag1	rs141526427	3	130100000	131090000	Within
Jag1	rs141526427	3	130110000	131090000	Within
Lekr1	rs34052812	2	157570000	157440000	Within
Mrps35	chr12_29140260	4	181430000	182340000	Within
Phf1	rs2464195	20	5530000	5440000	Outside
Cuta	rs2464195	20	5530000	5440000	Outside
Sppl3	rs17215231	12	47310000	47420000	Outside
Hnf1a	rs17215231	12	47410000	47420000	Outside
Trps1	rs13277568	7	90190000	90310000	Outside
Trps1	rs13277568	7	90200000	90310000	Outside
Trps1	rs13277568	7	90300000	90310000	Outside
Tiparp	rs34052812	2	157340000	157440000	Outside

TABLE 4 | Rat loci-gene interactions that recapitulate human breast cancer GWAS SNP-gene interaction from Baxter et al. (2018), identified with Capture-Hi-C. Shown is the name of the human SNP, the rat 10 kb bin that has the gene and the variant of the conserved loop.

Gene name	Human SNP	Chr	GeneContainingBin	SNPContainingBin	Within or outside WOS
Igfbp5	rs13387042	9	80160000	80440000	Within
Zmiz1	rs704010	16	1890000	1760000	Outside
Zmiz1	rs704010	16	1930000	1760000	Outside
Gtpbp3	rs8170	16	19900000	19790000	Outside
Rpl37a	rs16857609	9	80000000	80810000	Outside
Olfml3, Hipk1	rs11552449	2	206220000	206290000	Outside
Ovol1	rs3903072	1	220930000	220910000	Outside

screens and *Trps1* has been shown to repress transcription by interacting with multiple components of the nucleosome remodeling deacetylase complex (Witwicki et al., 2018). *Phf1* is part of the polycomb group of proteins that maintain repressive chromatin states and has been shown to be an activator of the p53 signaling pathway (**Figure 5B**) (Yang et al., 2013). p53 is a tumor suppressor that regulates cell growth and apoptosis (Yang et al., 2013). Both *Trps1* and *Phf1* have tumor suppressor properties, are associated with repressive chromatin and are up-regulated outside WOS.

We performed a similar analysis by leveraging a study that used Capture Hi-C to link breast cancer GWAS SNPs to genes (Baxter et al., 2018). The study investigated 41 breast cancer GWAS SNPs connected to genes. Of these 41, 16 mapped to a locus in rat and participated in 63 SNP-gene interactions in the human Capture Hi-C data. Seven of these SNP-gene interactions mapped to a corresponding locus-gene DCI in rat, one in a within WOS DCI and six in outside WOS DCIs (5 SNPs, 6 genes, **Table 4**). For the within WOS DCI, the locus corresponded to human SNP rs13387042 and gene *Igfbp5* (**Figure 5C**). *Igfbp5*, like the WOS-associated gene *Pappa* gene, is involved in IGF signaling and plays an important role in mammary development (Wyszynski et al., 2016). The interaction between

rs13387042 and *Igfbp5* is supported by previous studies in humans (Ghoussaini et al., 2014; Baxter et al., 2018). This result suggests that the mechanism by which variants interact with the *Igfbp5* promoter may be related to WOS.

Of the 6 genes that are interacting with SNP-associated loci outside WOS, two are differentially expressed, *Ovol1* (up-regulated outside WOS) and *Olfml3* (up-regulated inside WOS). *Ovol1* has been shown to induce mesenchymal to epithelial transition in human cancers (Roca et al., 2013) and is associated with rs3903072. This association is also supported by human eQTL studies in breast cancer, suggesting that rs3903072 may alter *Ovol1* expression (Li et al., 2014). Targeting *Olfml3* has been shown to suppress tumor growth and angiogenesis (Stalin et al., 2021). Of the non-DE genes, *Zmiz1* is a prognostic marker of multiple cancer types (Mathios et al., 2019), *Rpl37a* is a biomarker for response to neoadjuvant chemotherapy in non-metastatic locally advanced breast cancer (Carrara et al., 2021), and *Hipk1* has been shown to act as a tumor suppressor by activating p53 (Rey et al., 2013). In the outside WOS DCI locus-gene interactions, *Hipk1* interacts with rs11552449. This interaction is also supported in human follicular helper T cell Capture Hi-C data (Su et al., 2020). Infiltration of follicular helper T cells has also been shown to

predict breast cancer survival (Gu-Trantien et al., 2013). Overall, we were able to recover several human SNP locus-gene interactions in our dataset, which connected to genes implicated in cancer. The conservation of these long-range interactions in human and rat enable leveraging our dataset to study the human loci in rat as a model system.

DISCUSSION

The window of susceptibility (WOS) of breast cancer is an important period during which cancer risk due to environmental exposure is higher in women. The three-dimensional organization of the genome likely plays an important role in the transcriptional programs underlying the early stages of carcinogenesis (Henning et al., 2016; García-Nieto et al., 2017). However, little is known about these mechanisms within the WOS and how it differs outside the WOS. Here, we generated unique Hi-C and RNA-seq datasets for rats in and outside WOS and developed a computational approach, NE-MVNMF, that can unravel these differences.

Dynamics in three-dimensional genome organization can be studied at the level of individual loops as well as higher-order organizational units. However, the immediate impact on downstream gene expression due to these changes remains debated (van Steensel and Furlong, 2019). We demonstrated that differential chromatin interactions (DCIs) are associated with transcriptional differences between within WOS and outside WOS. Upregulated genes associated with differential interactions, which are higher in strength within WOS are specifically enriched for cell-cycle related terms compared to all up-regulated genes or genes associated with DCIs with counts higher outside the WOS. The cell cycle has been implicated in breast cancer susceptibility (Deng, 2006) and is often deregulated in breast cancer (Bower et al., 2017). Our results suggests that long-range regulation or deregulation of cell cycle genes could be important avenues for functional studies of breast cancer susceptibility.

A significant challenge in studying dynamics in 3D genome organization is detecting reliable changes between time points. This is difficult because of high sparsity of the data. To address this challenge, we developed a multi-view NMF approach with network enhancement to first enhance the Hi-C signal followed by identification of large-scale topological changes within WOS and outside WOS. The network enhancement smooths the matrix, which strengthens well-supported interactions and weakens poorly supported interactions. This allows MV-NMF to be more robust to noise and bias. Our results show that network enhancement to smooth the matrices before NMF leads to the identification of dynamic blocks that have larger changes in contact count overall and specifically larger changes in DCIs compared to static blocks. Closer inspection of dynamic blocks revealed many genes that are involved in mammary development and cancer-associated pathways.

We previously identified a WOS-specific interaction between the TCE and the *Pappa* gene (Henning et al., 2016). This interaction is conserved across human, rat and mouse. Therefore, we asked if we can identify similar conserved interactions by mapping human SNP-gene interactions to rat, which can then be followed up with in-depth molecular characterization in a model organism. We identified several examples of conserved locus-gene interactions by comparing our DCIs to two previous studies connecting breast cancer susceptibility SNPs to genes in human (Baxter et al., 2018; Zhang et al., 2020). For example, the SNP rs13387042, which falls in an enhancer region in human loops over a distance of 400 kb to *Igfbp5* (Wyszynski et al., 2016). We were able to map this locus onto a DCI within WOS rats connected to the rat ortholog of *Igfbp5*. Notably, similar to the previously validated WOS-associated *Pappa* gene, *Igfbp5* is also involved in mammary development and IGF signaling. This interaction, along with the other interactions identified in this study, will be a valuable resource for enabling deeper characterization of genetic variation and breast cancer that may have a similar age-specific window of susceptibility.

Our work can be extended in several ways. First, the addition of more time points would be useful in identifying more fine-grained dynamics of chromatin for entry and exit from the WOS. Second, the addition of one-dimensional regulatory signals would be beneficial in determining which enhancers and promoters are active within and outside WOS. In general, a more robust dataset can aid in gaining a more complete picture of the molecular mechanisms underlying WOS. On the methodological side, our approach could be extended to identify more complex patterns of change in 3D genome organization to handle more time points and heterogeneous samples. Taken together, our transcriptomic and 3D genome profiles of within WOS and outside WOS and our computational pipeline should be a useful resource for studying the role of 3D genome organization in the window of susceptibility for breast cancer.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the GEO database with the accession number GSE184285 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE184285). Source code (C++) for MVNMF is available here: http://github.com/Roy-lab/mvnmf.

ETHICS STATEMENT

The animal study was reviewed and approved by University of Wisconsin–Madison School of Medicine and Public Health Animal Care and Use Committee.

AUTHOR CONTRIBUTIONS

MG and SR conceived of the project. JH prepared the biological samples. BB and DC processed and analyzed the data. D-IL

developed the MVNMF pipeline. BB and D-IL analyzed the performance of NE-MVNMF and MVNMF. BB, D-IL, SR wrote the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.788318/full#supplementary-material

REFERENCES

Andrey, G., Schöpflin, R., Jerković, I., Heinrich, V., Ibrahim, D. M., Paliou, C., et al. (2017). Characterization of Hundreds of Regulatory Landscapes in Developing Limbs Reveals Two Regimes of Chromatin Folding. *Genome Res.* 27, 223–233. doi:10.1101/gr.213066.116

Ardakany, A. R., Ay, F., and Lonardi, S. (2019). Selfish: Discovery of Differential Chromatin Interactions via a Self-Similarity Measure. *Bioinformatics* 35, i145–i153. doi:10.1093/bioinformatics/btz362

Ay, F., Bailey, T. L., and Noble, W. S. (2014). Statistical Confidence Estimation for Hi-C Data Reveals Regulatory Chromatin Contacts. *Genome Res.* 24, 999–1011. doi:10.1101/gr.160374.113

Baxter, J. S., Leavy, O. C., Dryden, N. H., Maguire, S., Johnson, N., Fedele, V., et al. (2018). Capture Hi-C Identifies Putative Target Genes at 33 Breast Cancer Risk Loci. *Nat. Commun.* 9, 1028. doi:10.1038/s41467-018-03411-9

Bonev, B., and Cavalli, G. (2016). Organization and Function of the 3D Genome. *Nat. Rev. Genet.* 17, 661–678. doi:10.1038/nrg.2016.112

Bower, J. J., Vance, L. D., Psioda, M., Smith-Roe, S. L., Simpson, D. A., Ibrahim, J. G., et al. (2017). Patterns of Cell Cycle Checkpoint Deregulation Associated with Intrinsic Molecular Subtypes of Human Breast Cancer Cells. *Npj Breast Cancer* 3, 1–12. doi:10.1038/s41523-017-0009-7

Calderwood, S. K., and Gong, J. (2016). Heat Shock Proteins Promote Cancer: It's a Protection Racket. *Trends Biochem. Sci.* 41, 311–323. doi:10.1016/j.tibs.2016.01.003

Carrara, G. F. A., Evangelista, A. F., Scapulatempo-Neto, C., Abrahão-Machado, L. F., Morini, M. A., Kerr, L. M., et al. (2021). Analysis of RPL37A, MTSS1, and HTRA1 Expression as Potential Markers for Pathologic Complete Response and Survival. *Breast Cancer* 28, 307–320. doi:10.1007/s12282-020-01159-z

Chakraborty, A., and Ay, F. (2019). The Role of 3D Genome Organization in Disease: From Compartments to Single Nucleotides. *Semin. Cel Dev. Biol.* 90, 104–113. doi:10.1016/j.semcdb.2018.07.005

Costa, A. F., Campos, D., Reis, C. A., and Gomes, C. (2020). Targeting Glycosylation: A New Road for Cancer Drug Discovery. *Trends Cancer* 6, 757–766. doi:10.1016/j.trecan.2020.04.002

de Wit, E., Vos, E. S. M., Holwerda, S. J. B., Valdes-Quezada, C., Verstegen, M. J. A. M., Teunissen, H., et al. (2015). CTCF Binding Polarity Determines Chromatin Looping. *Mol. Cel.* 60, 676–684. doi:10.1016/j.molcel.2015.09.023

Deng, C.-X. (2006). BRCA1: Cell Cycle Checkpoint, Genetic Instability, DNA Damage Response and Cancer Evolution. *Nucleic Acids Res.* 34, 1416–1426. doi:10.1093/nar/gkl010

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012). Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. *Nature* 485, 376–380. doi:10.1038/nature11082

García-Nieto, P. E., Schwartz, E. K., King, D. A., Paulsen, J., Collas, P., Herrera, R. E., et al. (2017). Carcinogen Susceptibility Is Regulated by Genome Architecture and Predicts Cancer Mutagenesis. *EMBO J.* 36, 2829–2843. doi:10.15252/embj.201796717

Ghoussaini, M., Edwards, S. L., Michailidou, K., Nord, S., Cowper-Sal Lari, R., Desai, K., et al. (2014). Evidence that Breast Cancer Risk at the 2q35 Locus Is Mediated through IGFBP5 Regulation. *Nat. Commun.* 4, 4999. doi:10.1038/ncomms5999

Gu-Trantien, C., Loi, S., Garaud, S., Equeter, C., Libin, M., de Wind, A., et al. (2013). CD4+ Follicular Helper T Cell Infiltration Predicts Breast Cancer Survival. *J. Clin. Invest.* 123, 2873–2892. doi:10.1172/JCI67428

Henning, A. N., Haag, J. D., Smits, B. M. G., and Gould, M. N. (2016). The Non-coding Mammary Carcinoma Susceptibility Locus, Mcs5c, Regulates Pappa Expression via Age-Specific Chromatin Folding and Allele-Dependent DNA Methylation. *Plos Genet.* 12, e1006261. doi:10.1371/journal.pgen.1006261

Hermosilla, V. E., Hepp, M. I., Escobar, D., Farkas, C., Riffo, E. N., Castro, A. F., et al. (2017). Developmental SALL2 Transcription Factor: a New Player in Cancer. *Carcinogenesis* 38, 680–690. doi:10.1093/carcin/bgx036

Hnisz, D., Weintraub, A. S., Day, D. S., Valton, A.-L., Bak, R. O., Li, C. H., et al. (2016). Activation of Proto-Oncogenes by Disruption of Chromosome Neighborhoods. *Science* 351, 1454–1458. doi:10.1126/science.aad9024

Hou, C., Li, L., Qin, Z. S., and Corces, V. G. (2012). Gene Density, Transcription, and Insulators Contribute to the Partition of the Drosophila Genome into Physical Domains. *Mol. Cel.* 48, 471–484. doi:10.1016/j.molcel.2012.08.031

Hu, M., Deng, K., Selvaraj, S., Qin, Z., Ren, B., and Liu, J. S. (2012). HiCNorm: Removing Biases in Hi-C Data via Poisson Regression. *Bioinformatics* 28, 3131–3133. doi:10.1093/bioinformatics/bts570

Hug, C. B., and Vaquerizas, J. M. (2018). The Birth of the 3D Genome during Early Embryonic Development. *Trends Genet.* 34, 903–914. doi:10.1016/j.tig.2018.09.002

Imakaev, M., Fudenberg, G., McCord, R. P., Naumova, N., Goloborodko, A., Lajoie, B. R., et al. (2012). Iterative Correction of Hi-C Data Reveals Hallmarks of Chromosome Organization. *Nat. Methods* 9, 999–1003. doi:10.1038/nmeth.2148

Javierre, B. M., Burren, O. S., Wilder, S. P., Kreuzhuber, R., Hill, S. M., Sewitz, S., et al. (2016). Lineage-Specific Genome Architecture Links Enhancers and Non-Coding Disease Variants to Target Gene Promoters. *Cell* 167, 1369–1384.e19. doi:10.1016/j.cell.2016.09.037

Kim, J., He, Y., and Park, H. (2014). Algorithms for Nonnegative Matrix and Tensor Factorizations: a Unified View Based on Block Coordinate Descent Framework. *J. Glob. Optim.* 58, 285–319. doi:10.1007/s10898-013-0035-4

Koboldt, D. C., Fulton, R. S., McLellan, M. D., Schmidt, H., Kalicki-Veizer, J., McMichael, J. F., et al. (2012). Comprehensive Molecular Portraits of Human Breast Tumours. *Nature* 490, 61–70. doi:10.1038/nature11412

Kuhn, R. M., Haussler, D., and Kent, W. J. (2013). The UCSC Genome Browser and Associated Tools. *Brief. Bioinform.* 14, 144–161. doi:10.1093/bib/bbs038

Lee, D.-I., and Roy, S. (2021). GRiNCH: Simultaneous Smoothing and Detection of Topological Units of Genome Organization from Sparse Chromatin Contact Count Matrices with Matrix Factorization. *Genome Biol.* 22, 164. doi:10.1186/s13059-021-02378-z

Lee, D., and Seung, H. S. (2001). "Algorithms for Non-negative Matrix Factorization," in *Advances in Neural Information Processing Systems* (MIT Press). Available at: https://papers.nips.cc/paper/2000/hash/

f9d1152547c0bde01830b7e8bd60024c-Abstract.html (Accessed September 7, 2021).

Leng, N., Dawson, J. A., Thomson, J. A., Ruotti, V., Rissman, A. I., Smits, B. M. G., et al. (2013). EBSeq: an Empirical Bayes Hierarchical Model for Inference in RNA-Seq Experiments. *Bioinformatics* 29, 1035–1043. doi:10.1093/bioinformatics/btt087

Li, B., and Dewey, C. N. (2011). RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome. *BMC Bioinformatics* 12, 323. doi:10.1186/1471-2105-12-323

Li, Q., Stram, A., Chen, C., Kar, S., Gayther, S., Pharoah, P., et al. (2014). Expression QTL-Based Analyses Reveal Candidate Causal Genes and Loci across Five Tumor Types. *Hum. Mol. Genet.* 23, 5294–5302. doi:10.1093/hmg/ddu228

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* 326, 289–293. doi:10.1126/science.1181369

Liu, J., Wang, C., Gao, J., and Han, J. (2013). "Multi-View Clustering via Joint Nonnegative Matrix Factorization," in Proceedings of the 2013 SIAM International Conference on Data Mining. Editors J. Ghosh, Z. Obradovic, J. Dy, Z.-H. Zhou, C. Kamath, and S. Parthasarathy (Philadelphia, PA: Society for Industrial and Applied Mathematics), 252–260. doi:10.1137/1.9781611972832.28

Liu, Z., Zhan, Y., Tu, Y., Chen, K., Liu, Z., and Wu, C. (2015). PDZ and LIM Domain Protein 1(PDLIM1)/CLP36 Promotes Breast Cancer Cell Migration, Invasion and Metastasis through Interaction with α-Actinin. *Oncogene* 34, 1300–1311. doi:10.1038/onc.2014.64

Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8

Marquis, C., Fonseca, C. L., Queen, K. A., Wood, L., Vandal, S. E., Malaby, H. L. H., et al. (2021). Chromosomally Unstable Tumor Cells Specifically Require KIF18A for Proliferation. *Nat. Commun.* 12, 1213. doi:10.1038/s41467-021-21447-2

Mathios, D., Hwang, T., Xia, Y., Phallen, J., Rui, Y., See, A. P., et al. (2019). Genome-wide Investigation of Intragenic DNA Methylation Identifies ZMIZ1 Gene as a Prognostic Marker in Glioblastoma and Multiple Cancer Types. *Int. J. Cancer* 145, 3425–3435. doi:10.1002/ijc.32587

Mohibi, S., Zhang, J., Chen, M., and Chen, X. (2021). Mice Deficient in the RNA-Binding Protein Zfp871 Are Prone to Early Death and Steatohepatitis in Part through the P53–Mdm2 Axis. *Mol. Cancer Res.* 10, 1751–1762. doi:10.1158/1541-7786.MCR-21-0239

Paulsen, J., Liyakat Ali, T. M., Nekrasov, M., Delbarre, E., Baudement, M.-O., Kurscheid, S., et al. (2019). Long-range Interactions between Topologically Associating Domains Shape the Four-Dimensional Genome during Differentiation. *Nat. Genet.* 51, 835–843. doi:10.1038/s41588-019-0392-0

Reedijk, M., Odorcic, S., Chang, L., Zhang, H., Miller, N., McCready, D. R., et al. (2005). High-level Coexpression of JAG1 and NOTCH1 Is Observed in Human Breast Cancer and Is Associated with Poor Overall Survival. *Cancer Res.* 65, 8530–8537. doi:10.1158/0008-5472.CAN-05-1069

Rey, C., Soubeyran, I., Mahouche, I., Pedeboscq, S., Bessede, A., Ichas, F., et al. (2013). HIPK1 Drives P53 Activation to Limit Colorectal Cancer Cell Growth. *Cel. Cycle* 12, 1879–1891. doi:10.4161/cc.24927

Rhie, S. K., Perez, A. A., Lay, F. D., Schreiner, S., Shi, J., Polin, J., et al. (2019). A High-Resolution 3D Epigenomic Map Reveals Insights into the Creation of the Prostate Cancer Transcriptome. *Nat. Commun.* 10, 4154. doi:10.1038/s41467-019-12079-8

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* 26, 139–140. doi:10.1093/bioinformatics/btp616

Roca, H., Hernandez, J., Weidner, S., McEachin, R. C., Fuller, D., Sud, S., et al. (2013). Transcription Factors OVOL1 and OVOL2 Induce the Mesenchymal to Epithelial Transition in Human Cancer. *PLOS ONE* 8, e76773. doi:10.1371/journal.pone.0076773

Rodriguez-Antona, C., and Ingelman-Sundberg, M. (2006). Cytochrome P450 Pharmacogenetics and Cancer. *Oncogene* 25, 1679–1691. doi:10.1038/sj.onc.1209377

Rowley, M. J., and Corces, V. G. (2018). Organizational Principles of 3D Genome Architecture. *Nat. Rev. Genet.* 19, 789–800. doi:10.1038/s41576-018-0060-8

Sabnis, R. W. (2020). Novel KIF18A Inhibitors for Treating Cancer. *ACS Med. Chem. Lett.* 11, 2079–2080. doi:10.1021/acsmedchemlett.0c00470

Sarni, D., Sasaki, T., Irony Tur-Sinai, M., Miron, K., Rivera-Mulia, J. C., Magnuson, B., et al. (2020). 3D Genome Organization Contributes to Genome Instability at Fragile Sites. *Nat. Commun.* 11, 3613. doi:10.1038/s41467-020-17448-2

Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C.-J., Vert, J.-P., et al. (2015). HiC-Pro: an Optimized and Flexible Pipeline for Hi-C Data Processing. *Genome Biol.* 16, 259. doi:10.1186/s13059-015-0831-x

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., et al. (2012). Three-dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell* 148, 458–472. doi:10.1016/j.cell.2012.01.010

Song, R., Chang, R., Dai, C., Wu, Y., Guo, J., Qi, M., et al. (2017). SORBS1 Suppresses Tumor Metastasis and Improves the Sensitivity of Cancer to Chemotherapy Drug. *Oncotarget* 8, 9108–9122. doi:10.18632/oncotarget.12851

Stalin, J., Imhof, B. A., Coquoz, O., Jeitziner, R., Hammel, P., McKee, T. A., et al. (2021). Targeting OLFML3 in Colorectal Cancer Suppresses Tumor Growth and Angiogenesis, and Increases the Efficacy of Anti-PD1 Based Immunotherapy. *Cancers* 13, 4625. doi:10.3390/cancers13184625

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., et al. (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr. Protoc. Bioinformatics* 54, 1.30.1–1.30.33. doi:10.1002/cpbi.510.1002/cpbi.5

Su, C., Johnson, M. E., Torres, A., Thomas, R. M., Manduchi, E., Sharma, P., et al. (2020). Mapping Effector Genes at Lupus GWAS Loci Using Promoter Capture-C in Follicular Helper T Cells. *Nat. Commun.* 11, 3294. doi:10.1038/s41467-020-17089-5

Taberlay, P. C., Achinger-Kawecka, J., Lun, A. T. L., Buske, F. A., Sabir, K., Gould, C. M., et al. (2016). Three-dimensional Disorganization of the Cancer Genome Occurs Coincident with Long-Range Genetic and Epigenetic Alterations. *Genome Res.* 26, 719–731. doi:10.1101/gr.201517.115

Tang, Z., Luo, O. J., Li, X., Zheng, M., Zhu, J. J., Szalaj, P., et al. (2015). CTCF-mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* 163, 1611–1627. doi:10.1016/j.cell.2015.11.024

Terry, M. B., Michels, K. B., Michels, K. B., Brody, J. G., Byrne, C., Chen, S., et al. (2019). Environmental Exposures during Windows of Susceptibility for Breast Cancer: a Framework for Prevention Research. *Breast Cancer Res.* 21, 96. doi:10.1186/s13058-019-1168-2

van Steensel, B., and Furlong, E. E. M. (2019). The Role of Transcription in Shaping the Spatial Organization of the Genome. *Nat. Rev. Mol. Cel. Biol.* 20, 327–337. doi:10.1038/s41580-019-0114-6

Wang, B., Pourshafeie, A., Zitnik, M., Zhu, J., Bustamante, C. D., Batzoglou, S., et al. (2018). Network Enhancement as a General Method to Denoise Weighted Biological Networks. *Nat. Commun.* 9, 3108. doi:10.1038/s41467-018-05469-x

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., et al. (2014). The NHGRI GWAS Catalog, a Curated Resource of SNP-Trait Associations. *Nucl. Acids Res.* 42, D1001–D1006. doi:10.1093/nar/gkt1229

Witwicki, R. M., Ekram, M. B., Qiu, X., Janiszewska, M., Shu, S., Kwon, M., et al. (2018). TRPS1 Is a Lineage-Specific Transcriptional Dependency in Breast Cancer. *Cel Rep.* 25, 1255–1267.e5. doi:10.1016/j.celrep.2018.10.023

Wyszynski, A., Hong, C.-C., Lam, K., Michailidou, K., Lytle, C., Yao, S., et al. (2016). An Intergenic Risk Locus Containing an Enhancer Deletion in 2q35 Modulates Breast Cancer Risk by Deregulating IGFBP5 Expression. *Hum. Mol. Genet.* 25, 3863–3876. doi:10.1093/hmg/ddw223

Yang, Y., Wang, C., Zhang, P., Gao, K., Wang, D., Yu, H., et al. (2013). Polycomb Group Protein PHF1 Regulates P53-dependent Cell Growth Arrest and Apoptosis. *J. Biol. Chem.* 288, 529–539. doi:10.1074/jbc.M111.338996

Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., et al. (2020). Ensembl 2020. *Nucleic Acids Res.* 48, D682–D688. doi:10.1093/nar/gkz966

Yu, X., and Li, Z. (2015). TOX Gene: a Novel Target for Human Cancer Gene Therapy. *Am. J. Cancer Res.* 5, 3516–3524.

Zeng, C., Huang, W., Li, Y., and Weng, H. (2020). Roles of METTL3 in Cancer: Mechanisms and Therapeutic Targeting. *J. Hematol. Oncol.* 13, 117. doi:10.1186/s13045-020-00951-w

Zhang, Y., and Yang, Q. (2021). "A Survey on Multi-Task Learning," in IEEE Transactions on Knowledge and Data Engineering, 1. doi:10.1109/TKDE.2021.3070203

Zhang, Y., Manjunath, M., Yan, J., Baur, B. A., Zhang, S., Roy, S., et al. (2019). The Cancer-Associated Genetic Variant Rs3903072 Modulates Immune Cells in the

Tumor Microenvironment. *Front. Genet.* 10, 754. doi:10.3389/fgene.2019.00754

Zhang, H., Ahearn, T. U., Lecarpentier, J., Barnes, D., Beesley, J., Qi, G., et al. (2020). Genome-wide Association Study Identifies 32 Novel Breast Cancer Susceptibility Loci from Overall and Subtype-specific Analyses. *Nat. Genet.* 52, 572–581. doi:10.1038/s41588-020-0609-2

Zheng, H., and Xie, W. (2019). The Role of 3D Genome Organization in Development and Cell Differentiation. *Nat. Rev. Mol. Cel. Biol.* 20, 535–550. doi:10.1038/s41580-019-0132-4

Comparison of Capture Hi-C Analytical Pipelines

Dina Aljogol[1], I. Richard Thompson[2], Cameron S. Osborne[3] and Borbala Mifsud[1,4]*

[1]College of Health and Life Sciences, Hamad Bin Khalifa University, Doha, Qatar, [2]Qatar Biomedical Research Institute, Hamad Bin Khalifa University, Doha, Qatar, [3]Department of Medical and Molecular Genetics, King's College London, London, United Kingdom, [4]William Harvey Research Institute, Queen Mary University of London, London, United Kingdom

It is now evident that DNA forms an organized nuclear architecture, which is essential to maintain the structural and functional integrity of the genome. Chromatin organization can be systematically studied due to the recent boom in chromosome conformation capture technologies (e.g., 3C and its successors 4C, 5C and Hi-C), which is accompanied by the development of computational pipelines to identify biologically meaningful chromatin contacts in such data. However, not all tools are applicable to all experimental designs and all structural features. Capture Hi-C (CHi-C) is a method that uses an intermediate hybridization step to target and select predefined regions of interest in a Hi-C library, thereby increasing effective sequencing depth for those regions. It allows researchers to investigate fine chromatin structures at high resolution, for instance promoter-enhancer loops, but it introduces additional biases with the capture step, and therefore requires specialized pipelines. Here, we compare multiple analytical pipelines for CHi-C data analysis. We consider the effect of retaining multi-mapping reads and compare the efficiency of different statistical approaches in both identifying reproducible interactions and determining biologically significant interactions. At restriction fragment level resolution, the number of multi-mapping reads that could be rescued was negligible. The number of identified interactions varied widely, depending on the analytical method, indicating large differences in type I and type II error rates. The optimal pipeline depends on the project-specific tolerance level of false positive and false negative chromatin contacts.

Keywords: epigenetics, gene regulation, computational pipeline, capture Hi-C, chromatin organization

1 INTRODUCTION

The DNA fiber within the nucleus is assembled into an organized, multi-level architecture. During interphase, chromosomes occupy distinct territories that rarely interact (Cremer and Cremer 2010). Chromatin is further partitioned into hubs of active and inactive compartments, determined by their chromatin accessibility status, gene density and bound proteins (Rao et al., 2014). These compartments are built from smaller topologically associated domains (TADs), which serve as regulatory units, enclosing most chromatin loops within their boundaries (Dixon et al., 2012; Nora et al., 2012). Chromatin loops facilitate the communication of distant genomic regions by bringing them into physical proximity, including enhancers and their target promoters. Substantial evidence supports the importance of this organization in maintaining genome integrity and driving key biological processes, such as transcription (Osborne et al., 2004; Rao et al., 2014; Rhie et al., 2019; Akdemir et al., 2020; Cai et al., 2021). For instance, 3D genomic rearrangements allow genes to alternate between areas of active and repressed chromatin environments to regulate the circadian rhythm (Furlan-Magaril et al., 2021).

The 3D genome architecture can be investigated using either imaging or chromosome conformation capture (3C)-based methods. Imaging techniques are traditionally limited to studying a handful of loci at a time, even though recent developments in the field allow genome-scale studies (Su et al., 2020). 3C-based methods, on the other hand, have been used to study interactions genome-wide for more than a decade. 3C is a proximity ligation-based method, which was developed by Dekker et al. to study 'one to one' contacts using PCR amplification for detection (Dekker et al., 2002). Subsequently, several large-scale methods emerged, including the unbiased, genome-wide method, Hi-C, which leverages high-throughput sequencing to quantify all interactions simultaneously (Lieberman-Aiden et al., 2009). While Hi-C can provide information for all contacts, it requires deep sequencing to confidently identify true genomic interactions at higher resolution (Rao et al., 2014). To overcome this limitation and to focus on regulatory loops, library enrichment strategies, such as Capture Hi-C (CHi-C) (Mifsud et al., 2015) and Capture-C (Davies et al., 2016), have been applied. CHi-C uses sequence-specific RNA baits to further select regions of interest from a pool of ligated Hi-C contacts prior to sequencing. It has been widely used to capture promoter interactions with regulatory elements (Furlan-Magaril et al., 2021; Jung et al., 2019) and it has also been employed to assess disrupted genomic interactions of disease risk loci (Baxter et al., 2018; Song et al., 2019).

A typical Hi-C data analysis workflow includes the following steps: quality control and alignment of sequenced reads (Servant et al., 2015; Wingett et al., 2015; Zheng et al., 2019), optional binning of interactions, bias-correction (Imakaev et al., 2012) and performing a statistical test to identify valid (Mifsud et al., 2017) or functional interactions (Heinz et al., 2010; Hwang et al., 2015; Durand et al., 2016; Ron et al., 2017; Wolff et al., 2018; Kaul et al., 2020), which can be interrogated in downstream analyses (Lajoie et al., 2015). For each step, there is a growing selection of tools. While systematic comparisons of Hi-C analytical pipelines exist (Forcato et al., 2017; Pal et al., 2019), there is a lack of similar comparisons for CHi-C data.

Data from CHi-C experiments requires specialized software because CHi-C-specific biases, such as variable capture efficiency, are not accounted for by most Hi-C analysis tools. Furthermore, bait-bait interactions need to be treated separately from bait-other interactions. Ligation fragments that are targeted by baits on both ends have different capture probabilities compared to those targeted only on a single end.

The main decision points for CHi-C data analysis are choosing the method for alignment and filtering of the sequenced read-pairs and choosing the method for identifying interactions of interest. For alignment, most methods will utilize read pairs, where both ends are aligned uniquely to the genome, e.g., HiCUP and HiC-Pro (Servant et al., 2015; Wingett et al., 2015). Zheng et al. proposed an alternative method that rescues those multi-mapping read pairs that can be unambiguously assigned to an interaction, however, the benefit of this method for CHi-C has not been assessed (Zheng et al., 2019). For identification of interactions

of interest, there are a number of distinct strategies. GOTHiC aims to identify those interactions that are not experimental artefacts, but represent real contacts in the nucleus. It does not take genomic distance between the interacting fragments into account and it does not infer biologically relevant interactions (Mifsud et al., 2017). Although it was originally developed for Hi-C data, its visibility correction method, which uses all reads mapping to a fragment as the basis of correction, is applicable to bait-other interactions of CHi-C data as well. In combination with a random ligation sample, a modified version of the algorithm can be applied to bait-bait interactions, which uses a mixed additive/multiplicative model for visibility correction (Mifsud et al., 2015). Other methods aim to find functional interactions by assuming that contacts, which occur more often in the nucleus than other contacts spanning similar genomic distances, are biologically relevant. CHiCAGO's goal is to identify functional interactions by pinpointing those that show higher contact frequencies than would be expected by Brownian motion of the chromatin. It also corrects for visibility of a fragment by separating baits and other ends into groups of fragments with similar coverage (Cairns et al., 2016). CHiCANE calculates the significance of an interaction taking into account both the genomic distance between two bins and the "interactibility" of bait fragments. "Interactibility" is defined as the number of trans reads a bait fragment has (Holgersen et al., 2021). The above mentioned methods use global background measures to identify real or functional interactions, which do not take into account the local chromatin environment of a given bait fragment. In contrast, CHiCMaxima does not take into account the global properties of the CHi-C data set, but treats the contacts of each bait as a virtual 4C instead. It smoothes the read count profile of the bait and uses local maxima to find those fragments that form functional chromatin loops (Zouari et al., 2019). Here, we compare the performance of these various CHi-C data analysis pipelines in detecting reproducible interactions that are of potential biological relevance.

2 MATERIALS AND METHODS

2.1 CD34$^+$ CHi-C

2.1.1. CD34$^+$ Cell Collection, Purification and Fixation

CD34$^+$ cells were collected from the femoral heads of healthy donors who underwent total hip replacement surgery (in a consented study approved by the London - Westminster Research Ethics Committee - IRAS#220344). Bone marrow was extracted and irrigated in Iscove Modified Dulbecco Medium/10% Fetal calf serum. CD34$^+$ cells were isolated from the cell suspension using a Dynabeads CD34 Positive Isolation Kit (Invitrogen cat# 11301D). PBS-EDTA washed cells were fixed with 2% final concentration of formaldehyde for 10 min at room temperature. After quenching the fixation with 0.125M final concentration of glycine, CD34$^+$ cells were purified using CD34$^+$ MicroBeads (Miltenyi) according to manufacturer's

instructions. A 1 ml aliquot was used to assess the CD34[+] purity by FACS and the purity was determined to be above 90%.

2.1.2. Promoter Capture Hi-C
Hi-C library generation was carried out as described previously (Mifsud et al., 2015), with minor modifications. Briefly, after overnight digestion with HindIII at 37°C, DNA ends were labelled with biotin-14–dATP (Life Technologies) using a Klenow end-filling reaction. In nucleus ligation was performed by ligating together biotinylated DNA ends overnight using T4 DNA ligase (Invitrogen). After phenol: chloroform/ethanol purification DNA was quantified using Qubit, with a maximum of 40 μg taken forward. DNA was sheared to a peak concentration of ~ 400 bp, using the manufacturer's instructions (Covaris). Sheared DNA was then end-repaired, polyadenylated, and double size selected using AMPure XP beads to isolate DNA ranging from 250 to 550 bp in size. Ligation fragments marked by biotin were immobilized using MyOne Streptavidin C1 DynaBeads (Invitrogen) and ligated to paired-end adaptors (Illumina). Hi-C libraries were then amplified using PE PCR 1.0 and PE PCR 2.0 primers (Illumina) with 6 PCR amplification cycles.

Promoter capture was carried out with SureSelect target enrichment, using a custom-designed biotinylated RNA bait library and custom paired-end blockers according to the manufacturer's instructions (Agilent Technologies). The 120-mer baits were targeting both ends of HindIII restriction fragments that overlap with Ensembl promoters of protein-coding, noncoding, antisense, snRNA, miRNA and snoRNA transcripts, had a 25–65% GC content, their sequence contained no more than two consecutive Ns and were within 330 bp of the HindIII restriction fragment terminus. After library enrichment, a post-capture PCR amplification step was carried out using PE PCR 1.0 and PE PCR 2.0 primers with 4 PCR amplification cycles. CHi-C libraries were sequenced on the Illumina HiSeq 2000 platform for paired-end sequencing.

2.2 Tools and Datasets
Three replicates of GM12878 in solution ligation promoter capture Hi-C and three replicates each of iPSC and iPSC-derived cardiomyocyte in nucleus promoter capture Hi-C data were downloaded from ArrayExpress (E-MTAB-2323 and E-MTAB-6014, respectively) using fasterq-dump v2.9.6. GRCh37 reference genome and chromosome sizes were obtained from the UCSC genome browser.

H3K27ac, H3K4me1 and H3K4me3 peaks and DNase I hypersensitivity sites (DHS; GM12878, H1) from the Roadmap Epigenomics Consortium (2015), heart DHS from the ENCODE project (Consortium 2012) and Nuclease accessible sites (NAS; CD34[+]) from Gargiulo et al. (2009) were downloaded using the AnnotationHub v2.22.1 R BioConductor package for GM12878 (record numbers "AH29709", "AH29060", "AH29061", and "AH30743", respectively) and CD34[+] cells (record numbers "AH42424", "AH42192", "AH42194", and "AH5085", respectively), H1 cells (record numbers "AH29891", "AH28878", "AH28880", and "AH29873", respectively) and left ventricle/heart (record numbers "AH30592", "AH29554", "AH29555", and "AH25530", respectively). Significant

H3K27ac, H3K4me1, H3K4me3 and DHS peaks were defined as q-value< 0.05.

HiCUP v0.7.2 (Wingett et al., 2015), mHiC (Zheng et al., 2019), GOTHiC++ (based on (Mifsud et al., 2017), CHiCAGO (Cairns et al., 2016), CHiCANE ((Holgersen et al., 2021) and CHiCMaxima (Zouari et al., 2019) were downloaded from links summarized in **Figure 1A**.

2.3 Read Alignment and Filtering
2.3.1 HiCUP
An *in silico* 1-based HindIII digest profile of the hg19 reference genome was created using hicup_digester. This file represents all possible HindIII fragments in the genome and was used to identify CHi-C artifacts. HiCUP v0.7.2 was used with bowtie2 v2.4.2. (hg19) for the alignment step, and minimum and maximum di-tag ranges were set to 150 and 800 for the filtering step. All other parameters were kept as default. The final BAM output was filtered to include only read pairs where both ends have a mapping quality ≥10.

2.3.2 mHiC
mHiC was applied at four different resolutions: Restriction fragment level (RF), 10 kb, 100 kb and 1 MB with the default BWA aligner (v0.7.17-r1188). The 0-based HindIII digest profile supplied by mHiC was used for mapping the reads to restriction fragments. Parameters were adjusted to be consistent with the parameters used for HiCUP. We used 150 for the minimum and 800 for the maximum di-tag length. The chimeric read length threshold was adjusted to 20. The mapping quality threshold was reduced to 10. The unique and multi-read valid pairs (those that map to unique bins) were concatenated for further processing. The final normalization steps of mHiC were omitted. Valid read pairs were kept from the SAM output of step 2 and the SAM file was converted to BAM using samtools v1.9. for GOTHiC, CHiCAGO and CHiCANE input.

2.4 Identifying Significant Interactions
Significant chromatin contacts were identified at fragment resolution using four different software. Three compare the observed read counts for each interaction to a global background and one identifies significant contacts based on the local interaction profile (**Figure 1B**). Additionally, since the GOTHiC algorithm does not aim to identify functional interactions among those present in the nucleus, we defined bait-specific q-value thresholds for the GOTHiC results. Bait-specific q-value thresholds filter for interactions that are more significant than the majority of contacts a given bait makes. These are likely to represent functional loops.

2.4.1 CHiCAGO
CHiCAGO requires five input files: Rmaps represent all the possible fragments in the genome. Baitmaps represent intervals of fragments that were baited, as well as their bin ID relative to the rmap, and gene names within each captured fragment. The remaining three input files were created using chicagoTools makeDesignFiles.py script with its default settings for HindIII. For MboI-digested fragment we used binsize 1,500, minFragLen
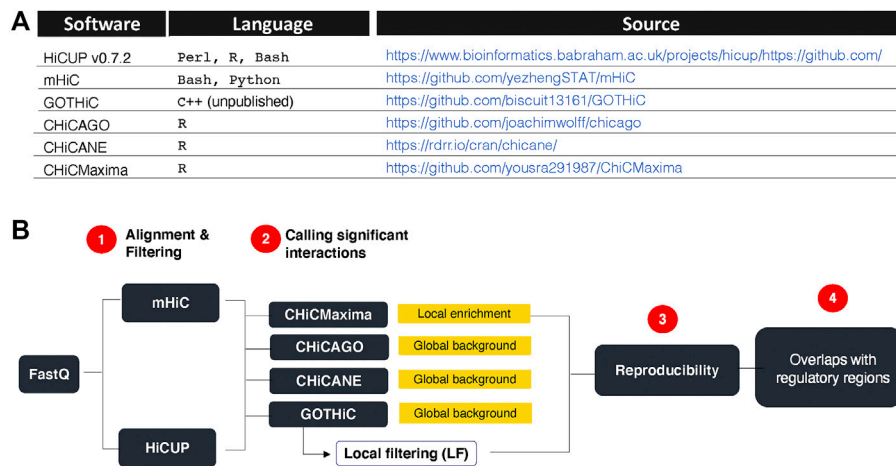
FIGURE 1 | Research summary. **(A)**. CHi-C analytical tools used and their sources. **(B)**. Strategy overview. HiCUP and mHiC were compared for their performance in mapping read pairs and filtering experimental artefacts. GOTHiC, CHiCMaxima, CHiCAGO and CHiCANE were compared for their ability to identify reproducible, biologically relevant interactions. Yellow boxes indicate the type of background model used by each tool. GOTHiC local filtering (LF) is an optional downstream filtering of GOTHiC globally significant interactions based on the local interaction profile of each bait.

75, maxFragLen 12,000 and maxLBrownEst 97,500. We converted BAM files to chinput format using chicagoTools bam2chicago.sh. Lastly, runChicago.R was executed with the same settings mentioned above. The significance threshold was set to a score ≥5. We also tested ≥10 and ≥15.

2.4.2 CHiCANE
Interactions files were created using prepare. data () with the default parameters and three input files: HiCUP/mHiC BAM files, and the baitmaps and rmaps created previously. We then executed chicane() using the interactions file as input. Significance threshold was set to q-value < 0.05. We also tested <0.01 and <0.001.

2.4.3 GOTHiC++
We executed gothic using the BAM files with default settings. Significance threshold was set to q-value < 0.05. We also tested <0.01 and <0.001. GOTHiC identifies interactions that are not due to random ligation events. In order to identify which one of the non-random interactions might be biologically relevant, we defined a per bait q-value threshold based on the slope of the cumulative significance [-log10 (q-value)] curve of the interactions each bait made. Briefly, significance values of all significant interactions of the bait were rounded and for each value we calculated the number of interactions with equal or higher significance. We took the derivative of this cumulative curve to set the threshold for the bait to the significance level, where the absolute slope is above 1.

2.4.4 CHiCMaxima
Interactions input files were created in the format specified in CHiCMaxima. IDs were defined as their bin ID relative to the rmap file. CHiCMaxima was used with default settings with a window size of 20 and 100 for HindIII- and MboI-digested samples, respectively. CHiCMaxima excludes genes with

insufficient coverage. Therefore, the output includes only valid interactions.

2.5 Downstream Analyses
We assessed the reproducibility of significant interactions two-fold. First, we overlapped the non-bait captured fragments for each bait across all replicates, then we investigated those that overlap with active chromatin. Interactions that were present in at least two replicates were considered reproducible. We also calculated the number of interactions, which pass a joint mean q-value or score threshold in at least two replicates for each tool.

To assess whether significant interactions are of biological relevance we calculated the proportion of identified promoter interacting fragments that harbour active chromatin regions. We overlapped the fragments, or the fragments extended on both sides with either 2.5 kb or 20 kb, with H3K27ac, H3K4me1, H3K4me3 and DNase I hypersensitivity sites using the GenomicRanges v1.42.0 R package.

3 RESULTS

3.1 The Effect of Multi-Mapping Reads
We analysed eleven promoter capture Hi-C (PCHi-C) data sets. Three replicates of the GM12878 cell line were in solution-ligation PCHi-C data sets with 93 million to 188 million sequenced read pairs. Two replicates prepared by in nucleus ligation from CD34[+] hematopoietic stem cells were sequenced more deeply and had 359 million and 579 million read pairs. Three replicates each of iPSC and iPSC-generated cardiomyocyte in nucleus ligation PCHi-C libraries were sequenced at similar depth with 368 million to 475 million reads (**Figure 2A**, **Supplementary Table S1**). Raw sequencing reads were mapped and filtered either by HiCUP, which only considers uniquely mapping reads, or by mHiC, which rescues those
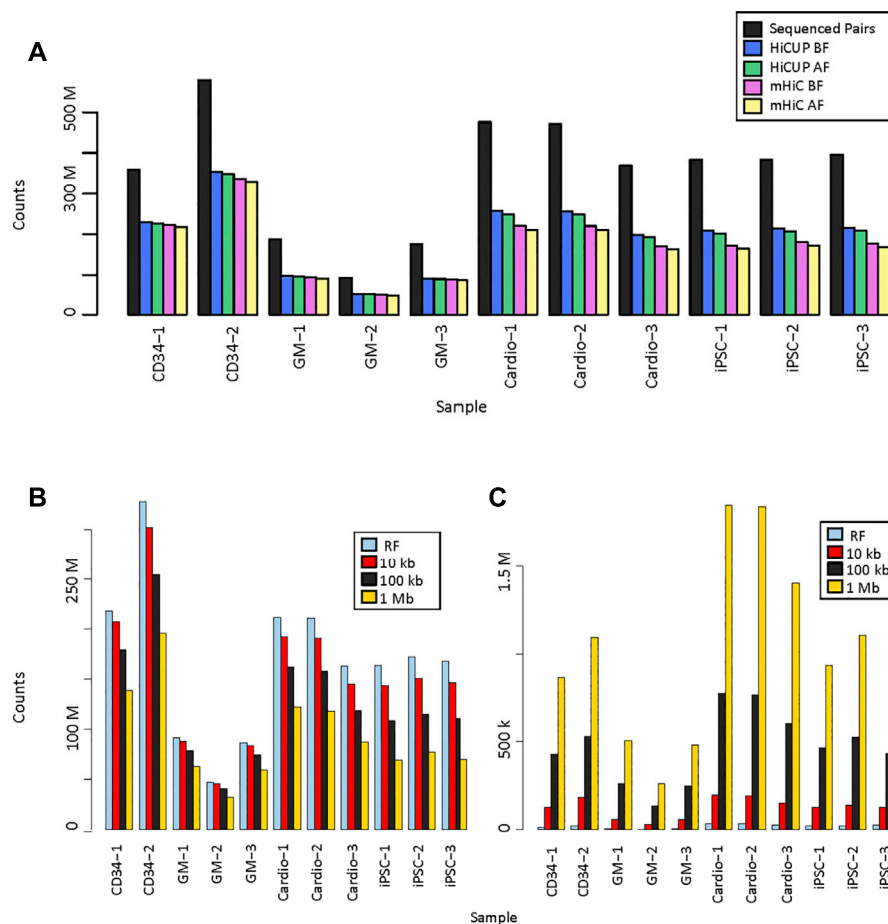
FIGURE 2 | Valid read pairs. **(A)**. Number of identified valid pairs using HiCUP and mHiC. Black bars indicate the total number of raw read pairs. Read pairs were filtered to keep only those with a mapping quality (MAPQ) ≥10. BF: number of mapped read pairs before MAPQ filtering. AF: number of mapped reads after MAPQ filtering. **(B)**. Number of uniquely mapping read pairs using mHiC at different resolutions. **(C)**. Number of rescued multi-mapping read pairs using mHiC at different resolutions. RF: restriction fragment.

multi-mapping reads that map to unique restriction fragments or interaction bins, depending on the resolution. HiCUP returned a slightly higher number of valid read-pairs than mHiC, which difference became more prominent in the MboI-digested samples. Each sample had 52–64% valid mapped read-pairs using HiCUP and 45–62% using mHiC. The difference remained the same when only those read-pairs were kept, where both ends had a good mapping quality (MAPQ ≥10) (**Figure 2A**, **Supplementary Table S1**.).

Zheng et al. showed that mHiC can rescue up to 20% of reads in Hi-C samples (Zheng et al., 2019), but we did not observe higher valid read counts when mapping these PCHi-C samples. In order to explore whether the lack of improved valid read proportion was due to the high, fragment-level resolution of PCHi-C, we calculated the number of valid unique and rescued multi-mapping reads at fragment level, 10 kb, 100 kb and 1 Mb resolutions (**Figures 2B,C**, **Supplementary Table S2**). The number of uniquely mapping reads decreased as the resolution decreased, because a larger proportion of the read pairs fell on the diagonal of the contact matrix, into a single bin, and those read

pairs were filtered out. The decrease was more pronounced for deeper sequenced samples and for shorter fragments (**Figure 2B**). The number of rescued multi-mapping read pairs was negligible at restriction fragment level resolution; at most 32,239 read pairs were rescued in the largest MboI-digested sample. This number did increase with the use of larger bins, however, it did not exceed 1.9M reads at 1 Mb resolution, which was only 0.6–1.6% of the uniquely mapping read pairs in the same samples (**Figure 2C**).

3.2 Reproducibility of Interactions

The numbers of identified significant interactions using HiCUP- or mHiC-aligned and filtered reads were similar. In general, there were up to 10% fewer interactions using mHiC-aligned reads (**Figure 3A**, **Supplementary Figure S1**). There was a 2–500-fold difference in the number of interactions identified by GOTHiC and CHiCANE. CHiCMaxima identified slightly more interactions than CHiCAGO and they both returned ~ 5–20 times as many interactions as CHiCANE (**Figure 3A**, **Supplementary Table S3A**). These differences were also apparent in the proportion of baited fragments with at least
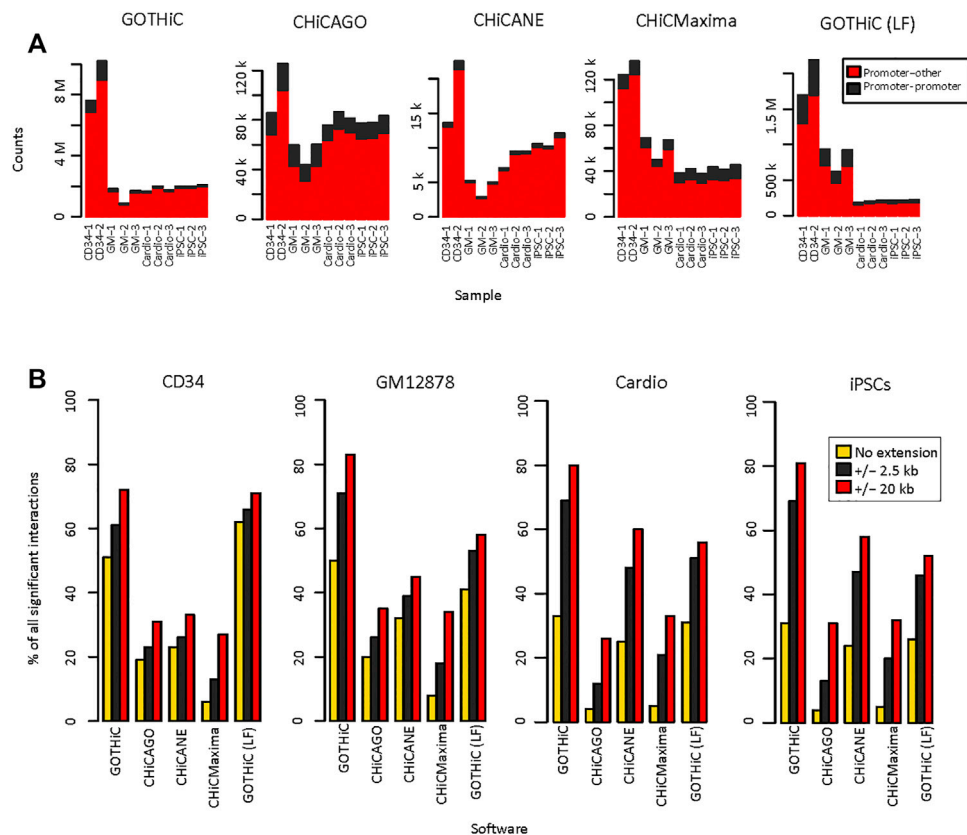
FIGURE 3 | Reproducibility of HiCUP-preprocessed significant interactions. **(A)**. Number of HiCUP-preprocessed significant interactions using GOTHiC (with and without LF), CHiCAGO, CHiCANE and CHiCMaxima. **(B)**. Bar plots represent the percentage of non-baited fragments that overlap in at least two replicates for each bait. Overlaps were studied for exact fragment-level interactions and interactions where non-baited fragment-ends were extended by 2.5 kb or 20 kb.

one identified interaction, which was 99–99.9% and 79–82% with GOTHiC and only 13–89% and 7–13% with CHiCANE for HindIII- and MboI-digested samples, respectively (CHiCAGO: 71–78.9% and 45–52%, CHiCMaxima: 86–99.6% and 32–36%) (**Supplementary Table S4**). The proportion of bait-bait interactions was highest in CHiCAGO, in the local-filtered GOTHiC and in CHiCMaxima for MboI-digested samples (**Figure 3A**, **Supplementary Figure S1**). The number of significant interactions in the 4-cutter-digested samples was equivalent to or lower than the number in the GM12878 datasets by GOTHiC, CHiCMaxima and GOTHiC (LF), despite the deeper sequencing of the iPSC and cardiomyocyte samples. CHiCAGO identified a similar number of significant interactions to those in the larger HindIII samples (**Figure 3A**, **Supplementary Figure S1**).

Reproducibility of exact interactions across replicates ranged from 4 to 8% using CHiCMaxima. It was 44–50% for GOTHiC (41–57% after local filtering). CHiCAGO showed 16–20% and CHiCANE interactions showed 22–32% reproducibility (**Figure 3B**, **Supplementary Table S5A**). However, when interactions between the exact fragments are not observed, it has been noted that interactions with neighbouring fragments are present in the replicates, therefore we calculated the reproducibility of interactions by extending the non-baited

fragments with 2.5 kb or 20 kb on each side. This resulted in a higher proportion of overlapping interactions in all tools, especially for 4-cutter digested samples. The most prominent increase was observed for GOTHiC, the proportion of reproducible interactions increased to 61–71% with the 2.5 kb extension (**Supplementary Table S6A**) and 70–83%% with the 20 kb extension (**Supplementary Table S7A**). CHiCMaxima showed the lowest reproducibility after extension as well (**Figure 3B**). In order to test whether the choice of threshold affected our results, we also filtered at q-values < 0.01 and 0.001 for CHiCANE, GOTHiC and GOTHiC (LF), and at scores ≥ 10 and 15 in CHiCAGO. CHiCMaxima does not have a scoring system and returns only local peaks. Using the second threshold, the number of significant interactions decreased by 11–22% in GOTHiC, 85–96% in CHiCAGO, 41–50% in CHiCANE and 0.3–7% in GOTHiC (LF). Using the third threshold, the number of identified interactions decreased by 28–34% in GOTHiC, 95–99.5% in CHiCAGO, 67–80% in CHiCANE and 0.5–13% in GOTHiC (LF) (**Supplementary Tables S3B,C**). These had a negligible effect on the reproducibility as the maximum increase was 0–5% in GOTHiC, 1–13% in CHiCAGO, 0–5% in CHiCANE and 0–4% in GOTHiC (LF) at the restriction fragment level (**Supplementary Tables S5B,C**). The increase was lower when extending for 2.5 kb (**Supplementary Tables S6B,C**) or 20 kb (**Supplementary Tables S7B,C**).

FIGURE 4 | Biologically relevant interactions. Bubble heat maps represent the proportion of interactions that overlap **(A)**. DHS or **(B)**. H3K4me1 peaks. The size of the circle represents the log10 number of promoter-other interactions identified by the different methods. **(C)**. A comparison between *BCL2* promoter interactions in GM12878 cells within ± 1 Mb as reported by each method. Dashed lines represent examples of interactions that are identified in all tools (magenta), all tools except CHiCANE (blue), only in GOTHiC (yellow). ENCODE ChIP-seq profiles for DHS, H3K4me1 and H3K27ac in GM12878 cell lines are shown in the bottom tracks. DHS: DNase I Hypersensitivity sites.

The reproducibility of potentially functional interactions, where the non-baited fragments overlapped with DNaseI hypersensitivity sites (DHS), H3K4me1, H3K4me3 or H3K27ac peaks, was higher than it was for all identified interactions using GOTHiC and its local-filtered interaction list but was equal or lower using the other methods (**Supplementary Figure S2**, **Supplementary Tables S8–S10**).

All Hi-C-type data, including capture Hi-C, are known to be prone to undersampling, therefore we tested the utility of using a joint mean q-value (GOTHiC, CHiCAGO and CHiCANE) and score (CHiCAGO) threshold for replicates. This resulted in a 0–17% increase in the total number of unique interactions identified across single replicates, the highest being in CHiCAGO. It indicates that, especially in CHiCAGO, many interactions that are significant in one replicate only, are near the threshold in another replicate (**Supplementary Table S5D**).

3.3 Interactions With Potential Biological Function

Finally, we assessed whether the identified interactions are of potential biological function, by overlapping the non-baited fragments with DHS (open chromatin), H3K4me1 and H3K27ac peaks (enhancer) and H3K4me3 (active promoter) from the respective cell types. A larger proportion of interactions overlapped with DHS peaks (13–64%) compared to H3K27ac (5–39%), H3K4me1 (5–56%) and H3K4me3 (5–30%) peaks. GOTHiC interactions had the lowest proportion of interactions overlapping with these features. CHiCAGO- and CHiCMaxima-identified interactions had on average a 1.6-fold higher proportion of functional interactions than GOTHiC-identified interactions and 1.2-fold higher than local filtered GOTHiC interactions. In general, CHiCANE showed the highest percentage of interactions overlapping active chromatin (**Figures 4A,B**). These differences are diminished when the non-baited fragments are extended. 2.5 kb-extended CHiCMaxima interactions have a higher proportion of functional interactions in 4-cutter digested samples than CHiCANE, but lower in 6-cutter digested ones (**Supplementary Tables S11–S13**).

Interactions made by the baited *BCL2* promoter demonstrate the above observations. Most methods identified a low number of interactions for this promoter, while GOTHiC, even after local filtering, found several interacting fragments. GOTHiC-identified interactions also spanned further than those pinpointed by other methods. CHiCANE interactions were the fewest and shortest. The bottom tracks show DHS, H3K4me1 and H3K27ac profiles in this region. Magenta highlights an interaction that was identified in all methods, blue highlights peaks overlapped with an interaction that was identified by all methods except CHiCANE, while yellow highlights those in GOTHiC-only interacting fragments (**Figure 4C**).

4 DISCUSSION

Recent advances in chromosome conformation capture technologies have enabled us to systematically investigate the spatial arrangement of chromatin within the nucleus. The increasing number of experimental approaches were accompanied by the development of computational pipelines to analyze resulting data and ensure reproducibility of research, but there is no standard method for the analysis of CHi-C data.

Here, we compared two software for the alignment and filtering of reads, HiCUP and mHiC. The former uses uniquely mapping reads only, while the latter keeps those multi-mapping reads that come from a single restriction fragment or genomic bin. At restriction fragment resolution the use of mHiC was not advantageous, in fact HiCUP returned more valid read pairs. This difference might have come from the different aligners used by the two methods, as HiCUP uses Bowtie2, while mHiC uses BWA. For this set of samples, we did not observe substantial benefit

from rescuing multi-mapping reads by mHiC at lower resolutions either. However, it might be useful with samples of lower sequencing quality as reads with mismatches are more prone to be misaligned.

Next, we compared methods used for identification of real or functional interactions. In addition to identifying interactions, CHiCAGO and CHiCANE provide R code for functional enrichment and visualization of the data and CHiCMaxima has a graphical interface which facilitates its use for less experienced users. These additional features might also influence the choice of tool. Here we focused on the reproducibility and specificity of regulatory interactions identified.

The most striking difference between these methods was the number of interactions identified. GOTHiC, which identifies those interactions that are not due to spurious ligation of DNA ends, unsurprisingly returns a magnitude higher number of interactions than the other methods, which define interactions as those that are more enriched than their local environment or than other interactions with similar genomic distance. When we aim to find functional interactions and filter GOTHiC results based on the local interaction profiles, there is about 0.1–0.69 of those interactions left, which is still much more than what we found by any other method. It is likely that this method has the highest false positive rate for functional interactions, but examples showed that many regulatory chromatin features are linked to promoters solely by GOTHiC. CHiCANE is the strictest of all four methods tested and a very high proportion of the interacting fragments is overlapping active chromatin, but it is likely to have a very high false negative rate, also indicated by the low proportion of baits with at least a single interaction. CHiCAGO and CHiCMaxima can be a good compromise between false positive and false negative rates, as these identify 4–18 times as many interactions as CHiCANE, and the proportion of those significant interactions that overlap with regulatory features is not much below CHiCANE's. Extending the interacting fragments with 2.5 kb or 20 kb on each side, increased both reproducibility and the proportion of regulatory interactions. However, the 20 kb-extension reduces the resolution of CHi-C beyond the size of an average regulatory element, which is not recommended for studying promoter-enhancer interactions.

In summary, the choice of method should depend on the tolerable level of false positive and false negative interactions, and this systematic comparison will help researchers identify the method best applicable to their projects.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ebi.ac.uk/arrayexpress/, E-MTAB-10701 https://www.ebi.ac.uk/arrayexpress/, E-MTAB-2323 and https://www.ebi.ac.uk/arrayexpress/ E-MTAB-6014.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by London - Westminster Research Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DA and BM. conceptualized the project. IT. developed C++ version of GOTHiC. CO. performed promoter capture Hi-C on CD34[+] cells. DA. performed the analyses under BM's supervision. DA and BM. wrote initial draft of the manuscript. DA, CO, and BM. revised the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.786501/full#supplementary-material

Supplementary Figure S1 | Number of mHiC-preprocessed significant interactions using GOTHiC (with and without LF), CHiCAGO, CHiCANE and CHiCMaxima.

Supplementary Figure S2 | Reproducibility of HiCUP-preprocessed regulatory interactions. Bar plots represent the percentage of **(A)**. DHS, **(B)**. H3K27ac, **(C)**. H3K4me1 or **(D)**. H3K4me3 overlapping non-baited fragments that overlap in at least two replicates for each bait. Overlaps were studied for exact fragment-level interactions and interactions where non-baited fragment-ends were extended by 2.5 kb or 20 kb. DHS: DNase I Hypersensitivity sites.

Supplementary Figure S3 | Biologically relevant interactions. Bubble heat maps represent the proportion of interactions that overlap **(A)**. H3K27ac or **(B)**. H3K4me3 peaks. The size of the circle represents the log10 number of promoter-other interactions identified by the different methods.

Supplementary Table S1 | Number of mapped valid read pairs using HiCUP and mHiC. Read pairs were filtered to keep only those with a mapping quality (MAPQ) >=10.

Supplementary Table S2 | Number of unique and multi-mapping read pairs using mHiC at different resolutions. RF: Restriction fragment.

Supplementary Table S3 | Number of significant interactions identified **(A)** at the default thresholds (q<0.05, CHiCAGO score>=5) and at stricter **(B)** q<0.01, score >=10 or **(C)** q<0.001, score >=15 thresholds.

Supplementary Table S4 | Number of baits with at least one significant interaction identified. Total number of baits used was 19,022 and 70,545 in HindIII- and MboI-digested samples, respectively.

Supplementary Table S5 | (A-C). Percentage of exact non-baited fragments detected at different significance thresholds that overlap in at least two replicates for each bait. **(D)**. Percentage of reproducible interactions, where the average q-value or CHiCAGO score of at least two replicates is below the default threshold.

Supplementary Table S6 | Percentage of non-baited fragments detected at different significance thresholds, where both ends have been extended by 2.5 kb, that overlap in at least two replicates for each bait.

Supplementary Table S7 | Percentage of non-baited fragments detected at different significance thresholds, where both ends have been extended by 20 kb, that overlap in at least two replicates for each bait.

Supplementary Table S8 | Percentage of regulatory interactions that overlap in at least two replicates for each bait. DHS: DNase I Hypersensitivity sites.

Supplementary Table S9 | Percentage of regulatory interactions, where both ends have been extended by 2.5 kb that overlap in at least two replicates for each bait. DHS: DNase I Hypersensitivity sites.

Supplementary Table S10 | Percentage of regulatory interactions, where both ends have been extended by 20 kb that overlap in at least two replicates for each bait. DHS: DNase I Hypersensitivity sites.

Supplementary Table S11 | Proportion of interactions which overlap **(A)**. DHS, **(B)**. H3K27ac, **(C)**. H3K4me1 or **(D)**. H3K4me3 peaks. DHS: DNase I Hypersensitivity sites.

Supplementary Table S12 | Proportion of interactions, in which both ends are extended by 2.5 kb that overlap **(A)**. DHS, **(B)**. H3K27ac, **(C)**. H3K4me1 or **(D)**. H3K4me3 peaks. DHS: DNase I Hypersensitivity sites.

Supplementary Table S13 | Proportion of interactions, in which both ends are extended by 20 kb that overlap **(A)**. DHS, **(B)**. H3K27ac, **(C)**. H3K4me1 or **(D)**. H3K4me3 peaks. DHS: DNase I Hypersensitivity sites.

REFERENCES

Akdemir, K. C., Le, V. T., Chandran, S., Li, Y., Verhaak, R. G., Beroukhim, R., et al. (2020). Disruption of Chromatin Folding Domains by Somatic Genomic Rearrangements in Human Cancer. *Nat. Genet.* 52 (3), 294–305. doi:10.1038/s41588-019-0564-y

Baxter, J. S., Leavy, O. C., Dryden, N. H., Maguire, S., Johnson, N., Fedele, V., et al. (2018). Capture Hi-C Identifies Putative Target Genes at 33 Breast Cancer Risk Loci. *Nat. Commun.* 9 (1), 1028. doi:10.1038/s41467-018-03411-9

Cai, Y., Zhang, Y., Loh, Y. P., Tng, J. Q., Lim, M. C., Cao, Z., et al. (2021). "H3K27me3-rich Genomic Regions Can Function as Silencers to Repress Gene Expression via Chromatin interactions.". *Nat. Commun.* 12 (1), 1–22. doi:10.1038/s41467-021-20940-y

Cairns, J., Freire-Pritchett, P., Wingett, S. W., Várnai, C., Dimond, A., Plagnol, V., et al. (2016). CHiCAGO: Robust Detection of DNA Looping Interactions in Capture Hi-C Data. *Genome Biol.* 17 (1), 127. doi:10.1186/s13059-016-0992-2

Consortium, E. P. (2012). An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature* 489 (7414), 57–74. doi:10.1038/nature11247

Cremer, T., and Cremer, M. (2010). Chromosome Territories. *Cold Spring Harbor Perspect. Biol.* 2 (3), a003889. doi:10.1101/cshperspect.a003889

Davies, J. O. J., Telenius, J. M., McGowan, S. J., Roberts, N. A., Taylor, S., Higgs, D. R., et al. (2016). Multiplexed Analysis of Chromosome Conformation at Vastly Improved Sensitivity. *Nat. Methods* 13 (1), 74–80. doi:10.1038/nmeth.3664

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing Chromosome Conformation. *science* 295 (5558), 1306–1311. doi:10.1126/science.1067799

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012). Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. *Nature* 485 (7398), 376–380. doi:10.1038/nature11082

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S. P., Huntley, M. H., Lander, E. S., et al. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cel Syst.* 3 (1), 95–98. doi:10.1016/j.cels.2016.07.002

Forcato, M., Nicoletti, C., Pal, K., Livi, C. M., Ferrari, F., and Bicciato, S. (2017). Comparison of Computational Methods for Hi-C Data Analysis. *Nat. Methods* 14 (7), 679–685. doi:10.1038/nmeth.4325

Furlan-Magaril, M., Ando-Kuri, M., Arzate-Mejía, R. G., Morf, J., Román-Figueroa, A., Tenorio-Hernández, L., et al. (2021). The Global and Promoter-Centric 3D Genome Organization Temporally Resolved during a Circadian Cycle. *Genome Biol.* 22 (1), 162. doi:10.1186/s13059-021-02374-3

Gargiulo, G., Levy, S., Bucci, G., Romanenghi, M., Fornasari, L., Beeson, K. Y., et al. (2009). NA-seq: a Discovery Tool for the Analysis of Chromatin Structure and Dynamics during Differentiation. *Develop. Cel.* 16 (3), 466–481. doi:10.1016/j.devcel.2009.02.002

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., et al. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime Cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cel.* 38 (4), 576–589. doi:10.1016/j.molcel.2010.05.004

Holgersen, E. M., Gillespie, A., Leavy, O. C., Baxter, J. S., Zvereva, A., Muirhead, G., et al. (2021). Identifying High-Confidence Capture Hi-C Interactions Using CHiCANE. *Nat. Protoc.* 16 (4), 2257–2285. doi:10.1038/s41596-021-00498-1

Hwang, Y.-C., Lin, C.-F., Valladares, O., Malamon, J., Kuksa, P. P., Zheng, Q., et al. (2015). HIPPIE: a High-Throughput Identification Pipeline for Promoter Interacting Enhancer Elements. *Bioinformatics* 31 (8), 1290–1292. doi:10.1093/bioinformatics/btu801

Imakaev, M., Fudenberg, G., McCord, R. P., Naumova, N., Goloborodko, A., Lajoie, B. R., et al. (2012). Iterative Correction of Hi-C Data Reveals Hallmarks of Chromosome Organization. *Nat. Methods* 9 (10), 999–1003. doi:10.1038/nmeth.2148

Jung, I., Schmitt, A., Diao, Y., Lee, A. J., Liu, T., Yang, D., et al. (2019). A Compendium of Promoter-Centered Long-Range Chromatin Interactions in the Human Genome. *Nat. Genet.* 51 (10), 1442–1449. doi:10.1038/s41588-019-0494-8

Kaul, A., Bhattacharyya, S., and Ay, F. (2020). Identifying Statistically Significant Chromatin Contacts from Hi-C Data with FitHiC2. *Nat. Protoc.* 15 (3), 991–1012. doi:10.1038/s41596-019-0273-0

Lajoie, B. R., Dekker, J., and Kaplan, N. (2015). The Hitchhiker's Guide to Hi-C Analysis: Practical Guidelines. *Methods* 72, 65–75. doi:10.1016/j.ymeth.2014.10.031

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *science* 326 (5950), 289–293. doi:10.1126/science.1181369

Mifsud, B., Tavares-Cadete, F., Young, A. N., Sugar, R., Schoenfelder, S., Ferreira, L., et al. (2015). Mapping Long-Range Promoter Contacts in Human Cells with High-Resolution Capture Hi-C. *Nat. Genet.* 47 (6), 598–606. doi:10.1038/ng.3286

Mifsud, B., Martincorena, I., Darbo, E., Sugar, R., Schoenfelder, S., Fraser, P., et al. (2017). Gothic, a Probabilistic Model to Resolve Complex Biases and to Identify Real Interactions in Hi-C Data. *PloS one* 12 (4), e0174744. doi:10.1371/journal.pone.0174744

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., et al. (2012). Spatial Partitioning of the Regulatory Landscape of the X-Inactivation centre. *Nature* 485 (7398), 381–385. doi:10.1038/nature11049

Osborne, C. S., Chakalova, L., Brown, K. E., Carter, D., Horton, A., Debrand, E., et al. (2004). Active Genes Dynamically Colocalize to Shared Sites of Ongoing Transcription. *Nat. Genet.* 36 (10), 1065–1071. doi:10.1038/ng1423

Pal, K., Forcato, M., and Ferrari, F. (2019). Hi-C Analysis: from Data Generation to Integration. *Biophys. Rev.* 11 (1), 67–78. doi:10.1007/s12551-018-0489-1

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* 159 (7), 1665–1680. doi:10.1016/j.cell.2014.11.021

Rhie, S. K., Perez, A. A., Lay, F. D., Schreiner, S., Shi, J., Polin, J., et al. (2019). A High-Resolution 3D Epigenomic Map Reveals Insights into the Creation

of the Prostate Cancer Transcriptome. *Nat. Commun.* 10 (1), 4154. doi:10.1038/s41467-019-12079-8

Ron, G., Globerson, Y., Moran, D., and Kaplan, T. (2017). Promoter-enhancer Interactions Identified from Hi-C Data Using Probabilistic Models and Hierarchical Topological Domains. *Nat. Commun.* 8 (1), 2237. doi:10.1038/s41467-017-02386-3

Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C. J., Vert, J. P., et al. (2015). HiC-Pro: an Optimized and Flexible Pipeline for Hi-C Data Processing. *Genome Biol.* 16 (1), 259. doi:10.1186/s13059-015-0831-x

Song, M., Yang, X., Ren, X., Maliskova, L., Li, B., Jones, I. R., et al. (2019). Mapping Cis-Regulatory Chromatin Contacts in Neural Cells Links Neuropsychiatric Disorder Risk Variants to Target Genes. *Nat. Genet.* 51 (8), 1252–1262. doi:10.1038/s41588-019-0472-1

Su, J.-H., Zheng, P., Kinrot, S. S., Bintu, B., and Zhuang, X. (2020). Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell* 182 (6), 1641–1659. e1626. doi:10.1016/j.cell.2020.07.032

Wingett, S., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P., et al. (2015). HiCUP: Pipeline for Mapping and Processing Hi-C Data. *F1000Research* 4, 1310. doi:10.12688/f1000research.7334.1

Wolff, J., Bhardwaj, V., Nothjunge, S., Richard, G., Renschler, G., Gilsbach, R., et al. (2018). Galaxy HiCExplorer: a Web Server for Reproducible Hi-C Data Analysis, Quality Control and Visualization. *Nucleic Acids Res.* 46 (W1), W11–W16. doi:10.1093/nar/gky504

Zheng, Y., Ay, F., and Keles, S. (2019). Generative Modeling of Multi-Mapping Reads with mHi-C Advances Analysis of Hi-C Studies. *Elife* 8, e38070. doi:10.7554/eLife.38070

Zouari, Y. B., Molitor, A. M., Sikorska, N., Pancaldi, V., and Sexton, T. (2019). ChiCMaxima: a Robust and Simple Pipeline for Detection and Visualization of Chromatin Looping in Capture Hi-C. *Genome Biol.* 20 (1), 102–119. doi:10.1186/s13059-019-1706-3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, orclaim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Fiber-Like Organization as a Basic Principle for Euchromatin Higher-Order Structure

Amir N. Zakirov [1,2], Sophie Sosnovskaya [1,2], Ekaterina D. Ryumina [1,3], Ekaterina Kharybina [1,2], Olga S. Strelkova [1], Oxana A. Zhironkina [1], Sergei A. Golyshev [1], Andrey Moiseenko [4] and Igor I. Kireev [1,2]*

[1]Department of Electron Microscopy, AN. Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia, [2]Chair of Cell Biology and Histology, Biology Faculty, Lomonosov Moscow State University, Moscow, Russia, [3]Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia, [4]Laboratory of Electron Microscopy, Biology Faculty, Lomonosov Moscow State University, Moscow, Russia

A detailed understanding of the principles of the structural organization of genetic material is of great importance for elucidating the mechanisms of differential regulation of genes in development. Modern ideas about the spatial organization of the genome are based on a microscopic analysis of chromatin structure and molecular data on DNA–DNA contact analysis using Chromatin conformation capture (3C) technology, ranging from the "polymer melt" model to a hierarchical folding concept. Heterogeneity of chromatin structure depending on its functional state and cell cycle progression brings another layer of complexity to the interpretation of structural data and requires selective labeling of various transcriptional states under nondestructive conditions. Here, we use a modified approach for replication timing-based metabolic labeling of transcriptionally active chromatin for ultrastructural analysis. The method allows pre-embedding labeling of optimally structurally preserved chromatin, thus making it compatible with various 3D-TEM techniques including electron tomography. By using variable pulse duration, we demonstrate that euchromatic genomic regions adopt a fiber-like higher-order structure of about 200 nm in diameter (chromonema), thus providing support for a hierarchical folding model of chromatin organization as well as the idea of transcription and replication occurring on a highly structured chromatin template.

Keywords: higher-order chromatin folding, euchromatin, replication, transcription, electron tomography

INTRODUCTION

In the interphase nucleus of higher eukaryotes, DNA displays up to 1,000-fold linear compaction by forming a complex with histones and a set of non-histone proteins—a chromatin. Despite many efforts aiming at elucidation of the DNA folding path in the nucleus, our understanding of how DNA is packaged into chromatin and adopts its conformation is still incomplete. The source of contradiction is rooted from the diversity of experimental techniques used to study chromatin organization at high spatial resolution. Initial attempts were made by analyzing permeabilized cells in order to improve chromatin contrast by removing soluble non-chromatin nuclear material (Zatsepina et al., 1983; Belmont et al., 1989). These studies revealed organization of chromatin into higher-order fibers of 100–130 nm in diameter—chromonema both in interphase and mitotic chromosomes. Careful ultrastructural analysis of cells entering and exiting mitosis demonstrated that

mitotic chromosome compaction/decompaction is achieved by sequential folding/unfolding of chromonema into the fibers of even higher thickness supporting the idea of hierarchical DNA compaction (Belmont and Bruce 1994; Kireeva et al., 2004). These data were criticized on the grounds of possible artifactual chromatin aggregation caused by chromatin-compacting agents (divalent cations and/or polyamines). In contrast, various alternative approaches including live nucleosome tracking (Maeshima et al., 2014), ChromEMT (Ou et al., 2017), and cryo-electron microscopy (Eltsov et al., 2008, 2018; Nishino et al., 2012) that better maintain native chromatin structure failed to demonstrate any signs of hierarchical folding motifs beyond a nucleosome fiber. This has led to formulation of a "polymer melt" model of chromatin organization further supported by the concept of phase separation as a driving force of chromatin organization (Erdel and Rippe, 2018; Mirny et al., 2019). However, regardless of the exact way of DNA folding in the nucleus, it is generally accepted that the local ratio of DNA packaging tightly correlates with transcriptional activity, rendering chromatin subdivided into two fractions—transcriptionally active and centrally located euchromatin and permanently silent or transcriptionally repressed heterochromatin, which preferentially occupies perinucleolar and peripheral areas of the cell nucleus (Solovei et al., 2016; van Steensel and Belmont, 2017). Different chromatin fractions bear specific molecular signatures, such as characteristic histone post-translational modifications, sets of non-histone proteins, and patterns or DNA methylation, which cumulatively contribute to the maintenance of their structural and functional states.

To facilitate transcription, euchromatin is maintained in a more decondensed ("open") state relative to heterochromatin as shown by both microscopy and biochemical assays, yet the degree of its "openness" apparently varies depending on the transcription level. On one extreme of the range lay highly transcribed chromosomal loci demonstrating complete chromatin unfolding, such as puffs in Diptera polytene chromosomes or loops in lampbrush chromosomes (Björk and Wieslander, 2015; Morgan, 2018). On the other hand, the majority of transcribed genes display much lower activity, raising the question whether euchromatin packaging displays variability depending on the transcription activity and what degree of compaction is most typical for it.

Direct imaging of euchromatic genomic loci at high resolution would make an ideal tool to answer this question. However, this approach faces several complications. First, identification of euchromatic loci merely by their relative positioning in the nucleus or by overall compaction state (Ou et al., 2017) may be misleading. Second, traditional ways of chromatin labeling [either by immunocytochemical approaches or by FISH (Boettiger et al., 2016)] require relatively mild fixation conditions required for probe penetration and antigene preservation, which are incompatible with maintenance of native chromatin structure, especially with FISH with its intrinsic harsh denaturation steps, while optimal fixation renders specific labeling ineffective and complicates the 3D analysis if applied to electron microscopy. Using an alternative approach based on analysis of transgenic loci labeled by the LacO/lac-repressor system and *in vivo* immunogold staining seems to preserve chromatin ultrastructure much better (Belmont et al., 2010). However, the artificial chromosomal loci may not faithfully recapitulate the behavior of endogenous genes upon transcription activation, while binding of tagged proteins to chromatin at high density required for good structural resolution may potentially disturb chromatin structure.

Previously, we adapted replicative labeling of DNA with 5-ethynyl-2′-deoxyuridine (EdU, Salic and Mitchison, 2008) to ultrastructural studies of chromatin reorganization during and after replication (Deng et al., 2016). Here, to specifically investigate structural organization of euchromatin with high spatial resolution and optimal structure preservation, we modified this approach based on prolonged replicative labeling to visualize long stretches of DNA (Visser and Aten, 1999) combining EdU labeling with biotin-streptavidin-mediated Nanogold detection scheme and electron tomography. We demonstrate here that the majority of early replicating euchromatin is arranged into higher-order fiber-like structures.

MATERIALS AND METHODS

Cell Labeling and Fixation
HT1080 cells were plated on glass coverslips 1 day before the experiment. For labeling of replicated DNA, cells were incubated in 10 μM EdU (Thermo Fisher Scientific) for 2 h. Cells were fixed with prewarmed (37°C) 2.5% glutaraldehyde on cacodylate buffer (pH 7,2) for 1 h. After washing in PBS with 5 mM $MgCl_2$ (PBS∗) three times for 5 min, free aldehyde groups were quenched with 20 mM glycine in PBS∗ (2 × 10 min). Cells were next permeabilized in 0.1% Triton X-100 in PBS∗ (PBS∗T) twice for 20 min and blocked in 1% BSA in PBS∗ for 1 h.

EdU Detection
EdU was detected according to the Click-IT EdU Imaging Kit protocol with AlexaFluor488-azide or biotin-azide for 40 min. After EdU detection, cells were washed again with BSA-PBS*T buffer 3 times for 5 min, and then washed with deionized water 3 times for 5 min. Streptavidin-Nanogold (Nanoprobes) in BSA-PBS∗ T buffer (1:500) was then added to biotin-azide samples overnight and thoroughly washed with BSA-PBS*T buffer. For cells stained with AlexaFluor488-azide, processing for TEM included detection of AlexaFluor488 with mouse monoclonal antibodies against AlexaFluor488 (Thermo-Fisher) and Nanogold-conjugated goat anti-mouse Fab-fragments (Nanoprobes) at 1:400 dilution for 24 h with the same washes as for streptavidin-Nanogold staining. Cells were post-fixed with 1% glutaraldehyde in PBS buffer for 30 min, washed with 20 mM glycine in PBS∗ and deionized water and then free aldehyde groups were additionally quenched with $NaBH_4$ (1 mg/ml) for 20 min and cells were extensively washed with several changes of deionized water. Fluorescently labeled samples were imaged with Eclipse Ti-E inverted microscope (Nikon) using 60 x 1.4 NA objective and

appropriate filter sets. Z-stacks were recorded with Neo sCMOS camera (Andor) and deconvolved using NIS-Elements 5.3 software package.

Ag-Amplification

Nanogold particles (1.4 nm) were silver-enhanced as described previously (Hainfeld and Furuya, 1992; Kireev et al., 2008). This procedure results in deposition of silver on the surface on Nanogold particles and formation of larger (10–20 nm) silver particles with an Au core, which are readily detected at low-magnification TEM. Briefly, 5 ml of 30% acacia powder solution in deionized water was mixed with 2 ml of 1 M MES (pH = 6.1) in a foil-wrapped 50-ml tube and mixed thoroughly for 30 min by slowly rocking the tube. Right before the procedure, 1.5 ml of freshly prepared 0.2% N-propyl gallate (Fisher) in deionized water was added to acacia powder mix and tube rocked for about 3 min, then 1.5 ml of freshly prepared 0.7% silver lactate in deionized water was added and a mix was rocked for another 2 min. In a dark room, the reaction mix was applied to the cells for 3–4 min and immediately washed out with several changes of deionized water.

Dehydration, Epon Embedding, and Sectioning

Samples were dehydrated in a series of graded ethanol solutions and embedded in Epon 812 (Sigma-Aldrich). Coverslips were then removed by repetitive placement of the samples in liquid N_2 and boiling water. The cells in early S-phase (replication pattern 1 or 2) were located under the bright-field microscope with 20 x lens and the blocks were manually trimmed with the razor blade. Sections that are 250–350 nm thick were cut on the Ultracut E ultramicrotome (Leica) in that serial sections were picked onto Formvar-coated 1-mm single-slot grids. Grids were either stained with 5% aqueous uranyl acetate for 20 min or left unstained, and then carbon-coated.

Electron Tomography

Electron tomography data were acquired with a JEOL JEM-2100 200kV LaB6 transmission electron microscope, equipped with a Gatan GIF Quantum ER energy filter and SerialEM software (Mastronarde, 2005). Images were recorded in EFTEM mode with 20 eV energy-selecting slit, positioned at zero loss, near-parallel condenser illumination conditions, and 0.8 µm defocus. The tilt series was one-axis and one-directional from −60 to +60° with 2° steps. No dose-limiting procedures were carried out, but the tomography region was pre-illuminated with high electron dose. To estimate the resolution, we generated the pair of tomograms from even and odd tilt series subsets independently. The worst resolution estimate is 9 nm at 0.143 FSC criterion. The FSC was calculated across all tomograms including low SNR areas that do not contain nanoparticles.

Tomogram Analysis

The tomogram reconstruction was performed with IMOD software (Kremer et al., 1996) following the standard workflow. Some Ag particles were selected as fiducials for

accurate image alignment. All the subsequent steps were performed in ImageJ. During the first step, the default threshold was applied to the reconstructed tomogram to convert grayscale 3D-stack into a binary image. Then, the "3D Simple segmentation" plugin of the binary image was applied (with default threshold) to a binary image to identify 3D particles as objects.

Subsequent clusterization of those objects is necessary for further analysis of chromatin structure. First, local density was calculated with the "3D Density" plugin. This plugin requires two parameters: "number of neighbors" determines the number of closest neighbors to compute for each pixel, and "radius" determines the radius of expansion from the particle center (hence the radius of resulting fiber). 3D density was calculated for different radii with the number of neighbors set to 40 (variation of both parameters lead to somewhat similar results, so we chose to fix the number of neighbors and iteratively adjust the radius). The number of individual clusters was then plotted against the density calculation radius. The plateau on this plot indicates drastic change in density of golden particles, which can be interpreted as an edge of the object. Usually, two obvious thresholds could be clearly seen. 3D density maps calculated with those threshold radii were further converted to binary images using default brightness threshold. The resulting objects were considered as required chromatin structures.

In order to obtain some metrics of those fibers, first the "3D Distance map" plugin was used. It calculates minimum distance to the edge of the objects for every pixel and assigns it as brightness of this pixel. Since we were interested in overall fiber diameter, we needed to extract central pixels—axis—of the objects. This was achieved by the "Skeletonize 3D" plugin. Then, the skeletonized image and 3D distance map were combined, so that the final image contains only central axial pixels with their intensities indicating the radius (distance from the center to the edge of the object). Density histograms of these images show the distribution of fiber radii.

RESULTS

An important feature discriminating eu- and heterochromatin is replication timing. Microscopic analysis of the spatial distribution of replication foci during S-phase identified a set of specific patterns. These patterns appear in a rather strict temporal order during the genome replication with early patterns colocalizing with euchromatin while late patterns are similar to the distribution of heterochromatin (Nakamura et al., 1986; Manders et al., 1996; Zink et al., 1999). Tight correlation of the replication timing and replication patterns with transcriptional state was further confirmed by molecular approaches (Pope et al., 2013; Zhao et al., 2020). It was also shown that replicative domains (RDs) overlap with topologically associating domains (TADs) identified by Hi-C approach and representing DNA packaging units. The neighboring RDs replicate sequentially so that replication waves started at multiple points along the chromosome at the onset of the
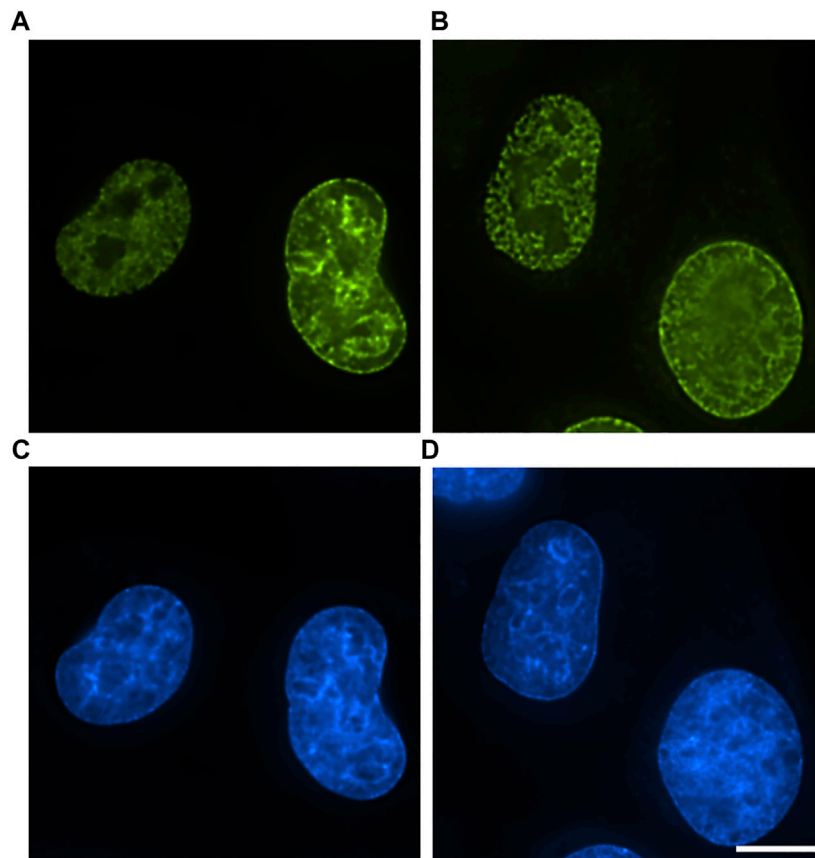
FIGURE 1 | Comparison of labeling efficiency and probe penetration into EdU-labeled cells. **(A)** 10 min formaldehyde fixation, detection with AlexaFluor-azide. **(B)** 1 h glutaraldehyde fixation, detection with biotin-azide–streptavidin-AlexaFluor. Bar, 10 μm. **(C,D)** DAPI staining. Bar 10 μm.

S-phase spread bidirectionally involving first euchromatic regions and then heterochromatin (Pope et al., 2014; Dileep et al., 2015).

In our previous works, we adopted EdU labeling protocol to electron microscopic detection of replicated DNA by using anti-AlexaFluor488 antibodies (Zhironkina et al., 2015; Deng et al., 2016). Using this approach, we obtained high-density labeling of replication foci after permeabilization. We used high-Mg buffer for stabilization of higher-order chromatin structure (Belmont et al., 1989), which allowed us to visualize chromonema fibers with high contrast and demonstrate that replication occurs on highly condensed chromatin template (Deng et al., 2016).

We also noted that early S-phase replication foci at the ultrastructural level represented densely labeled segments of chromonema fiber, suggesting that transcriptionally active chromatin might also fold into higher-order structures. However, since pre-fixation permeabilization of cells in chromatin-condensing conditions may possibly induce chromatin hyper-condensation or even artifactual aggregation, we decided to further investigate this question by developing an experimental approach for ultrastructural replicative labeling under conditions that maximally preserve native chromatin structure.

First, we explored the possibility of application of antibody-mediated EdU labeling on cells fixed directly with gluteraldehyde. We found that gluteraldehyde fixation is widely used for electron microscopy of cells and tissues and superbly preserving cellular ultrastructure does not interfere significantly with Click-reaction if special care is taken to quench free aldehyde groups (**Figure 1B**). However, direct application of anti-fluorochrome antibodies demonstrated that glutaraldehyde fixation creates a diffusional barrier to antibodies as we observed clear gradient of labeling efficiency with peripheral chromatin labeled rather densely but the labeling density rapidly fading towards the nuclear interior (**Supplementary Figure 2**). To avoid this problem, we substituted fluorochrome-azide with biotin-azide and used Streptavidin-Nanogold for labeling. Reduction of the probe size and introduction of one-step labeling procedure allowed to obtain a uniform labeling throughout the nucleus while maintaining labeling intensity roughly at the same level with substantial gain in S/N ratio even in glutaraldehyde-crosslinked cells (**Figures 1A,B**).

Next, we decided to test the hypothesis whether early-replicated euchromatin is organized into higher-order fibrillar structures. Since direct fixation does not allow to easily identify loosely packed chromatin on the background of non-chromatin nuclear components, we relied upon prolonged replicative labeling in order
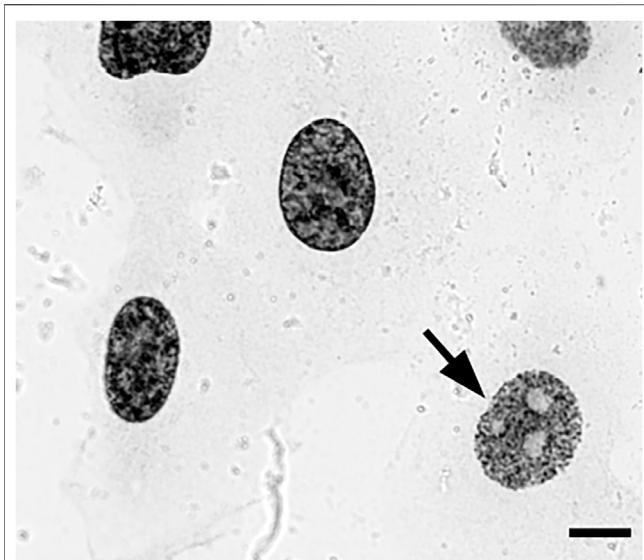
FIGURE 2 | EdU pulse-labeled HT1080 cells after biotin-azide streptavidin-Nanogold detection and embedding demonstrate clear replication patterns en bloc under a transmitted light microscope, enabling selection of cells with early-S patterns (arrow) for sectioning. Bar, 10 μm.

to visualize long DNA stretches representing several neighboring RDs and trace its folding based exclusively on amplified Nanogold particle distribution. For this reason, we chose 2-h EdU pulses, which would give a continuous labeling of at least three neighboring RDs [provided an average replication timing of a single domain in early-S estimated between 45 and 60 min (Ma et al., 1998)] while staying within a time frame of euchromatin replication that occurs during the first 3 h of S-phase in HT1080 cells (Deng et al., 2016). Since the estimates of a chromonema fiber (and RDs) size fall in the range between 100 and 200 nm (Zatsepina et al., 1983; Belmont et al., 1989; Belmont and Bruce 1994; Kireeva et al., 2004; Kireev et al., 2008; Rego et al., 2008; Cseresnyes et al., 2009; Baddeley et al., 2010; Su et al., 2020), in order to accurately measure the size of higher-order chromatin structures in 3D, we employed electron tomography of 300-nm plastic sections. The cells in early S-phase were selected for sectioning based on the pattern of RD distribution (**Supplementary Figures S1A–E**). The corresponding patterns are easily detected by bright-field microscopy after Ag amplification (**Figure 2**) and angular projections were collected from equatorial sections of the nuclei. Already on raw images, distribution of replicative label over the higher-order fibrillar chromatin structures becomes clearly seen (**Figures 3A,B**). The fibers were almost uniformly distributed over the nuclear interior excluding the nucleolus, as expected for euchromatin. Provided high labeling density, we could measure the fiber thickness by first calculating the silver-enhanced Au nanoparticle 3D density map on tomographic reconstructions using the ImageJ 3D Density plugin (**Figure 4**). The radius for counting the neighboring particles was determined by iteratively calculating 3D density map and measuring the number of objects after segmentation of the resulting map using the default threshold. A plateau on plots of the number of objects as a function of the radius chosen indicates a drastic change in density of golden particles (**Supplementary Figure**

S3), which can be interpreted as an edge of the object. These radii were subsequently used for further calculations. In the majority of samples collected from five tomograms, two plateaus were detected (**Supplementary Figure S4**), indicating the existence of either two size classes or, most probably, variability in thickness within a single fibrillar structure.

Next, local thickness of the 3D objects representing segments of labeled chromatin higher-order structures was calculated by applying 3D Distance map and Skeletonize2D/3D plugins (**Figure 4**). The resulting histograms demonstrate the modal radius of the labeled fibers, calculated separately for two neighborhood radii, to be 74,5 nm (S.D. 26,77 nm, SE 0.2 nm) and 90,02 nm (S.D. 28,71 nm, SE 0.2 nm), which correspond to ~150–180 nm fiber diameter.

DISCUSSION

In our previous work aimed at the analysis of transcription-dependent chromatin rearrangement using engineered chromosome loci containing inducible genes, we observed expected unfolding of BAC arrays upon transcription induction (Hu et al., 2009). However, estimations of linear compaction ratios based on measurements of FISH signals and direct immunoelectron visualization of the transgenes suggested that overall packing ratio
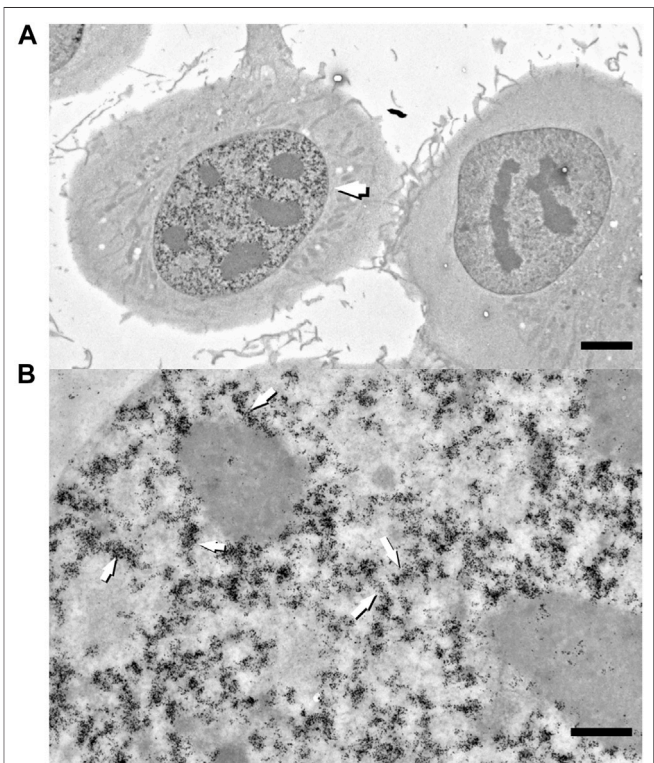
FIGURE 3 | Raw zero-tilt image of a 250-nm-thick section of HT1080 cell in early S-phase labeled with 2 h pulse of EdU with subsequent detection with biotin-azide–streptavidin-Nanogold **(A)**. Segments of labeled fiber-like chromatin structures (arrows) are randomly distributed throughout nuclear interior. Bar, 1 μm **(A)** and 0.5 μm **(B)**.
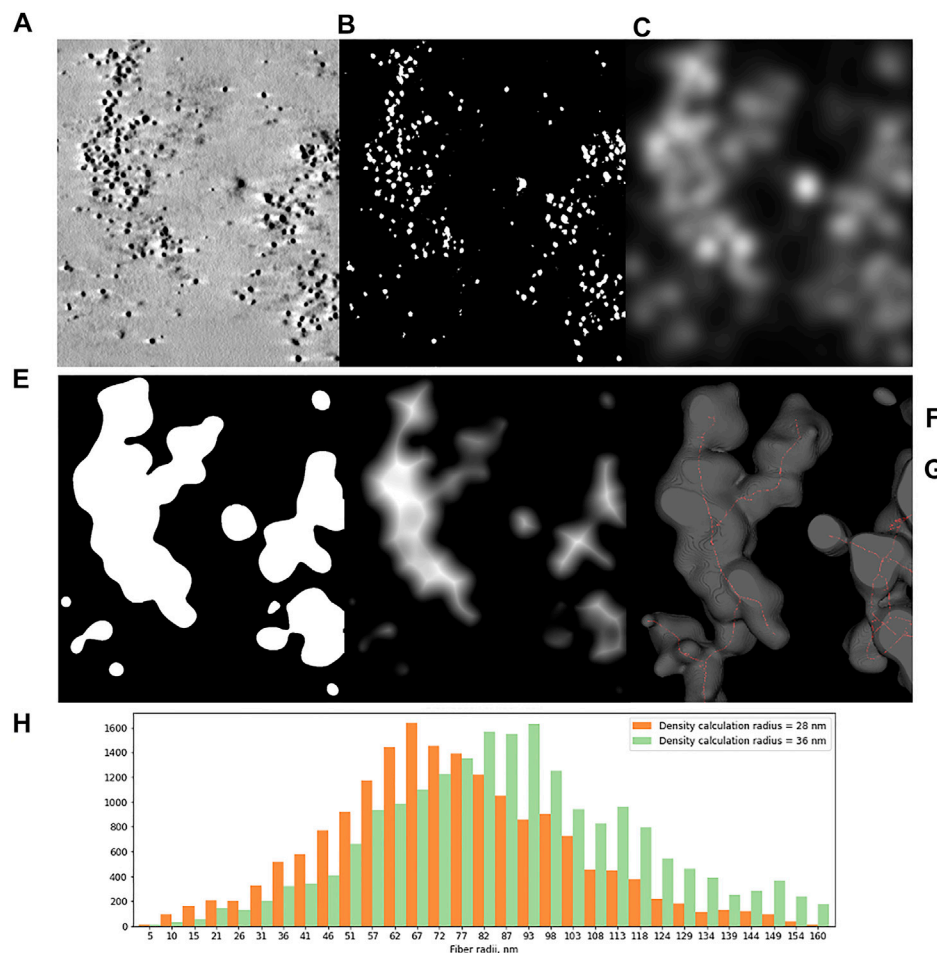
FIGURE 4 | Tomographic slice of HT1080 cell nucleus labeled with 2 h pulse of EdU and detected with biotin-azide and streptavidin-Nanogold **(A)** and main steps of image analysis. **(B–F) (B)** Segmentation of Ag-enhanced Au particles. **(C)** 3D density calculation for different radii. The plot demonstrates total number of clusters depending on density calculation radius, green lines indicating the thresholds used for further calculations. **(D)** 3D density map. **(E)** Thresholded 3D density map. **(F)** 3D Distance map; Bar, 500 nm. **(G)** Histograms of local thickness distribution, calculated for two clustering thresholds show modal radii of higher-order chromatin fibers between 75 and 90 nm.

(400- to 1,000-fold) of the transcribed loci remains well above expected for nucleosome fiber. This suggests that transcription may occur on highly condensed chromatin template. Obvious technical limitations of the approaches used and a desire to extend our study to endogenous loci have inspired us to design a method that would (1) allow for selective labeling of total euchromatin, (2) minimally perturb native chromatin structure, and (3) be easily compatible with high-resolution 3D analysis.

We decided to make use of metabolic DNA labeling with alkyne derivatives of nucleotides (EdU; Salic and Mitchison, 2008). When applied in the early S-phase, EdU is incorporated into transcriptionally active DNA, thus labeling euchromatin fraction. Since EdU detection with Click-chemistry does not require any DNA-perturbing treatments (denaturation enzymatic digestion, etc.), it is more compatible with high-resolution imaging compared to previously used halogenated nucleotides (Visser et al., 2000; Jaunin and Fakan, 2002). In our first attempt, we used a rather mild fixation protocol on pre-extracted nuclei and found that replication label in euchromatin at sites of replication and well after the replication is completed is distributed over distinct higher-order fiber-like structures (Deng et al., 2016). However, suboptimal fixation conditions used may potentially generate artifacts (Amiad-Pavlov et al., 2021) despite the attempts to stabilize chromatin structure with Mg^{++}, so we decided to switch to more robust fixation protocol using glutaraldehyde in order to reproduce this label distribution in nearly intact chromatin. Generation of electron contrast for visualization of the label by TEM may be problematic since glutaraldehyde fixation, while optimally preserving nuclear ultrastructure (Fussner et al., 2012), limits large probe accessibility to the nuclear interior. Indeed, we mentioned dramatic differences in probe penetration between small fluorescently labeled azides and secondary gold-conjugated antibodies. DAB photooxidation as an alternative to gold particles (Ngo et al., 2016), although offering more uniform distribution in glutaraldehyde-fixed cells, gives much lower contrast and requires higher density labeling, which may disturb native chromatin

organization. Thus, we switched to a multi-step protocol, which employs smaller-size components at each step, cumulatively achieving uniform labeling of chromatin throughout the nucleus. For this purpose, we used the biotin-streptavidin system to link EdU to Nanogold particles with subsequent silver enhancement for high contrast labeling detectable both at low-magnification optical microscopy and TEM. The ability to discern label distribution with bright-field optical microscopy greatly facilitated pre-selection of nuclei with labeled euchromatin for sectioning and tomography. This protocol has several advantages over previously published ones (Vogel et al., 1990; Jaunin and Fakan, 2002; Philimonenko et al., 2004; Koberna et al., 2005; Deng et al., 2016). Glutaraldehyde fixation ensures optimal preservation of chromatin near-native structure. Click-chemistry provides simple and extremely selective labeling of replicated DNA, without the need of DNA denaturation prerequisite for BrdU detection with antibodies used in previous reports. The use of streptavidin-Nanogold conjugates provides better penetration efficiency even into glutaraldehyde-fixed samples due to the relatively small size of the probe. Overall our protocol allows for high-contrast high-efficiency pre-embedding labeling compatible with various 3D-electron microscopy techniques.

This protocol has allowed us to visualize continuously labeled segments of the genome corresponding to euchromatin, in three dimensions. Although relatively large size of silver particles does not allow for tracing nucleosome chain folding paths as offered by ChromEMT (Ou et al., 2017) or cryoET (Eltsov et al., 2018), we achieved a satisfactory labeling density for visualization of higher-order chromatin structures more than 100 nm thick. Surprisingly, we found that the majority of chromatin labeled in early S-phase is forming fiber-like structures of the thickness ranging from 130 to 200 nm. These estimations closely correlate to the measurements obtained in the cells permeabilized in chromatin-stabilizing conditions (Belmont and Bruce 1994; Kireeva et al., 2004; Rego et al., 2008), as well as both *in vivo* and *in situ* labeled engineered chromosome loci (Strukov et al., 2003; Kireev et al., 2008; Hu et al., 2009), further supporting the idea of hierarchical folding principle of chromatin organization and suggesting that this multi-step folding is characteristic not only for highly condensed repressed heterochromatin, but also for transcriptionally active euchromatin. At the first glance, this idea contradicts a generally accepted principle of correlation between transcription activity and chromatin compaction. However, various experimental approaches have demonstrated that chromatin displays a high degree of local dynamics, which would allow for both accessibility of the genes for transcription-related *trans*-factors and local and temporal DNA conformational changes required for transcription and, more broadly, for any type of activity involving DNA (replication, repair, etc.) (Deng et al., 2016; Nozaki et al., 2018). We can propose this dynamics to occur within higher-order chromatin domains, which are aligned in the cell nucleus into chromonema fibers. One can argue that apparent fiber-like structures may result from an exclusion of euchromatic domains from chromosome territories, as proposed by the model of interchromatin domain (Cremer et al., 2020). However, high-resolution FISH analysis of contiguous DNA segment containing transcriptionally active genes

rather suggests a fiber-like folding (Volpi et al., 2000; Muller et al., 2004; Goetze et al., 2007; Hu et al., 2009).

Chromatin dynamics experiments, as well as high-resolution tracing of nucleosome chains *in situ* and x-ray scattering, suggest the absence of any regular folding of 10-nm nucleosome fiber (Fussner et al., 2012; Nishino et al., 2012; Ricci et al., 2015; Nozaki et al., 2017; Ou et al., 2017). Yet, higher-order chromatin fiber-like structures may be formed by irregularly packed nucleosome chains, represented by a series of DNA loops of varying size, whose state of compaction is controlled by phase-separation mechanisms and/or a combination of loop-forming activities. Recent studies including Hi-C and single-molecule imaging suggested the dynamic nature of the loops with cohesin complex serving as both a loop-forming and a loop-maintaining factor (Gassler et al., 2017; Wutz et al., 2017; Nuebler et al., 2018; Davidson et al., 2019). A dynamic loop extrusion process further contributes to the variability in overall linear parameters of the chromonema fibers, partially explaining a rather wide range of thicknesses measured on electron tomograms. Fluctuations of local transcription activity may represent another source of variability in chromonema structure, which would require more accurate assessment of chromatin folding of individual genes at high resolution and maximal structural preservation.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

Conceptualization: IK and OZ; methodology: OZ and OS; investigation: AZ, SS, ER, EK, and SG; electron tomography: AM, ER, AZ, and SG; data analysis: ER and AZ; writing and editing: IK and OZ.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.784440/full#supplementary-material

REFERENCES

Amiad-Pavlov, D., Lorber, D., Bajpai, G., Reuveny, A., Roncato, F., Alon, R., et al. (2021). Live Imaging of Chromatin Distribution Reveals Novel Principles of Nuclear Architecture and Chromatin Compartmentalization. *Sci. Adv.* 7 (23), eabf6251. doi:10.1126/sciadv.abf6251

Baddeley, D., Chagin, V. O., Schermelleh, L., Martin, S., Pombo, A., Carlton, P. M., et al. (2010). Measurement of Replication Structures at the Nanometer Scale Using Super-resolution Light Microscopy. *Nucleic Acids Res.* 38, e8. doi:10.1093/nar/gkp901

Belmont, A. S., and Bruce, K. (1994). Visualization of G1 Chromosomes: a Folded, Twisted, Supercoiled Chromonema Model of Interphase Chromatid Structure. *J. Cel Biol.* 127, 287–302. doi:10.1083/jcb.127.2.287

Belmont, A. S., Braunfeld, M. B., Sedat, J. W., and Agard, D. A. (1989). Large-scale Chromatin Structural Domains within Mitotic and Interphase Chromosomes *In Vivo* and *In Vitro*. *Chromosoma* 98, 129–143. doi:10.1007/BF00291049

Belmont, A. S., Hu, Y., Sinclair, P. B., Wu, W., Bian, Q., and Kireev, I. (2010). Insights into Interphase Large-Scale Chromatin Structure from Analysis of Engineered Chromosome Regions. *Cold Spring Harbor Symp. Quant. Biol.* 75, 453–460. doi:10.1101/sqb.2010.75.050

Björk, P., and Wieslander, L. (2015). The Balbiani Ring Story: Synthesis, Assembly, Processing, and Transport of Specific Messenger RNA-Protein Complexes. *Annu. Rev. Biochem.* 84, 65–92. doi:10.1146/annurev-biochem-060614-034150

Boettiger, A. N., Bintu, B., Moffitt, J. R., Wang, S., Beliveau, B. J., Fudenberg, G., et al. (2016). Super-resolution Imaging Reveals Distinct Chromatin Folding for Different Epigenetic States. *Nature* 529, 418–422. doi:10.1038/nature16496

Cremer, T., Cremer, M., Hübner, B., Silahtaroglu, A., Hendzel, M., Lanctôt, C., et al. (2020). The Interchromatin Compartment Participates in the Structural and Functional Organization of the Cell Nucleus. *BioEssays* 42, 1900132. doi:10.1002/bies.201900132

Cseresnyes, Z., Schwarz, U., and Green, C. M. (2009). Analysis of Replication Factories in Human Cells by Super-resolution Light Microscopy. *BMC Cel Biol* 10, 88. doi:10.1186/1471-2121-10-88

Davidson, I. F., Bauer, B., Goetz, D., Tang, W., Wutz, G., and Peters, J.-M. (2019). DNA Loop Extrusion by Human Cohesin. *Science* 366, 1338–1345. doi:10.1126/science.aaz3418

Deng, X., Zhironkina, O. A., Cherepanynets, V. D., Strelkova, O. S., Kireev, I. I., and Belmont, A. S. (2016). Cytology of DNA Replication Reveals Dynamic Plasticity of Large-Scale Chromatin Fibers. *Curr. Biol.* 26, 2527–2534. doi:10.1016/j.cub.2016.07.020

Dileep, V., Rivera-Mulia, J. C., Sima, J., and Gilbert, D. M. (2015). Large-Scale Chromatin Structure-Function Relationships during the Cell Cycle and Development: Insights from Replication Timing. *Cold Spring Harb. Symp. Quant. Biol.* 80, 53–63. doi:10.1101/sqb.2015.80.027284

Eltsov, M., MacLellan, K. M., Maeshima, K., Frangakis, A. S., and Dubochet, J. (2008). Analysis of Cryo-Electron Microscopy Images Does Not Support the Existence of 30-nm Chromatin Fibers in Mitotic Chromosomes *In Situ*. *Proc. Natl. Acad. Sci.* 105, 19732–19737. doi:10.1073/pnas.0810057105

Eltsov, M., Grewe, D., Lemercier, N., Frangakis, A., Livolant, F., and Leforestier, A. (2018). Nucleosome Conformational Variability in Solution and in Interphase Nuclei Evidenced by Cryo-Electron Microscopy of Vitreous Sections. *Nucleic Acids Res.* 46, 9189–9200. doi:10.1093/nar/gky670

Erdel, F., and Rippe, K. (2018). Formation of Chromatin Subcompartments by Phase Separation. *Biophysical J.* 114, 2262–2270. doi:10.1016/j.bpj.2018.03.011

Fussner, E., Strauss, M., Djuric, U., Li, R., Ahmed, K., Hart, M., et al. (2012). Open and Closed Domains in the Mouse Genome Are Configured as 10-nm Chromatin Fibres. *EMBO Rep.* 13, 992–996. doi:10.1038/embor.2012.139

Gassler, J., Brandão, H. B., Imakaev, M., Flyamer, I. M., Ladstätter, S., Bickmore, W. A., et al. (2017). A Mechanism of Cohesin-dependent Loop Extrusion Organizes Zygotic Genome Architecture. *EMBO J.* 36, 3600–3618. doi:10.15252/embj.201798083

Goetze, S., Mateos-Langerak, J., Gierman, H. J., de Leeuw, W., Giromus, O., Indemans, M. H. G., et al. (2007). The Three-Dimensional Structure of Human Interphase Chromosomes Is Related to the Transcriptome Map. *Mol. Cel. Biol.* 27 (12), 4475–4487. doi:10.1128/MCB.00208-07

Hainfeld, J. F., and Furuya, F. R. (1992). A 1.4-nm Gold Cluster Covalently Attached to Antibodies Improves Immunolabeling. *J. Histochem. Cytochem.* 40 (2), 177–184. doi:10.1177/40.2.1552162

Hu, Y., Kireev, I., Plutz, M., Ashourian, N., and Belmont, A. S. (2009). Large-scale Chromatin Structure of Inducible Genes: Transcription on a Condensed, Linear Template. *J. Cel Biol.* 185, 87–100. doi:10.1083/jcb.200809196

Jaunin, F., and Fakan, S. (2002). DNA Replication and Nuclear Architecture. *J. Cel. Biochem.* 85, 1–9. doi:10.1002/jcb.10115

Kireev, I., Lakonishok, M., Liu, W., Joshi, V. N., Powell, R., and Belmont, A. S. (2008). *In Vivo* immunogold Labeling Confirms Large-Scale Chromatin Folding Motifs. *Nat. Methods* 5, 311–313. doi:10.1038/nmeth.1196

Kireeva, N., Lakonishok, M., Kireev, I., Hirano, T., and Belmont, A. S. (2004). Visualization of Early Chromosome Condensation: A Hierarchical Folding, Axial Glue Model of Chromosome Structure. *J. Cel Biol.* 166, 775–785. doi:10.1083/jcb.200406049

Koberna, K., Ligasová, A., Malínský, J., Pliss, A., Siegel, A. J., Cvačková, Z., et al. (2005). Electron Microscopy of DNA Replication in 3-D: Evidence for Similar-Sized Replication Foci throughout S-phase. *J. Cel. Biochem.* 94 (1), 126–138. doi:10.1002/jcb.20300

Kremer, J. R., Mastronarde, D. N., and McIntosh, J. R. (1996). Computer Visualization of Three-Dimensional Image Data Using IMOD. *J. Struct. Biol.* 116, 71–76. doi:10.1006/jsbi.1996.0013

Ma, H., Samarabandu, J., Devdhar, R. S., Acharya, R., Cheng, P.-c., Meng, C., et al. (1998). Spatial and Temporal Dynamics of DNA Replication Sites in Mammalian Cells. *J. Cel Biol.* 143, 1415–1425. doi:10.1083/jcb.143.6.1415

Maeshima, K., Imai, R., Tamura, S., and Nozaki, T. (2014). Chromatin as Dynamic 10-nm Fibers. *Chromosoma* 123, 225–237. doi:10.1007/s00412-014-0460-2

Manders, E. M. M., Stap, J., Strackee, J., van Driel, R., and Aten, J. A. (1996). Dynamic Behavior of DNA Replication Domains. *Exp. Cel Res.* 226, 328–335. doi:10.1006/excr.1996.0233

Mastronarde, D. N. (2005). Automated Electron Microscope Tomography Using Robust Prediction of Specimen Movements. *J. Struct. Biol.* 152, 36–51. doi:10.1016/j.jsb.2005.07.007

Mirny, L. A., Imakaev, M., and Abdennur, N. (2019). Two Major Mechanisms of Chromosome Organization. *Curr. Opin. Cel Biol.* 58, 142–152. doi:10.1016/j.ceb.2019.05.001

Morgan, G. T. (2018). Imaging the Dynamics of Transcription Loops in Living Chromosomes. *Chromosoma* 127, 361–374. doi:10.1007/s00412-018-0667-8

Müller, W. G., Rieder, D., Kreth, G., Cremer, C., Trajanoski, Z., and McNally, J. G. (2004). Generic Features of Tertiary Chromatin Structure as Detected in Natural Chromosomes. *Mol. Cel. Biol.* 24, 9359–9370. doi:10.1128/MCB.24.21.9359-9370.2004

Nakamura, H., Morita, T., and Sato, C. (1986). Structural Organizations of Replicon Domains during DNA Synthetic Phase in the Mammalian Nucleus. *Exp. Cel Res.* 165, 291–297. doi:10.1016/0014-4827(86)90583-5

Ngo, J. T., Adams, S. R., Deerinck, T. J., Boassa, D., Rodriguez-Rivera, F., Palida, S. F., et al. (2016). Click-EM for Imaging Metabolically Tagged Nonprotein Biomolecules. *Nat. Chem. Biol.* 12, 459–465. doi:10.1038/nchembio.2076

Nishino, Y., Eltsov, M., Joti, Y., Ito, K., Takata, H., Takahashi, Y., et al. (2012). Human Mitotic Chromosomes Consist Predominantly of Irregularly Folded Nucleosome Fibres without a 30-nm Chromatin Structure. *EMBO J.* 31, 1644–1653. doi:10.1038/emboj.2012.35

Nozaki, T., Imai, R., Tanbo, M., Nagashima, R., Tamura, S., Tani, T., et al. (2017). Dynamic Organization of Chromatin Domains Revealed by Super-Resolution Live-Cell Imaging. *Mol. Cel.* 67, 282–293.e7. doi:10.1016/j.molcel.2017.06.018

Nozaki, T., Hudson, D. F., Tamura, S., and Maeshima, K. (2018). Dynamic Chromatin Folding in the Cell," in *Translational Epigenetics, Nuclear Architecture and Dynamics*. Editor C. Lavelle and J. -M. Victor (Cambridge, MA: Academic Press), 101–122. doi:10.1016/B978-0-12-803480-4.00004-1

Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N., and Mirny, L. A. (2018). Chromatin Organization by an Interplay of Loop Extrusion and Compartmental Segregation. *Proc. Natl. Acad. Sci. USA* 115, E6697–E6706. doi:10.1073/pnas.1717730115

Ou, H. D., Phan, S., Deerinck, T. J., Thor, A., Ellisman, M. H., and O'Shea, C. C. (2017). ChromEMT: Visualizing 3D Chromatin Structure and Compaction in Interphase and Mitotic Cells. *Science* 357, eaag0025. doi:10.1126/science.aag0025

Philimonenko, A. A., Jackson, D. A., Hodný, Z., Janáček, J., Cook, P. R., and Hozák, P. (2004). Dynamics of DNA Replication: an Ultrastructural Study. *J. Struct. Biol.* 148 (3), 279–289. doi:10.1016/j.jsb.2004.08.001

Pope, B. D., Aparicio, O. M., and Gilbert, D. M. (2013). SnapShot: Replication Timing. *Cell* 152, 1390–1390.e1. doi:10.1016/j.cell.2013.02.038

Pope, B. D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., et al. (2014). Topologically Associating Domains Are Stable Units of Replication-Timing Regulation. *Nature* 515, 402–405. doi:10.1038/nature13986

Rego, A., Sinclair, P. B., Tao, W., Kireev, I., and Belmont, A. S. (2008). The Facultative Heterochromatin of the Inactive X Chromosome Has a Distinctive Condensed Ultrastructure. *J. Cel Sci.* 121, 1119–1127. doi:10.1242/jcs.026104

Ricci, M. A., Manzo, C., García-Parajo, M. F., Lakadamyali, M., and Cosma, M. P. (2015). Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes *In Vivo. Cell* 160, 1145–1158. doi:10.1016/j.cell.2015.01.054

Salic, A., and Mitchison, T. J. (2008). A Chemical Method for Fast and Sensitive Detection of DNA Synthesis *In Vivo. Proc. Natl. Acad. Sci.* 105, 2415–2420. doi:10.1073/pnas.0712168105

Solovei, I., Thanisch, K., and Feodorova, Y. (2016). How to rule the nucleus: divide et impera. *Curr. Opin. Cel Biol.* 40, 47–59. doi:10.1016/j.ceb.2016.02.014

Strukov, Y. G., Wang, Y., and Belmont, A. S. (2003). Engineered Chromosome Regions with Altered Sequence Composition Demonstrate Hierarchical Large-Scale Folding within Metaphase Chromosomes. *J. Cel Biol.* 162, 23–35. doi:10.1083/jcb.200303098

Su, Q. P., Zhao, Z. W., Meng, L., Ding, M., Zhang, W., Li, Y., et al. (2020). Superresolution Imaging Reveals Spatiotemporal Propagation of Human Replication Foci Mediated by CTCF-Organized Chromatin Structures. *Proc. Natl. Acad. Sci. USA* 117, 15036–15046. doi:10.1073/pnas.2001521117

van Steensel, B., and Belmont, A. S. (2017). Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* 169, 780–791. doi:10.1016/j.cell.2017.04.022

Visser, A. E., and Aten, J. A. (1999). Chromosomes as Well as Chromosomal Subdomains Constitute Distinct Units in Interphase Nuclei. *J. Cel Sci.* 112 (Pt 19), 3353–3360. doi:10.1242/jcs.112.19.3353

Visser, A. E., Jaunin, F., Fakan, S., and Aten, J. A. (2000). High Resolution Analysis of Interphase Chromosome Domains. *J. Cel Sci.* 113 (Pt 14), 2585–2593. doi:10.1242/jcs.113.14.2585

Vogel, W., Mehnert, K., and Pentz, S. (1990). Demonstration of Chromosome Replication by BrdU Antibody Technique and Electron Microscopy. *Hum. Genet.* 84 (3), 237–240. doi:10.1007/BF00200566

Volpi, E. V., Chevret, E., Jones, T., Vatcheva, R., Williamson, J., Beck, S., et al. (2000). Large-scale Chromatin Organization of the Major Histocompatibility Complex and Other Regions of Human Chromosome 6 and its Response to Interferon in Interphase Nuclei. *J. Cel Sci.* 113, 1565–1576. doi:10.1242/jcs.113.9.1565

Wutz, G., Várnai, C., Nagasaka, K., Cisneros, D. A., Stocsits, R., Tang, W., et al. (2017). Topologically Associating Domains and Chromatin Loops Depend on Cohesin and are Regulated by CTCF, WAPL, and PDS5 Proteins*EMBO J.* 36, 3573–3599. doi:10.1101/177444

Zatsepina, O. V., Polyakov, V. Y., and Chentsov, Y. S. (1983). Chromonema and Chromomere. *Chromosoma* 88, 91–97. doi:10.1007/BF00327327

Zhao, P. A., Sasaki, T., and Gilbert, D. M. (2020). High-resolution Repli-Seq Defines the Temporal Choreography of Initiation, Elongation and Termination of Replication in Mammalian Cells. *Genome Biol.* 21, 76. doi:10.1186/s13059-020-01983-8

Zhironkina, O. A., Kurchashova, S. Y., Brattseva, A. L., Cherepaninets, V. D., Strelkova, O. S., Belmont, A. S., et al. (2015). Overcoming Steric Hindrances during Replication of Peripheral Heterochromatin. *Cell Tiss. Biol.* 9, 110–118. doi:10.1134/S1990519X15020121

Zink, D., Bornfleth, H., Visser, A., Cremer, C., and Cremer, T. (1999). Organization of Early and Late Replicating DNA in Human Chromosome Territories. *Exp. Cel Res.* 247, 176–188. doi:10.1006/excr.1998.4311

Application of the 3C Method to Study the Developmental Genes in *Drosophila* Larvae

Oleg V. Bylino[1], Airat N. Ibragimov[1,2], Filomena Anna Digilio[3], Ennio Giordano[4] and Yulii V. Shidlovskii[1,5]*

[1]Department of Gene Expression Regulation in Development, Institute of Gene Biology, Russian Academy of Sciences, Moscow, Russia, [2]Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Gene Biology, Russian Academy of Sciences, Moscow, Russia, [3]Research Institute on Terrestrial Ecosystems (IRET), UOS Naples-CNR, Naples, Italy, [4]Department of Biology, Università di Napoli Federico II, Naples, Italy, [5]Department of Biology and General Genetics, I.M. Sechenov First Moscow State Medical University, Moscow, Russia

A transition from one developmental stage to another is accompanied by activation of developmental programs and corresponding gene ensembles. Changes in the spatial conformation of the corresponding loci are associated with this activation and can be investigated with the help of the Chromosome Conformation Capture (3C) methodology. Application of 3C to specific developmental stages is a sophisticated task. Here, we describe the use of the 3C method to study the spatial organization of developmental loci in *Drosophila* larvae. We critically analyzed the existing protocols and offered our own solutions and the optimized protocol to overcome limitations. To demonstrate the efficiency of our procedure, we studied the spatial organization of the developmental locus *Dad* in 3rd instar *Drosophila* larvae. Differences in locus conformation were found between embryonic cells and living wild-type larvae. We also observed the establishment of novel regulatory interactions in the presence of an adjacent transgene upon activation of its expression in larvae. Our work fills the gap in the application of the 3C method to *Drosophila* larvae and provides a useful guide for establishing 3C on an animal model.

Keywords: chromatin conformation capture, distal interaction, larvae, chromatin, enhancer, promoter, *Drosophila*

1 INTRODUCTION

The study of the basis of changes in the repertoire of active genes associated with the implementation of specific development programs is an important task of modern developmental biology. One of the basic changes of gene expression lies in the events that occur with a chromatin template. Current views on gene activity suggest that significant changes in chromatin conformation of corresponding loci accompany the developmental processes regardless of whether the development is discrete, like

Abbreviations: 3C, Chromosome Conformation Capture; BAC, bacterial artificial chromosome; BAP170, Brahma-associated protein 170 kD; BMP, bone morphogenic protein; DTT, dithiothreitol; DBD, DNA binding domain; EB, extraction buffer; EDTA, ethylenediaminetetraacetic acid; FA, formaldehyde; qPCR, quantitative polymerase chain reaction; PBS, phosphate-buffered saline; FA, formaldehyde; PFA, paraformaldehyde; Ph/Chl, Phenol/Chloroform; PHF10, PHD finger protein 10; PIC, protease inhibitor cocktail; RB, restriction buffer; RE, restriction enzyme; RLF, relative ligation frequency; RNAPII, RNA polymerase II; RT, room temperature; RT-qPCR, reverse transcription-quantitative PCR; S2 cells, Schneider 2 cells; SAYP, supporter of activation of yellow protein; SDS, sodium dodecylsulfate; TGF-β, Transforming growth factor beta; WT, wild type.

in *Drosophila*, or continuous, like in mammals, and can be investigated using Chromosome Conformation Capture (3C) methods (Du et al., 2017; Flyamer et al., 2017; Ke et al., 2017; Ogiyama et al., 2018; Sun et al., 2019; Collombet et al., 2020).

Cell cultures and embryos are the focus of many studies performed on *Drosophila* using C-methods. 3C protocols suitable for cell cultures are widespread, while experiments with embryos are less common (Sexton et al., 2012; Webber et al., 2013; Stadler et al., 2017; Hug and Vaquerizas, 2018; Erceg et al., 2019), and only a few experiments have been performed with individual tissues, for example, salivary glands of wandering 3rd instar larvae (Eagen et al., 2015), wing imaginal discs (Li, 2016; Vizcaya-Molina et al., 2018), eye–antennal imaginal discs (Loubiere et al., 2020), imaginal discs (not specified) (Bernardo et al., 2014), fat body (Bernardo et al., 2014), and larval brain (Tolhuis et al., 2011). The same situation is observed in studies where whole larvae are used. There is a significant gap regarding the 3C procedure for the whole 3rd instar *Drosophila* larvae. This stage is of special interest since 3rd instar larvae have the largest and best developed imaginal discs and histoblast nests, which can be considered as non-specialized precursors of terminally differentiated adult cells that give origin to tissues of an adult fly during metamorphosis (Gilbert, 2010). Thus, understanding the differences in the state of genes between the last larval stage and adult flies, including the level of the spatial organization of chromatin, is very important for understanding how one stage of development turns into another, the terminal stage of an adult insect.

The study of the larval stages is a separate and rather difficult task. For example, *Drosophila* larvae of the 1st, 2nd, and middle 3rd instars live in fly food, and larvae of the late 3rd instar appear on the walls of fly tubes only transiently, before pupation. It is, therefore, difficult to collect a large amount of 3rd instar larvae. It is also difficult to collect large amounts of 1st and 2nd instar larvae, which are rather small in size.

A common feature of the abovementioned experimental works with *Drosophila* tissues is that the step of tissue fixation and homogenization yields a cellular material (cell suspension), which can further be processed using any type of the 3C protocol, including an *in situ*/in-nucleus ligation protocol (Comet et al., 2011; Rao et al., 2014; Nagano et al., 2015a, Nagano et al., 2015b, Nagano et al., 2017; Flyamer et al., 2017; Stadler et al., 2017; Bylino et al., 2021; Ulianov et al., 2021), a tethered ligation protocol (Kalhor et al., 2012; Eagen et al., 2015, Eagen et al., 2017; Gabdank et al., 2016), or a dilution ligation protocol (Dekker et al., 2002; Tolhuis et al., 2002; Lieberman-Aiden et al., 2009; Comet et al., 2011; Stadhouders et al., 2013; Ulianov et al., 2016; El-Sharnouby et al., 2017; Vermeulen et al., 2020). Therefore, it is of primary interest, first, to carefully consider the initial stages of the 3C procedure, including extraction of larvae from fly food, sorting them by age, and preparing a cell material from them.

Here, we described for the first time a detailed procedure of the 3C experiment with whole *Drosophila* larvae. Important initial steps of handling larvae, selecting the developmental stages, and preparing a cell material, were thoroughly considered. Next, the subsequent stages of preparation of experimental 3C and control

BAC libraries, preparation of a calibration curve, and analysis of interactions are carefully described. To validate the results, electrophoretic pictures of the resulting 3C libraries are shown and raw qPCR data are provided to demonstrate that the libraries are well-amplified. Finally, to prove the efficiency of our procedure, we provided the experimental results obtained to determine the interactions between regulatory elements of the developmental locus *Dad* in S2 cells and wild-type (WT) and transgenic larvae. Our work fills the gap in the application of the 3C method to *Drosophila* larvae and provides a useful guide for establishing 3C on an animal model.

2 RESULTS

2.1 Collection, Sorting, and Homogenization of *Drosophila* Larvae and Preparation of a Larval Cell Material

2.1.1 Extraction of Total Larvae From Fly Food

Individual collection of late 3rd instar larvae from the walls of fly cultivation tubes is laborious and requires a large number of tubes with flies of the same developmental stage. It is more rational to extract all larvae from food with the help of 20% (585 mM) or 0.8 M sucrose. Upon adding 20% sucrose, the larvae float up with the liquid part of the food and can be collected individually from the surface after stirring the contents with a spatula with a groove to facilitate the process.

If a large amount of larvae is required, a system of three sieves of different sizes can be used (**Figure 1A**). The sieves make it possible to separate larvae of different stages and, most importantly, to isolate the largest larvae of the 3rd instar, which precedes the adult stage. These larvae are the principal focus of this study. A 125-μm sieve retains 1st instar larvae, embryos, and water. An 800-μm sieve retains raisins, pieces of agarized food, dead adult flies, and the largest 3rd instar larvae. A 315-μm sieve retains early 3rd instar larvae and 2nd instar larvae.

2.1.2 Mass collection of 3rd and 2nd Instar Larvae for Homogenization

If it is not necessary to separate the 3rd and 2nd instar larvae, larvae can be picked up from a 315 μm sieve using a spatula with a groove into a little glass with 20% sucrose. Then the larvae are transferred from the glass into a 100-μm nylon cell strainer (Corning Falcon, cat. no. 352360) and washed with the EW buffer. After washing, the entire mass of the 3rd and 2nd instar larvae are transferred using a paintbrush to a Dounce homogenizer of an appropriate volume and homogenized with pestle A to produce a cell material.

A disadvantage of this approach is that some amount of semolina or corn meal particles (depending on what cereal is used to prepare fly food) is collected together with the larvae when collection is carried out with a spatula with a groove from a 315-μm sieve. The particles sediment with cell material after homogenization to produce a uniform mixture of fixed cells and cereal particles. Experiments end in failure when performed with such a mixture. We concluded that it is necessary to separate fixed
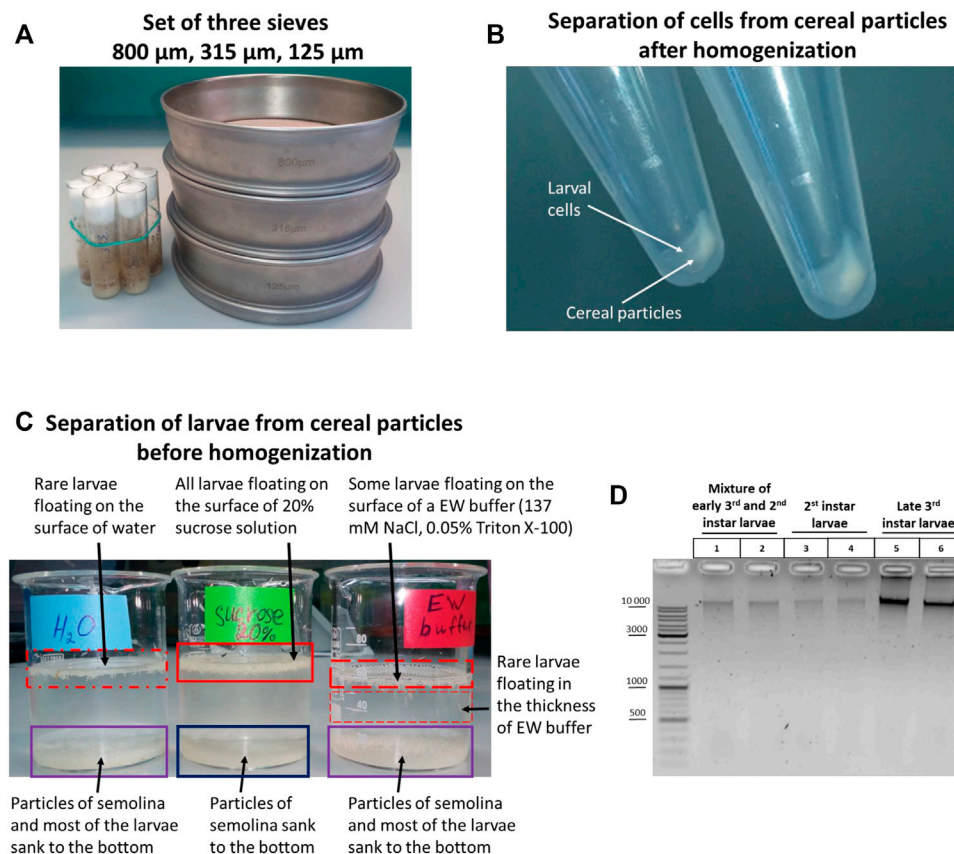
FIGURE 1 | Initial stages of the 3C protocol: collection, sorting, and separation of larvae from fly food before and after homogenization. **(A)** TEST SIEVE Retsch GmbH 200 mm × 50 mm, 800 μm, 315 μm, 125 μm in comparison with standard vials for *Drosophila* cultivation. **(B)** Separation of larval cell material from cereal particles after homogenization. Total larvae were extracted from fly food as described in Section 2.1.1. Then the sucrose solution containing the liquid part of the food with the larvae was poured from fly vials into a set of 3 sieves depicted in **Figure 1A**, and the larvae were processed to obtain a cell material as described in *Processing of larvae extracted from fly food and preparation of cellular material* in *Materials and Methods*. After dissolving the cell material in ice-cold 1X PBS, the larval cell material was either i) applied onto a discontinuous sucrose gradient (0.8, 1.6, and 2.3 M sucrose steps) and centrifuged at 1,000 g at +4°C or ii) centrifuged through a 20% (0.584 M)/ 0.8 M sucrose cushion at 5,000 g. **(C)** Separation of larvae from fly food before homogenization. Larvae were extracted from fly food, as shown in **(B)**, and collected as described in Section 2.1.2. Then the larvae were transferred with a paintbrush from a 100-μm cell strainer into 3 different solutions: water, 20% sucrose, and EW buffer. The larvae floating on the surface of a liquid or deep in the solution are marked with a red rectangle; the density of the rectangle shading symbolizes the number of larvae. A blue rectangle shows the particles of cereal that sunk to the bottom. A purple rectangle shows a mixture of larvae and particles of cereal. In this experiment, the separation of larvae and cereal particles was performed by letting the solutions with larvae and cereal particles stand at RT for 30 min. **(D)** Chromatin integrity control obtained from WT Canton S larvae at different stages of development. Larvae were extracted from fly food, as shown in **(B)**. Then 3rd instar larvae were individually collected and washed as described in *Individual collection of 3rd instar larvae for homogenization*. A mixture of early 3rd and 2nd instar larvae was collected, as shown in **(C)**. The 2nd instar larvae were collected individually with tweezers from the mixture of 3rd and 2nd instar larvae from a 315-μm sieve. After that, the larvae were processed, as shown in **(B)**. Larval cell material (25 mg) was taken and centrifuged, the supernatant was removed, and the 3C protocol was performed until obtaining the chromatin integrity control (see *Protocol of the 3C experiment with Drosophila larvae* in the **Supplementary Material**). 1/10 of the sample volume (2.5 mg) was taken as a control, and the control was processed according to the 3C protocol except that, to isolate DNA, 500 μl of an extraction buffer (EB) (see *Materials and Methods*) containing 30 mM EDTA and 0.2 mg/ml Proteinase K was added in each tube instead of 1X T4 DNA ligase buffer and the DNA was dissolved in 25 μl of 10 mM Tris-HCl, pH 8.0. Five μl of the DNA preparation was examined by electrophoresis. Two biological replicates were performed for each type of larvae.

cell material from cereal particles after homogenization or, alternatively, to cleanly separate larvae from fly food before homogenization. Accordingly, we tried first to separate cell material from cereal particles after homogenization and found at the post-homogenization step that this method is inefficient, be it performed with a sucrose cushion (not shown) or a discontinuous sucrose gradient (**Figure 1B**); that is, the larval cell material and homogenized cereal particles co-sedimented simultaneously in both cases. Thus, it is necessary to separate larvae from fly food at an earlier stage before homogenization.

Next, we tried to separate larvae from fly food before homogenization and observed that when a mixture of larvae and cereal particles was left to stand for 20–30 min in a glass with 20% sucrose, but not the EW buffer or water, the cereal particles sank to the bottom, while live larvae float on the surface of 20% sucrose (**Figure 1C**). Then the larvae can be collected from the surface of 20% sucrose with a spatula with a groove in a 100-μm nylon cell strainer and washed with EW before homogenizing. In order to speed up the separation process, the larvae in 20% sucrose can be transferred from a glass into 50-ml tubes and centrifuged at 2,500 g

for 1 min. In this case, the larvae slightly sink into the thickness of 20% sucrose and quickly float back to the surface after centrifugation, while the precipitated cereal particles do not. We concluded that the approach of separating larvae and cereal particles before the stage of larva homogenization proved to be efficient. Another option that would help to separate larvae from cereal particles is using fly food cooked without semolina or corn meal and, preferably, without raisins (the addition of cereals and raisins to fly food does not affect the total number of flies produced nor their rate of development, unpublished data).

2.1.3 Individual collection of 3rd Instar Larvae for Homogenization

In the case where only 3rd instar larvae are required, but not a mixture of 2nd and 3rd instar larvae, 3rd instar larvae are individually collected with tweezers from an 800-µm sieve in a 100-ml beaker glass with 20% sucrose. The solution with larvae was poured into a 100-µm cell strainer, and the larvae were washed with EW buffer. Approximately 100 larvae or a little more (no more than 150) were necessary to collect. This number of larvae is easy to homogenize in a 7-ml Dounce homogenizer with pestle A, which disrupts tissues into individual cells. Filtration through a 40-µm cell strainer yields ~50 mg of cell material, which can serve to make two replicates of 25 mg each. We observed that 200 or more larvae were hardly homogenized in a 7-ml Dounce homogenizer (there was risk of breaking the homogenizer or pestle), and a 15-ml Dounce homogenizer should be used with this number of larvae.

As shown in **Figure 1D**, 50 mg of cell material obtained from 2nd instar larvae contained a lower amount of DNA than 50 mg of cell material obtained from 3rd instar larvae; a mixture of 2nd and 3rd instar larvae contained an intermediate amount of DNA. Since larval growth and an increase in cell size are achieved primarily *via* endoreplication in *Drosophila* larvae, that is, cell growth and DNA replication occur in the absence of cell division, most larval tissues are composed mainly of polyploid cells (Edgar and Orr-Weaver, 2001; Edgar and Nijhout, 2004; Lee et al., 2009; Zielke et al., 2013; Ren et al., 2020). Therefore, a higher degree of cell polytenization in 3rd instar larvae is apparently responsible for the extraction of a greater DNA amount from the same amount of cell material. We used only 3rd instar larvae in subsequent experiments.

Alternatively, embryos can be grown in large jars of a large volume, for example, 200 ml, and 3rd instar larvae that have crawled onto the walls can be collected with a paintbrush or tweezers or rinsed off the walls with water. Extraction with sucrose will be especially convenient for those who want to get larvae of earlier instars and separate them into size fractions by age.

2.2 Processing of Larvae Extracted From Fly Food and Preparation of Cell Material

This chapter is given in the beginning of the Materials and Methods section (see **Supplementary Material**).

Two important methodological observations were made while preparing 3C libraries from wild-type (Canton S) and mutant (transgenic) larvae:

(i) The DNA yield from 10 mg of WT larval cell material pellet was approximately 10–15 times lower than from the same quantity of S2 cells. Although most *Drosophila* larval tissues are composed mainly of polyploid cells (Edgar and Orr-Weaver, 2001; Edgar and Nijhout, 2004; Lee et al., 2009; Zielke et al., 2013; Ren et al., 2020), we suppose that the pellet that forms after filtration through a 40-µm cell strainer and that is analyzed may contain a sufficiently large number of cells of the imaginal disc and histoblast nest (mitotic cells), diploid in nature, as well as extracellular material, such as remnants of the milled cuticle. This may be one of the reasons that cell material obtained from larvae contains less DNA than the same quantity of S2 cells. Another reason is that S2 cells have an altered karyotype and their ploidy varies from 1 to more than 8, with an average karyotype of 2X; 4A corresponding to tetraploids (Zhang et al., 2010; Lee et al., 2015). Therefore, it is very likely that a tetraploid cell culture of the same quantity contains more DNA than milled larval tissues.

(ii) The amount of DNA isolated from the same quantity of cell material obtained from mutant lines carrying transgenic insertions significantly varied from line to line and was generally reduced (at least by a factor of 2–3) as compared with the WT Canton S strain. This may be due to the poor genetic background in the mutant lines or impaired expression of the *Dad* gene. *Dad* expression is required for inhibiting the BMP/Dpp/TGF-β signal transduction pathway (Tsuneizumi et al., 1997; Marquez et al., 2001; Sharifkhodaei and Auld, 2021). Direct suppression of endomitosis and endoreplication has been found to occur upon stimulation of the BMP/Dpp/TGF-β pathway (Kuter et al., 1992), although an opposite situation is observed, for example, in nematodes, where the BMP/Dpp/TGF-β pathway stimulates endoreplication (Nystrom et al., 2002; Lozano et al., 2006). Nevertheless, these data taken together implicate the BMP/Dpp/TGF-β pathway in controlling endoreplication

2.3 Processing of the Cell Material Prepared From *Drosophila* Larvae

Taking into account the abovementioned observations, the amount of starting cell material to prepare a 3C library should be at least 25 mg in the case of WT larvae and at least 50 mg in the case of mutant larvae. It is necessary to check the DNA yield for each particular mutant line since the yield may be significantly lower than that for WT larvae. A comprehensive protocol of the 3C experiment with *Drosophila* larvae, from the processing of cell material to obtain a purified 3C library to statistical analysis of the 3C experiment results, is presented in the **Supplementary Material**. The part devoted to processing cell material covers steps 1 through 25 of the protocol and comprises the following sections:

  I. Cell lysis
 II. Nucleoplasm release and chromatin treatment with heat
III. Digestion of DNA in nuclei

IV. Ligation of DNA in nuclei

V. Reversion of cross-links and isolation of a 3C library

VI. Treatment of the 3C library with RNase

VII. Purification of the 3C library on magnetic beads and DNA analysis

The stages of reversion of cross-links and isolation of a 3C library, treatment of 3C library DNA with RNase, and purification of the 3C library on magnetic beads and DNA analysis can be used as independent protocols and have been thoroughly discussed by Bylino et al. (2021).

2.4 Preparation of a Control Library (Random Ligation Library) for Constructing a Calibration Curve

The next step in the 3C procedure is to obtain a control library consisting of Sau3A (BspI)-digested and then randomly ligated DNA fragments of BAC(s) overlapping the locus (loci) of interest. If the locus contains transgenic sequences, the BAC can be mixed in an equimolar ratio with a plasmid carrying such sequences. A control library is necessary for constructing a calibration curve, which is used to calculate the relative ligation frequencies (RLFs) of the fragments of interest in 3C samples. We studied the conditions of BAC cultivation, isolation, purification, restriction digestion, and random ligation of the resulting restriction fragments. These stages are described in detail in section 2 of the **Supplementary Material** (subsections 2.1, 2.2, 2.3, and 2.4). This section can be used as an independent protocol to grow and purify BAC and to prepare a random ligation library. The part devoted to the preparation of a random ligation BAC library covers steps 26 through 34 of the protocol in the **Supplementary Material** and comprises the following sections:

VIII. Preparation of a random ligation BAC library for constructing a calibration curve

The main conclusions from our experiments and optimization of the above stages of the protocol are as follows:

(i) Induction of BAC replication (CHORI321 library pCC1BAC series) occurs in the presence of L-arabinose in the culture medium but not in the presence of inactivated tetracycline hydrochloride or chlortetracycline.

(ii) In the absence of a replication inducer, the BAC behaves as an ordinary multicopy plasmid despite the presence of a single-copy origin of the F-factor, increasing in copy number in the presence of chloramphenicol (Cm). The BAC copy number rapidly decreases when Cm is absent or consumed in the medium, and the BAC yield becomes very low.

(iii) The BAC can be efficiently isolated with a usual plasmid isolation kit.

(iv) Purification of the BAC on AMPure XP beads appears to be more efficient than standard EtOH precipitation (without the addition of glycogen).

(v) Equimolar mixing of the BAC and the plasmid leads to an equally represented ligation of the transgene and the BAC fragments in the resulting control library.

2.5 Construction of a Calibration Curve and Analysis of the 3C Library

The next step of the 3C procedure is analyzing the ligation frequencies in the experimental 3C library (described in detail for S2 cells by Bylino et al. (2021) and for larvae in the **Supplementary Material**). Measurement of the number of ligation events between spatially adjacent DNA regions in genome-wide 3C (Hi-C) (Lieberman-Aiden et al., 2009; Rao et al., 2014; Belaghzal et al., 2017; Akgol Oksuz et al., 2021; Lafontaine et al., 2021) or in 3C covering a few (Capture-C) (Hughes et al., 2014; Davies et al., 2016; Golov et al., 2020a; Hua et al., 2021) or more (promoter capture Hi-C) (Mifsud et al., 2015; Schoenfelder et al., 2018a) regions of interest is based on the results of sequencing on the Illumina platform (Davies et al., 2017; Grob and Cavalli, 2018). Its prohibitive cost means that such datasets are reasonable to obtain when a large number of interactions is to be analyzed or when interaction partners are unknown. However, when a limited number of interactions between several DNA regions is of interest, the ligation frequency can be determined using real-time qPCR. To achieve a greater sensitivity of the method, TaqMan probes are used instead of SYBR Green–based detection. A qPCR analysis of the ligation frequencies utilizes a calibration curve, which is obtained from 10X dilutions of a random ligation mixture of control DNA digested with the same restriction endonuclease (RE) or a RE that has the same cleavage site. The calibration curve can be quantitative, that is, based on purified PCR products, which are used in known concentrations, overlap the restriction sites of interest, and are mixed in equimolar amounts before random ligation, or semi-quantitative, that is, based on random ligation of a digested BAC that covers the region of interest (relative ligation frequencies) (Gavrilov et al., 2013). When a transgenic construct is present in a region of interest, a plasmid(s) containing the cloned region(s) or PCR amplicons of transgenic regions can be combined with the BAC in equimolar amounts before preparing a random ligation library mix (Shidlovskii et al., 2021).

Before analyzing the ligation frequencies between restriction fragments of interest in experimental 3C samples, several important preliminary experiments should be performed: 1) to optimize the conditions of qPCR with TaqMan probes, which will be used to determine the ligation frequencies in experimental 3C libraries (optional); 2) to study the ligation frequencies in random ligation libraries (optional); 3) to determine the linear range of amplification by testing several dilutions of a random ligation library; and 4) to establish the amount of 3C DNA libraries required to fit the calibration curve (in our case, this was done with a model of S2 cells and then with a model of WT larvae). These stages are described in detail in section 3 of the **Supplementary Material** (subsections 3.1, 3.2, 3.3, and 3.4). The part devoted to the preparation and preliminary testing of

the calibration curve covers steps 26 through 35 of the protocol and comprises the following sections:

IX. Preparation of the calibration curve based on a random ligation library

X. Preliminary testing of dilutions of 3C samples against the calibration curve

The main conclusions that can be drawn from our experiments and optimization of the above stages of the protocol are as follows:

(i) The detection threshold of the PCR product in the reaction increases in direct proportion to the amount of the TaqMan probe added to the reaction. The greater the probe amount, the earlier occurs the threshold of 3C product detection (an increase in probe concentration from 0.05 to 1 pM/µl leads to the appearance of a signal 4.5 ± 0.5 Ct earlier). Therefore, if the DNA content in the 3C library is low and a limited amount of the starting material is available for preparing the 3C library, then increasing the concentration of the TaqMan probe in the reaction up to 1 pM/µl makes it possible to detect the product without generating additional amounts of the 3C library.

(ii) When a plasmid containing the transgene and a BAC containing a genomic locus of interest are combined in equimolar proportion, a mixture of restriction fragments is obtained, in which the ligation products of different regions are equally represented.

(iii) A calibration curve prepared in a range from 1 ng/µl to 100 fg/µl from 10X dilutions of a random ligation library as described in **section 4** and in the protocol for larvae in the **Supplementary Material** (step 35) can be routinely used to determine the cross-linking frequency in an experimental 3C library. Each dilution is used in 4 technical replicates in an amount of 5 µl per PCR mixture (then the template DNA concentrations will be 5 ng–500 pg–50 pg–5 pg–500 fg per PCR mixture, respectively). A linear amplification region corresponds to a 10X dilution range from 10 ng/µl to 100 fg/µl. The step between two calibration dilutions is 3.5–4.0 Ct. It is convenient to select 4 dilutions to be used in the experiment in parallel with the 3C samples.

(iv) 3C libraries prepared from 10 mg of S2 cells according to the protocol described by Bylino et al. (2021) are possible to dilute by a factor of 10–15 (to a concentration of 1.24–6.23 ng/µl) without impairing the reliability of contact determination. With such a concentration of 3C libraries in the PCR mixture, the calibration curve is built from dilutions of 100 pg–10 pg–1 pg–100 fg per µl (the product appears between the 10 pg/µl and 1 pg/µl dilutions of the calibration curve). The quantity of S2 cells required for the preparation of a 3C library is 10 mg and may be diminished even to 5 mg if necessary. A twofold increase in the DNA concentration of the 3C library in PCR gives an increase of 1.75–2.0 Ct for S2 cells and for WT larvae.

(v) At least 5–10 ng of the 3C library per PCR mixture prepared from WT larvae is sufficient for detecting the ligation

products between DNA regions of interest at 100 pg–10 pg–1 pg–100 fg dilutions of the calibration curve per µl (the product appears between the 1 pg/µl and 100 fg/µl dilutions). For additional depth of the calibration curve in the case of a limited amount of 3C library DNA from larvae, a 10 fg/µl dilution can be introduced into the range of the calibration curve. The dilution is set in 8, rather than 4, technical replicates. However, the dilution does not always fall within the linear range of amplification.

(vi) At least 12–46 ng of 3C library DNA prepared from mutant larvae is necessary to take in a PCR mixture for reliable detection of ligation products (the product appears between the 1 pg/µl and100 fg/µl or between the 10 and 1 pg/µl calibration curve dilutions, depending on the mutant line and the amount of DNA taken into the PCR mixture).

X. The stages of preparation of the qPCR master mix and the general arrangement of a qPCR experiment with 3C samples, normalization of ligation frequencies, calculation, and statistical analysis of the 3C experiment results are described in detail in section 4 of the **Supplementary Material** (subsections 4.1, 4.2, and 4.3). The part devoted to the preparation and preliminary testing of the calibration curve covers steps 36 through 39 of the protocol in the **Supplementary Material** and comprises the following sections of the protocol: Preparation of the PCR master mix and the general arrangement of a qPCR experiment with 3C samples.

XI. Normalization of ligation frequencies

XII. Calculation and statistical analysis of the 3C experimental results

2.6 Critical Analysis of the Existing 3C Protocols for Whole *Drosophila* Larvae. Optimal Parameters, Peculiarities, and Advice on Preparing and Processing Cell Material Obtained From Larvae

Although a sort of universal 3C protocol has been previously published to describe the processing of collected adult flies, pupae, and embryos and to prepare cell material from them (Comet et al., 2011), no detailed procedure has been reported yet to allow 3C investigations in *Drosophila* whole larvae. Prior to our study, only one study was published to describe 3C with whole *Drosophila* larvae (Bieli et al., 2015). A critical analysis of the very brief procedure described by Bieli et al. (2015) revealed several issues that would require significant changes. Based on our own experience with larva processing and treatment of the cell material obtained from them and from S2 cells (Bylino et al., 2021; Shidlovskii et al., 2021), we suggest the following important improvements to the previously published procedure:

(i) Fix larvae simultaneously with homogenization in a Dounce homogenizer with FA used at a concentration not exceeding 0.5% for 10 min (we found that fixation with 1.8% FA for 20 min is excessive). We observed that tissues of larvae are over-fixed even when fixation is

performed with 0.5% FA for 25 min and that DNA of over-fixed cells is poorly cut by a restriction endonuclease (RE) and is not extracted with Ph/Chl, remaining in the non-lysed cells (Bylino et al., 2021).

(ii) Quench FA with glycine used at an equimolar concentration or in a slight excess to FA (Comet et al., 2011; Sexton et al., 2012), keeping in mind two reactive groups of FA vs. one group of glycine. For example, we used glycine at 666 (equimolar) and 800 (slight excess) mM for 1% FA (666 mM molarity by the number of reactive groups). Do not quench with a strong molar deficiency of glycine to FA (with 0.125 M glycine) since quenching is not likely to be complete in these conditions (Splinter et al., 2012) or, if doing so, immediately proceed to the next step after quenching without storing the material (even after flash freezing since cell fixation may still proceed during thawing).

(iii) Do not store fixed larval cell material inactivated with glycine and washed with ice-cold 1X PBS on ice for several days, for example, until controls #1 (chromatin integrity) and #2 (chromatin restriction digestion) are ready to understand if chromatin is not degraded and well digested. Fixation followed by keeping the cell material on ice does not allow maintaining the DNA integrity. Degraded DNA was isolated from the cell material and formed a smear, which disappeared after treatment with bovine RNase A, which digests DNA, especially in a degraded form, according to our results (discussed by Bylino et al. (2021)). Instead, proceed to the cell lysis stage immediately without storing the cell material.

(iv) Do not use a too high RCF (10,000 g and higher), but centrifuge larvae and cells at an RCF not exceeding 5,000–7,500 g. It has been shown that centrifugation at more than 8,000 g leads to broken, sheared nuclei (Louwers et al., 2009).

(v) Do not use mechanical force when treating the nuclei, and handle the nuclei gently. For example, do not pass the nuclei through a syringe needle because this can affect the integrity of the nuclei and DNA. Instead, use a 40-μm cell strainer to filter the homogenate to obtain a cellular material (Bylino et al., 2021).

(vi) Use chromatin treatment modes at 65°C (65°C with SDS for 5–10 min and 37°C with Triton X-100 for 15 min) as it is preferable for larval cells. We observed that the regimens of chromatin treatment with SDS at 37°C (37°C with SDS for 10 min and Triton X-100 for 15 min or 37°C with SDS for 1 h and Triton X-100 for 1 h), which provide for more efficient ligation in the case of S2 cells (Bylino et al., 2021), did not give the same improvement in the case of larval cell material.

(vii) After chromatin heat treatment at 37/65°C in the presence of SDS/Triton X-100, wash the nuclei with 1X restriction buffer (RB) before restriction digestion (Flyamer et al., 2017) since, even sequestered with Triton X-100, SDS is able to hamper the RE function at high concentrations (Louwers et al., 2009) and only a few REs can tolerate the conditions of 0.3% SDS

sequestrated with 1.8–2% Triton X-100 (Splinter et al., 2012; van de Werken et al., 2012).

(viii) Omit the step of inactivation of RE at 65°C for 10 min in the presence of high SDS concentrations (~1.2–1.3%) as it negatively affects the structure of nuclei (Nagano et al., 2015b). Instead, wash off the nuclei from RE with 1X T4 DNA ligase buffer before DNA ligation (Nagano et al., 2013; Nagano et al., 2015a; Nagano et al., 2015b, Nagano et al., 2017).

(ix) Do not ligate chromatin in the presence of 0.1% SDS even sequestered with 1% Triton X-100. We observed that 0.1% SDS even sequestered with 1% Triton X-100 gives strong inhibition during in-nucleus DNA ligation (Bylino et al., 2021), although ligation in a solution of plasmid DNA cleaved by 6 bp cutter in the presence of 0.1% SDS sequestered with 1% Triton X-100 appear to be efficient (Gavrilov et al., 2013). It has previously been proposed to wash the nuclei before DNA ligation (Flyamer et al., 2017; Golov et al., 2020b). We examined this important issue and found that at least a triple washing of larvae nuclei suspension with 1X T4 ligase buffer at this stage efficiently prevents inhibition and does not lead to DNA degradation (Bylino et al., 2021).

(x) After Ph/Chl extraction, process 3C library preparations with RNase I and then purify them on AMPure XP paramagnetic beads (SPRI technology) in order to remove RNase and to additionally purify the libraries before PCR determination of ligation frequencies. We observed that RNase I is efficient in removing RNA impurities from DNA preparations and the «safest» RNase for DNA treatment (Bylino et al., 2021). Use a T4 DNA ligase concentration of at least 0.025 WeissU/μl. We determined that the T4 DNA ligase concentration of 0.0025 WeissU/μl is inefficient during in-nucleus ligation and that an efficient concentration range is from 0.025 to 0.25 WeissU/μl (Bylino et al., 2021).

The abovementioned steps were tested in our previous studies (Bylino et al., 2021; Shidlovskii et al., 2021) and in this work and were found to ensure the preservation of DNA integrity at all stages. Our improvements allow the stable detection of distant DNA site interactions in 3C libraries prepared from at least 25 (WT larvae) or 50 (transgenic mutant larvae) mg of starter material. For a more detailed discussion of the steps described earlier, see the main part and the supplementary in the study by Bylino et al. (2021).

2.7 Analysis of Distal Interactions in the Developmental Locus *Dad* in *Drosophila* Larvae

2.7.1 Detection of the *Dad* Enhancer Interactions in *Drosophila* Wild-type Larvae. Comparison of the 3C Profiles in Wild-type Larvae and S2 Cells

To validate our optimizations of the 3C procedure for *Drosophila* larvae, we chose the developmental locus *Dad* (daughters against
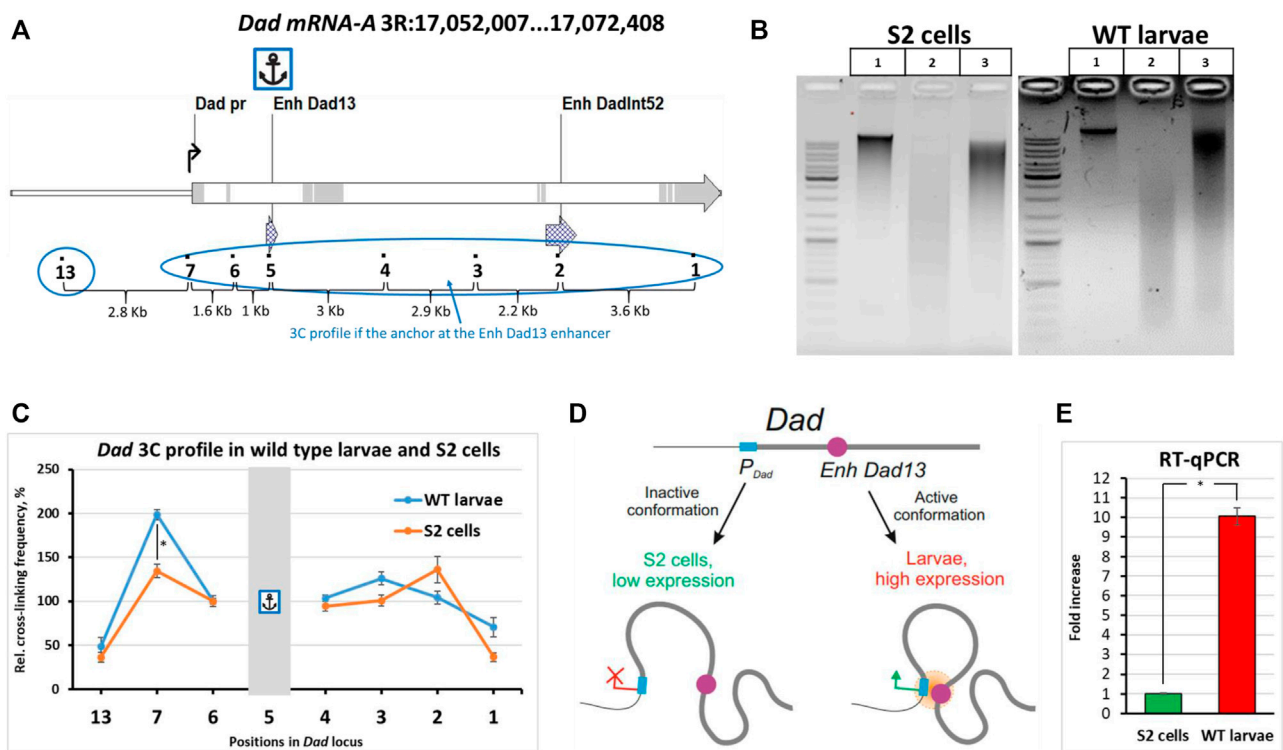
FIGURE 2 | Comparison of the 3C profiles of live wild-type *Drosophila* larvae and S2 cells. **(A)** The model locus *Dad* with an anchor on the enhancer *Dad13* in WT *Drosophila* larvae and S2 cells is shown. Designations are as shown in **Supplementary Figures S3A**. Exons are highlighted grey; introns are shown white. **(B)** A representative example of the 3C libraries prepared from WT larvae and S2 cells. Larvae were extracted from fly food, as shown in **Figure 1B**. One hundred WT larvae were individually collected and washed, as shown in **Figure 1D**, and then processed to obtain a cell material, as shown in **Figure 1B**. Larval cell material (25 mg) was taken for 3C library preparation. The cell material was centrifuged, the supernatant was removed, and a 3C library and controls #1 and 2 were prepared as described in *Protocol of the 3C experiment with Drosophila larvae* in the **Supplementary Material**. The 3C libraries and controls of S2 cells were prepared as shown in **Supplementary Figure S6** except that control #1 and the ligation mixture were purified using a 1.5X volume of AMPure XP beads and control #2 was purified using a 2X volume of AMPure XP beads. For electrophoresis, in the case of S2 cells, 200 ng of DNA was used for controls #1,2 and the ligation mixture. In the case of larvae, the DNA amount resolved in the gel was 25 ng for control #1, 50 ng for control #2, and 75 ng for the ligation mixture. Lane 1, control #1 (chromatin integrity control); lane 2, control #2 (chromatin digestion control); lane 3, ligation mixture (3C sample). **(C)** Comparison of the 3C profile between live WT Canton S larvae and S2 embryonic cells. Six independent biological replicates of the 3C library were analyzed for larvae and for S2 cells. The concentrations of all 3C libraries were made equal according to Qubit readings before measurements. The frequency of ligation of the anchor fragment with the adjacent fragment (point #6) was arbitrarily taken as 100%, and the values of all other experimental points were calculated proportionally. The relative ligation frequencies (RLFs) of experimental samples were normalized to RLFs within the constitutively expressed *RpII* locus. Error bars indicate SEMs from six independent biological replicates of the 3C library. Each experimental point for each 3C library was studied in 4 technical PCR replicates, and the data were averaged. One-tailed Student's t-test was used for comparison between groups to calculate the reliability of the revealed differences. An asterisk indicates the significance level: $*p < 0.05$, n = 12. **(D)** A model illustrating spatial differences in the location of regulatory elements of the developmental *Dad* gene in cultured cells and live *Drosophila* larvae. **(E)** Evaluation of the expression level of the *Dad* gene in S2 cells and wild-type larvae. Total RNA was prepared, and the mRNA level was measured as described in the Isolation of total RNA and RT-qPCR section of Materials and Methods. For reverse transcription, the same amount of RNA was taken from S2 cells and wild-type Canton S larvae. The *Dad* gene expression level was normalized to the *β-tubulin 56D* gene expression level. The two genes showed similar amplification efficiencies. The *Dad* mRNA content was calculated using the ΔΔCt method. In addition to the ΔΔCt, a calibration curve prepared from S2 cell DNA as described in Materials and Methods was used in the experiment. The results obtained using the calibration curve and ΔΔCt were similar. The *Dad* expression level in S2 cells was arbitrarily taken as 1. Three independent total RNAs were prepared from S2 cells and from larvae, and each RT was studied in 4 PCR technical replicates. Error bars indicate SDs of 4 PCR technical measurements from three independent biological replicates of total RNA. One-tailed Student's t-test was used to evaluate the reliability of differences in between-group comparisons. An asterisk indicates the significance level: $*p < 0.001$, n = 6.

decapentaplegic) as a model (**Figure 2A**). The *Dad* gene is a regulatory response gene in the BMP/Dpp/TGF-β pathway and encodes the receptor-binding protein important for negative feedback in the transmission of the signal from receptors activated by the Dpp ligand (Tsuneizumi et al., 1997; Marquez et al., 2001; Sharifkhodaei and Auld, 2021). The BMP/Dpp/TGF-β signal transduction pathway is important for the growth, proliferation, and differentiation of cells of imaginal discs

(Peterson and O'Connor, 2014; Upadhyay et al., 2017). The *Dad* gene has two enhancers, the proximal *Dad13* enhancer (the main enhancer) and the distal *DadInt52* enhancer (shadow enhancer) (Weiss et al., 2010; Neal et al., 2019). Both enhancers are located in the introns of the gene (**Figure 2A**). We studied the interactions of the enhancer *Dad13* with the regions inside the *Dad* locus in WT larvae and compared the resulting 3C profile with the profile obtained for S2 embryonic cells.
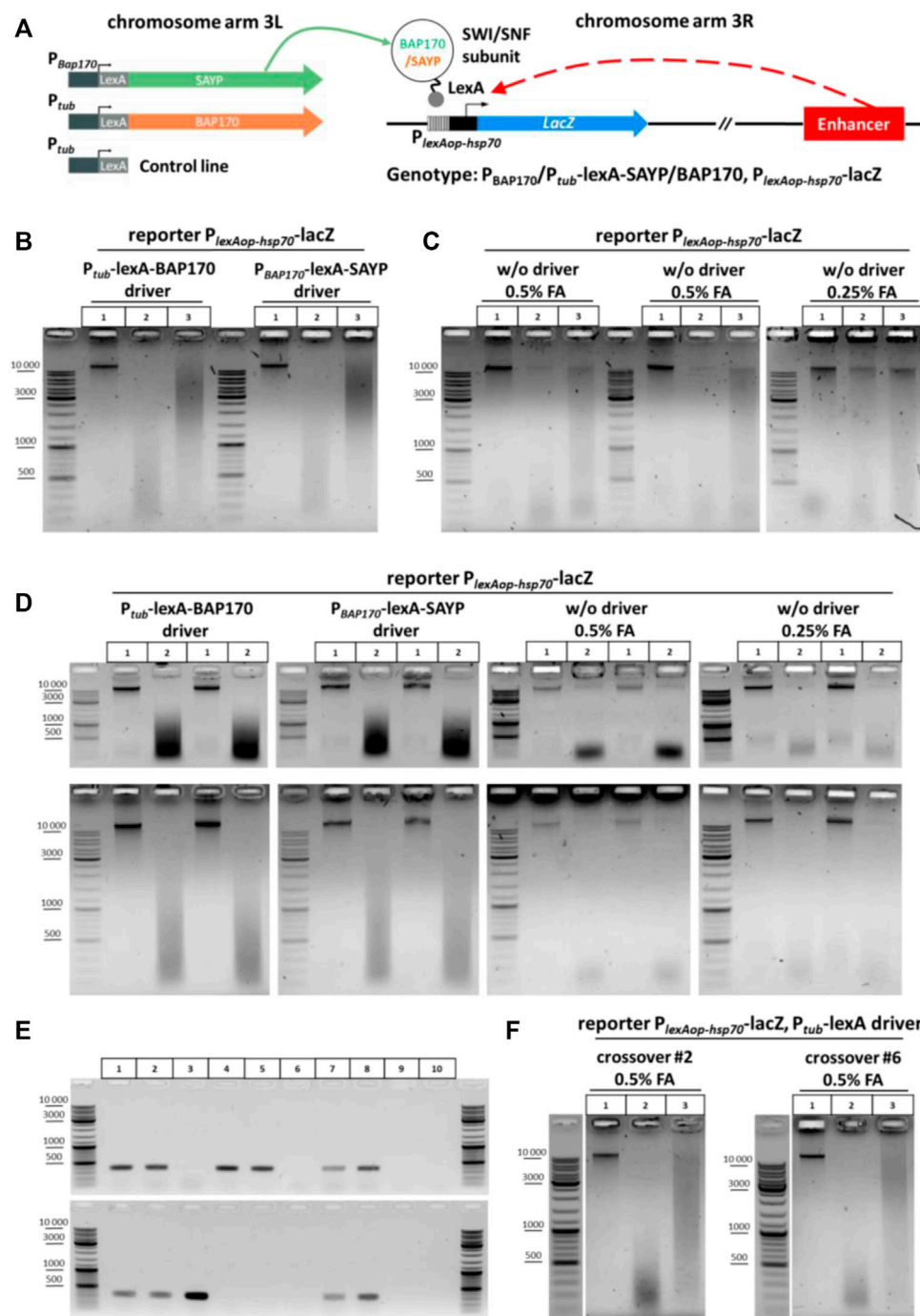
FIGURE 3 | General scheme of the experiment with transgenic larvae, screening of crossovers, and 3C libraries obtained from the transgenic larvae. **(A)** Genetic structure of transgenic flies and a general arrangement of the experiment with transgenic larvae used to determine the role of SAYP or the BAP170 Brahma/SWI/SNF subunits in enhancer–promoter interactions. The SWI/SNF subunits BAP170 and SAYP tethered to the reporter promoter induce activation of the reporter in an adjacent enhancer-dependent manner. **(B)** 3C libraries prepared from larvae carrying the reporter $P_{lexAop-hsp70}$-lacZ transgene in the presence of either P_{tub}-lexA-BAP170 or P_{BAP170}-lexA-SAYP driver. The combinations of transgenes are indicated above the electrophoretic pictures. The experiment was performed as shown in **Figure 2B** for larvae except that 150 larvae were collected and 50 mg of larval cell material was taken for 3C library preparation. Lane 1, chromatin integrity control; lane 2, chromatin digestion control; lane 3, ligation mixture (3C sample). At least seven independent 3C libraries were prepared for each genotype. A representative example is shown for two 3C libraries. **(C)** 3C libraries prepared from larvae carrying the $P_{lexAop-hsp70}$-lacZ reporter only. 3C libraries and controls were prepared as shown in **(A)**. Lane 1, chromatin integrity control; lane 2, chromatin digestion control; lane 3, ligation mixture (3C sample). At least seven independent 3C libraries were prepared. Three biological replicates obtained after fixation with different FA concentrations are presented. **(D)** Electrophoretic analysis of chromatin integrity controls and restriction digestion controls prepared from larvae carrying the $P_{lexAop-hsp70}$-lacZ reporter with or without the P_{tub}-lexA-BAP170 or P_{BAP170}-lexA-SAYP driver. The controls were

(*Continued*)

Individually collected 3rd instar Canton S larvae were used. 3C libraries were prepared and analyzed as described earlier. Electrophoretic analysis of the 3C libraries is given in **Figure 2B**.

When we addressed the 3C profile, we found that the profile of *Dad13* enhancer interactions in larvae differed from that obtained for cultured cells. Although the upstream and downstream regions interacted with the enhancer weakly in both larvae and S2 cells, the promoter region was significantly closer to the enhancer in larvae than in S2 cells (**Figure 2C**, **Supplementary Figures S10 and S11**). The situation did not differ when a smaller amount of larval cell material was used to prepare the 3C library (**Supplementary Figure S12C**). This means that our data fall within the linear sensitivity range of the 3C method. Thus, our data suggest that there are differences in the proximity of the enhancer to the promoter between WT larval cells and cultured embryonic S2 cells (**Figure 2D**). The differences may be associated with the *Dad* expression level. Indeed, *Dad* is expressed at a low level in cultured cell lines and is active in 3rd instar larvae, according to databases (Chintapalli et al., 2007; modENCODE Consortium et al., 2010; Larkin et al., 2021). To check this experimentally, we performed RT-qPCR using equal amounts of S2 and WT 3rd instar larval cell material and found that the *Dad* expression level in cultured S2 cells was approximately 10 folds lower than in larval cells. Thus, the differences in the enhancer–promoter proximity within the *Dad* locus correlate with the gene activity. The inactive state of the *Dad* locus in S2 cells can be accounted for by its localization in the B compartment in S2 cells and embryos (Sexton et al., 2012; Li et al., 2015; Cubeñas-Potts et al., 2017; Eagen et al., 2017).

2.7.2 3C Profile of the *Dad* Locus in the Presence of an Adjacent Transgene

Description of the experimental reporter system. We have previously described the enhancer-trap system in flies that contains a reporter transgene P$_{lexAop\text{-}hsp70}$-lacZ activated by a nearby enhancer in a Brahma complex (SWI/SNF)-dependent manner (Shidlovskii et al., 2021). The flies carry a P-element insertion of the reporter *lacZ* gene into the 5′ region of the *Dad* gene in the 3R arm of the third chromosome (**Figure 3A**). The reporter was placed under the control of the minimal *hsp70* promoter fused with the operator sites for the DNA-binding domain (DBD) of the LexA protein (P$_{lexAop\text{-}hsp70}$-lacZ). The flies also have a second transgene, which is integrated in the attP2 site in the 3L arm of the same chromosome (Markstein et al., 2008; Pfeiffer et al., 2008, Pfeiffer et al., 2010), thus expressing a fusion

protein that consists of one of the signature subunits of the Brahma complex (SAYP or BAP170) and DBD of the LexA protein (**Figure 3A**). These subunits are conservative (SAYP is a homologue of PHF10 in humans; BAP170 is a homologue of ARID2 in humans and RSC9 in yeast) and specify the PBAP/PBAF subtype of the Brahma (SWI/SNF) chromatin remodeling complex. The driver transgene in 3L was placed under the control of P$_{tub}$ (BAP170 fusion) or P$_{BAP170}$ (SAYP fusion) promoter. As was previously shown, targeted recruitment of the SAYP-lexA or BAP170-lexA fusion to the LexA operator sites in the *lacZ* promoter was accompanied by reporter gene activation in an enhancer-dependent manner with a pattern similar to that of the endogenous *Dad* gene (Shidlovskii et al., 2021). In this regard, it is of interest to elucidate whether such activation is accompanied by spatial convergence of the captured enhancer and the reporter gene promoter. For this purpose, a 3C library was prepared from 3rd instar larvae of fly lines carrying the P$_{lexAop\text{-}hsp70}$-lacZ reporter transgene with any of the two other transgenes that coded for a fusion of either SAYP or BAP170 with the LexA DBD (P$_{tub}$-lexA-BAP170 or P$_{BAP170}$-lexA-SAYP driver transgenes). Larvae that carried only the P$_{lexAop\text{-}hsp70}$-lacZ reporter transgene were used as a control.

Genetic background affects digestion by restriction enzyme in 3C procedure. The chromatin of the P$_{tub}$-lexA-BAP170- and P$_{BAP170}$-lexA-SAYP-containing lines was digested normally with DpnII and efficiently ligated after that (**Figure 3B**), whereas the chromatin of the line containing only the P$_{lexAop\text{-}hsp70}$-lacZ transgene was inefficiently digested with DpnII in parallel experiments, leading to poor ligation in the case of this line (**Figure 3C**, left and central panels). Halving the concentration of the fixing agent did not significantly improve digestion and ligation (**Figure 3C**, right panel). Moreover, the yields of DNA and especially total RNA were significantly reduced in the case of the P$_{lexAop\text{-}hsp70}$-lacZ-only line as compared with the driver-containing lines (compare **Figure 3B**, two left and two right panels). We concluded that the genetic background prevents efficient digestion and ligation of chromatin in the case of the P$_{lexAop\text{-}hsp70}$-lacZ-only line. It is possible that the RE site is methylated at A or C in this line and its digestion is thus blocked (Kunert et al., 2003; Zhang et al., 2015; Deshmukh et al., 2018), but this assumption requires further study. Thus, these important observations suggest that the genetic background can influence the success of the 3C procedure with different fly lines. To overcome the difficulties with the genetic background of the control P$_{lexAop\text{-}hsp70}$-lacZ line, we used meiotic crossing over between chromosomes as a powerful genetic technique for
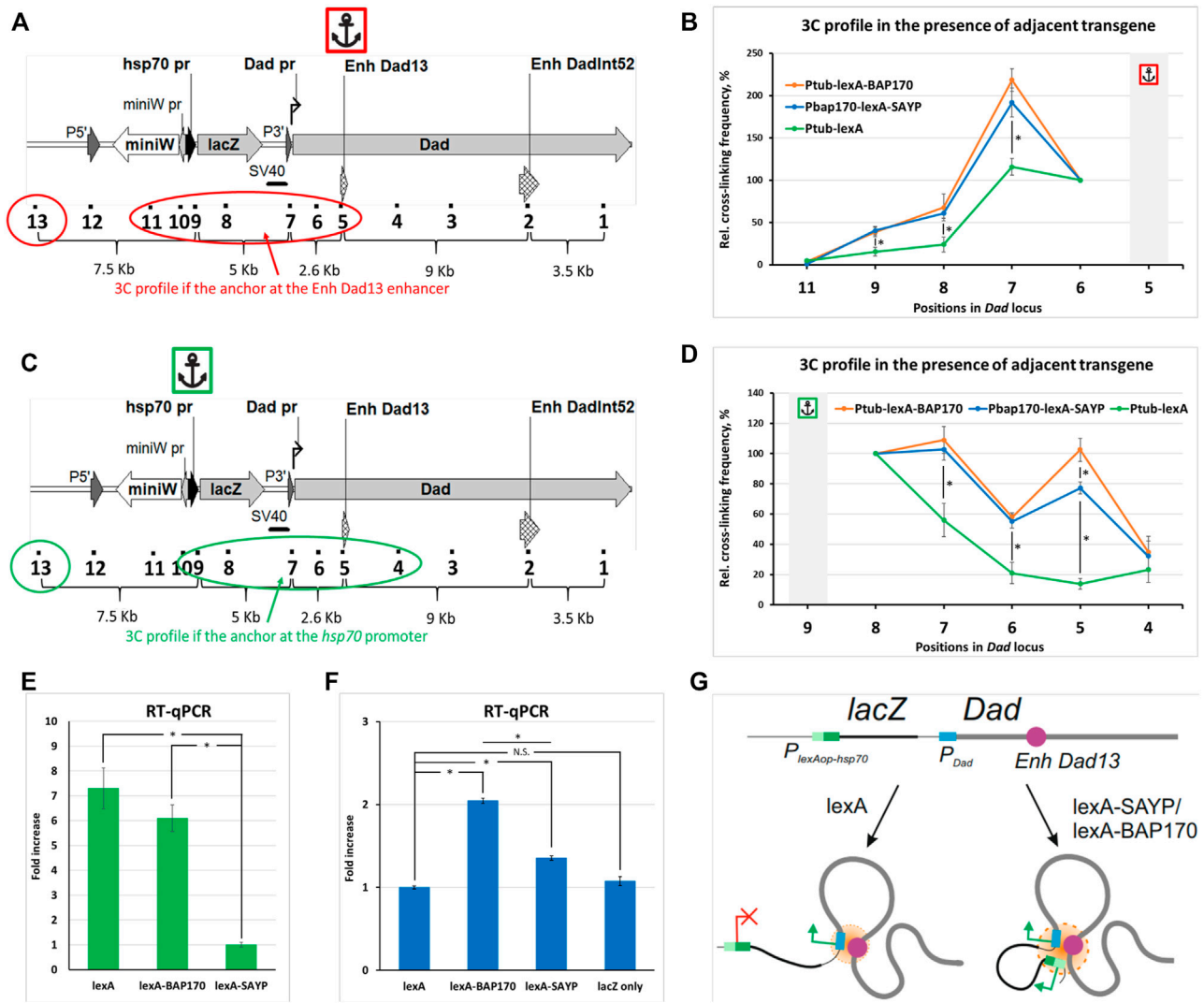
FIGURE 4 | Comparison of the 3C profiles in the *Dad* locus in the presence of an adjacent transgene. **(A)** The model locus *Dad* with the adjacent transgene is shown. Designations are as shown in **Supplementary Figures S3A and S6**. A red circle and oval encompass the points under study. An anchor icon indicates the location of the anchor primer at the *Dad13* enhancer. **(B)** Interaction profiles for indicated positions in *Dad* locus with the adjacent reporter transgene were obtained in the presence of the P$_{tub}$-lexA-BAP170, P$_{BAP170}$-lexA-SAYP, or control P$_{tub}$-lexA driver transgene with the anchor at the *Dad13* enhancer. Extraction, collection, and processing of larvae and 3C library preparation were done as shown in **Figure 2B** for larvae except that 150 larvae were collected and 50 mg of larval cell material was taken for 3C library preparation. Seven independent 3C libraries were prepared and analyzed for P$_{tub}$-lexA-BAP170 and P$_{BAP170}$-lexA-SAYP larvae and 4, for control P$_{tub}$-lexA larvae (two for each crossover line #2 and #6, respectively). The 3C experiment was done as shown in **Figure 2C** except that error bars indicate SEMs from 7 or 4 independent biological replicates of the 3C library for each case, respectively. Each experimental point for each 3C library was studied in 2–4 technical PCR replicates, and the data were averaged. Point 10 was excluded from the analysis since points 9 and 10 are adjacent. The RLFs of experimental samples were normalized to RLFs within the constitutively expressed *RpII* locus and to point #13 (intergenic spacer). The values between the two normalizations were averaged. An asterisk indicates the significance level: *$p < 0.05$, $n = 12$. **(C)** The same as shown in **(A)**, but the anchor is at the P$_{lexAop-hsp70}$-lacZ promoter and the points under study are encompassed in green. **(D)** The same as shown in **(B)**, but the anchor is at the P$_{lexAop-hsp70}$-lacZ promoter. The number of the prepared and analyzed 3C libraries is the same as shown in **(B)**. The frequency of ligation of the anchor fragment with the adjacent fragment (point #8) was arbitrarily taken as 100%, and the values of all other experimental points were calculated proportionally. An asterisk indicates the significance level: *$p < 0.05$, $n = 12$. **(E)** RT–qPCR analysis of the expression level of the driver transgenes in mutant larvae. mRNA expression of P$_{tub}$-lexA-BAP170, P$_{BAP170}$-lexA-SAYP, and control driver P$_{tub}$-lexA was studied using primers annealing to the LexA domain-coding sequence. The following genotypes were studied: P$_{lexAop-hsp70}$-lacZ, P$_{tub}$-lexA-BAP170; P$_{lexAop-hsp70}$-lacZ, P$_{BAP170}$-lexA-SAYP; P$_{lexAop-hsp70}$-lacZ, P$_{tub}$-lexA. RT–qPCR was carried out as shown in **Figure 2E** except that only the ΔΔCt method was used for calculation and four independent total RNAs were prepared for each driver line. The expression level in the line carrying the P$_{lexAop-hsp70}$-lacZ reporter and the P$_{BAP170}$-lexA-SAYP driver was arbitrarily taken as 1. An asterisk indicates the significance level: *$p < 0.05$, $n = 8$. **(F)** RT–qPCR analysis of the expression level of the reporter transgene *lacZ* in mutant larvae. RT–qPCR was done as shown in **(E)**. The following genotypes were studied: P$_{lexAop-hsp70}$-lacZ, P$_{tub}$-lexA; P$_{lexAop-hsp70}$-lacZ, P$_{tub}$-lexA-BAP170; P$_{lexAop-hsp70}$-lacZ, P$_{BAP170}$-lexA-SAYP, and P$_{lexAop-hsp70}$-lacZ only. RT–qPCR was done as shown in **Figure 2E** except that only the ΔΔCt method was used for calculation and 4 independent total RNAs were prepared for each line. The expression level in the line carrying the P$_{lexAop-hsp70}$-lacZ reporter and P$_{tub}$-lexA driver was arbitrarily taken as 1. N.S., non-significant ($p > 0.05$). An asterisk indicates the significance level: *$p < 0.05$, $n = 8$. **(G)** A model illustrating the role of the tethered (t) SAYP/BAP170 in distal interactions in the *Dad* locus carrying an adjacent transgene. The endogenous *Dad* promoter is regulated by the endogenous Dad enhancer, the two elements always forming a contact. In the absence of tSAYP/BAP170, the transgene does not establish contact with the *Dad* gene. Recruitment of SAYP/BAP170 to the reporter promoter induces its recruitment into a joint regulatory hub with the *Dad* enhancer and *Dad* promoter.

separating harmful mutations and diluting the genetic background (see a recombination scheme in **Supplementary Figure S13**). Two independent lines carrying the P$_{lexAop-hsp70}$-lacZ reporter transgene in the presence of the P$_{tub}$-lexA control driver were obtained. The presence of both transgenes in crossover flies was confirmed by PCR (**Figure 3E**). Then 3C libraries were prepared in parallel for each new control line. Representative examples of the resulting 3C libraries are given in **Figure 3F**.

Recruitment of SAYP and BAP170 to the reporter promoter induces its convergence to the endogenous enhancer. To verify that LexA-BAP170/LexA-SAYP-mediated activation of the *lacZ* reporter is accompanied by spatial convergence of the captured enhancer and the reporter gene promoter, we compared the 3C profiles of the *Dad* locus with the adjacent P$_{lexAop-hsp70}$-lacZ transgene in the control line, in the presence of the P$_{tub}$-lexA driver, with the 3C profiles obtained for the two experimental lines, in the presence of the P$_{tub}$-lexA-BAP170 or P$_{BAP170}$-lexA-SAYP driver. We observed that the ligation frequency between the *lacZ* reporter promoter and *Dad* enhancer in the control line in the presence of P$_{tub}$-lexA driver transgene was significantly lower than in lines with the LexA-SAYP/BAP170 drivers, either in direct (the anchor primer was located on the *Dad* enhancer) (**Figures 4A and B**) or in reciprocal (the anchor primer was located on the reporter transgene promoter) (**Figures 4C and D**) experiments. We concluded that transcription activation mediated by the tethering of the Brahma complex subunit SAYP or BAP170 to the reporter promoter is accompanied by spatial convergence of the captured enhancer and the reporter *lacZ* gene promoter. This correlates well with a *Dad*-like expression pattern of *lacZ* in the presence of the LexA-BAP170 or LexA-SAYP (Shidlovskii et al., 2021).

Increased expression of the driver is accompanied by an increase in the ligation frequency of the enhancer with the promoter and enhanced reporter expression. When comparing the 3C profiles of the P$_{tub}$-lexA-BAP170- and P$_{BAP170}$-lexA-SAYP-containing driver lines in a direct experiment, no significant differences were found between the lines and, in both the lines, the strongest interaction of the *Dad13* enhancer was observed with the endogenous *Dad* promoter, which is closer to the *Dad13* enhancer, than the *lacZ* promoter is (**Figures 4A and B**), much the same as in WT larvae (**Figure 2C**). When comparing the 3C profiles of the P$_{tub}$-lexA-BAP170 and P$_{BAP170}$-lexA-SAYP lines in a reciprocal experiment, a significant difference in the ligation frequency of the *lacZ* promoter and *Dad13* enhancer was found between the lines (point #5). The difference might result from different expression levels of the driver under the control of different promoters, P$_{tub}$ (constitutive strong promoter of the *α1-tubulin* gene) or P$_{BAP170}$ (−373/+135 bp). To check, we measured the abundance of the driver mRNA in the P$_{tub}$-lexA-BAP170 and P$_{BAP170}$-lexA-SAYP lines and the control P$_{tub}$-lexA line by RT-qPCR. We observed that the expression of P$_{tub}$-lexA-BAP170 was indeed significantly higher than that of P$_{BAP170}$-lexA-SAYP and was comparable to that in the control P$_{tub}$-lexA line (**Figure 4E**). Thus, there is a

direct relationship between the level of lexA-SAYP/BAP170 drivers and the ligation frequency of the *lacZ* promoter with the *Dad13* enhancer.

To verify that the difference in expression of the P$_{tub}$-lexA-BAP170 and P$_{BAP170}$-lexA-SAYP transgenic constructs results in differences in expression of the *lacZ* reporter gene, we measured the *lacZ* mRNA level in the P$_{tub}$-lexA-BAP170/P$_{BAP170}$-lexA-SAYP lines, the control P$_{tub}$-lexA-containing line, and the P$_{lexAop-hsp70}$-lacZ only line. Indeed, transcriptional activation of *lacZ* correlated with the driver type: significant differences in *lacZ* mRNA expression level were found between the P$_{tub}$-lexA-BAP170 and P$_{BAP170}$-lexA-SAYP lines (**Figure 4F**). Both P$_{tub}$-lexA-BAP170 and P$_{BAP170}$-lexA-SAYP driver lines differed in *lacZ* expression from the control P$_{tub}$-lexA-containing line, and no significant difference was found between the P$_{tub}$-lexA-containing line and the P$_{lexAop-hsp70}$-lacZ only line. The control line, which carried only the P$_{lexAop-hsp70}$-lacZ transgene, showed almost no activation of transcription (**Figure 4F**). We concluded that greater expression of the driver results in more pronounced upregulation of reporter gene expression. At the same time, low expression of the P$_{BAP170}$-lexA-SAYP driver was enough to induce interactions between the *lacZ* reporter promoter and the *Dad13* enhancer.

Thus, recruitment of the SAYP/BAP170 subunit to the reporter gene promoter mediates its interaction with the enhancer, and this is accompanied by changes in chromatin fiber conformation.

We hypothesized that the activation of expression is apparently accompanied by the association of compatible (co-regulated in our case) regulatory elements into a complex (chromatin hub), which probably shares transcription factors and other common components of the transcription machinery (**Figure 4G**). Our data also suggest that recruitment of the PBAP chromatin remodeling complex to the promoter may be a prerequisite for establishing the interaction with enhancers in *Drosophila*.

3 DISCUSSION

The principles of developmental gene functioning are an important problem in current biology. The expression of developmental genes is tightly linked to their conformation. Conformational changes in developmental genes can proceed simultaneously with the activation of their expression (Schoenfelder et al., 2015a; Bonev et al., 2017; Freire-Pritchett et al., 2017; Rubin et al., 2017; Schoenfelder et al., 2018b; Novo et al., 2018), or a specific conformation can pre-exist and serves as a scaffold to facilitate the activation (Jin et al., 2013; Ghavi-Helm et al., 2014; Schoenfelder et al., 2015b; Bonev et al., 2017; Cruz-Molina et al., 2017; Rubin et al., 2017; Comoglio et al., 2018; Ing-Simmons et al., 2021). Interactions between *cis*-regulatory elements in both cases can be investigated using 3C-based methods. In the present work, we investigated the 3C procedure for whole *Drosophila* larvae using a developmental gene *Dad* (daughters against *decapentaplegic*) as a model. We give a complete description of the 3C experiment on *Drosophila*

larvae, including isolation of larvae from fly food, their sorting by age, fixation, homogenization, preparation of cellular material, preparation of experimental 3C and control libraries, analysis of interactions, and calculation of the experimental results.

3.1 The 3C Procedure for Whole Larvae and a *Dad* Gene Model

The 3C protocol has previously been optimized with S2 cells (Bylino et al., 2021). Here, we applied the procedure to *Drosophila* 3rd instar larvae. The steps that we adapted from the protocols previously developed for cultured cells and zygotes of mammals (Rao et al., 2014; Flyamer et al., 2017; Golov et al., 2020b) (see **Supplementary Figure S6** legend) also proved applicable to *Drosophila* primary cells obtained from whole 3rd instar larvae. The quality of the resulting 3C libraries was acceptable, as demonstrated by electrophoresis (**Figure 2B** and **Figures 3B and F**). To further validate our procedure, we used it to study the conformation and distal interactions between regulatory elements in the developmental locus *Dad* in WT larvae. We observed that the conformation differed between cultured S2 cells (a repressive conformation) and live larvae (an active conformation) (**Figure 2D**), and this observation correlated well with *Dad* expression level in larvae and S2 cells (**Figure 2E**). The *Dad* gene codes for a negative feedback regulator of the BMP/Dpp/TGF-β signaling pathway. Its low expression in S2 cells is in good agreement with the literature data that S2 cells produce a weak response to treatment with the Dpp ligand since these cells do not express some components of the BMP/Dpp/TGF-β pathway (Cherbas et al., 2011; Neal et al., 2019). Low or altered expression of developmental genes in cultured cells is a general problem typical not only for *Drosophila* but also for mammals and can be overcome at least partly by using primary cell cultures (Antequera et al., 1990; Kitsis and Leinwand, 1992; Zaitseva et al., 2006).

3.2 Enhancers of the *Dad* Locus

The *Dad* locus contains two functional enhancers: *Dad13* is the first enhancer in the chain (a primary enhancer) and the *DadInt52* enhancer downstream (a shadow enhancer) (Weiss et al., 2010). The *Dad13* enhancer is thought to be the primary element that drives the expression of the gene (Weiss et al., 2010; Neal et al., 2019). According to our data, a two-humped curve is absent in the graph in **Figure 2C**, when an anchor is on the *Dad13* enhancer, indicating that the *Dad13* enhancer is not in complex with the *Dad* promoter and *DadInt52* enhancer. This is in good agreement with the leading role that the first enhancer in a chain of gene enhancers plays in regulating expression in higher eukaryotes and their evolutionary precursors (Song et al., 2019; Bylino et al., 2020). The finding additionally correlates well with observations by Weiss et al. (2010) and Neal et al. (2019), who showed that the *Dad13* enhancer provides a proper transcriptional response and drives the expression of the reporter gene in a pattern virtually identical to that of endogenous *Dad*, while *DadInt52*-driven expression is weak and only partially overlaps that of *Dad13*. Altogether, our data emphasize that the *DadInt52* enhancer is a shadow enhancer and plays an

auxiliary or redundant role in comparison with the *Dad13* enhancer. However, it cannot be ruled out that *DadInt52* is involved in the regulation at a developmental stage other than embryos (Weiss et al., 2010) or 3rd instar larvae (this work). Our results concerning the leading role of the first enhancer are in good agreement with the results by Bieli et al. (2015), who have studied the enhancer–promoter interactions in whole larvae for the developmental gene *apterous*. They have similarly found that the *apterous* enhancer closest to the promoter (the first in the chain) interacts with the promoter more strongly than the second enhancer (the second in the chain). Moreover, the leading enhancer interacted with the promoter more strongly than with the control region situated downstream of the *apterous*, like in our case (point #13). This suggests that the mode of regulation is universal for developmental genes and emphasizes the leading role of the first enhancer in a chain of enhancers.

3.3 3C Experiment With WT and Mutant Larvae

Having studied the interactions in cultured cells and WT larvae, we applied our procedure to the mutant fly lines that carried a reporter transgene adjacent to the *Dad* gene and demonstrated conformational changes in the *Dad* locus upon activation of transgene expression (**Figures 4B and D**). The induction of expression was achieved through the recruitment (targeted tethering) of SAYP or BAP170 subunits of the SWI/SNF remodeling complex to the reporter promoter and subsequent assembly of the SWI/SNF complex (Shidlovskii et al., 2021). Recruitment of SAYP/BAP170 to the *lacZ* promoter leads to the discrimination of this promoter among the other promoters of the transgene (P-element promoter and *mini-white* promoter) and specific activation of *lacZ* expression following a *Dad*-like pattern (Shidlovskii et al., 2021), thus, providing an example of specificity of enhancer–promoter communication (Galouzis and Furlong, 2022). Accordingly, we hypothesized that the targeted recruitment of the specific SWI/SNF subunits to the promoter is accompanied not only by an increase in gene expression but also by spatial convergence of the regulatory elements. We observed that this was the case (**Figures 4B and D**). At the same time, surprisingly, the recruitment of the SAYP/BAP170 resulted in more pronounced interactions of the *Dad13* enhancer not only with the reporter promoter but also with the *Dad* promoter (**Figure 4B**) and enhanced interaction of the two promoters with each other (**Figure 4D**). We hypothesized that both promoters and the *Dad13* enhancer combine to form a ternary complex (chromatin hub), which possesses a more stable conformation compared with a dual *Dad* promoter-enhancer complex and provides a basis for a higher ligation frequency (**Figure 4F**). Alternatively, increased ligation between regulatory elements may be interpreted as an activation chromatin hub, where enhancement of both promoter activation might take place due to the looser chromatin structure induced by targeted recruitment of SAYP/BAP170 and subsequently of Brahma complex. However, we do not know whether the *Dad13* enhancer activates both promoters simultaneously or sequentially or whether this activation occurs in different cells

[discussed in (Peng and Zhang, 2018)]. The first scenario is supported by many observations of the co-localization of regulatory elements of co-expressed genes within the same transcription factory (Osborne et al., 2004; Mitchell and Fraser, 2008; Sutherland and Bickmore, 2009; Rieder et al., 2012). Such genes usually share the components of the transcription apparatus, and co-expression of two transgenes/endogenes closely located in the genome and regulated with the same enhancer and competition between two promoters for one enhancer is well-documented in the literature (Raj et al., 2006; Papantonis et al., 2010; Bartman et al., 2016; Fukaya et al., 2016; Lim et al., 2018; Stavreva et al., 2019).

3.4 Competition Between Two Promoters for the Same Enhancer

Competition between the *lacZ* and *Dad* promoters can also occur in order to establish separate contacts with the enhancer. Indeed, mutually inhibitory co-expression of genes has been demonstrated when two different sequentially located globin genes are regulated by one enhancer (LCR) *in cis* (Deng et al., 2014; Bartman et al., 2016). It strongly resembles the organization of the *Dad* locus under study. In this model, one would expect an "enhancer hijacking" phenomenon. In the case of hijacking of the *Dad13* enhancer by the transgenic promoter, the 3C profile would show a decrease in interaction with the endogenous *Dad* promoter, which would be accompanied by a corresponding increase in interaction with the transgenic promoter. However, it was not detected. Moreover, in our case, *lacZ* is apparently expressed at a significantly lower level than the *Dad* gene (data not shown). Due to this, it is not possible to reliably determine by RT-qPCR whether an increase in *lacZ* expression in the presence of SAYP/BAP170 is accompanied by a corresponding decrease in *Dad* endogene expression. The presence of the *lacZ* transgene has just a slightly negative effect on *Dad* expression (data not shown) and may be caused by disruption of *Dad* promoter structure (hypomorphic mutation) rather than by enhancer hijacking. Thus, our data indicate that, even if the reporter promoter hijacks some of the *Dad13* enhancer activity, it does so in a very weak manner.

3.5 Activation of RNAPII Promoter by RSC/PBAP/PBAF Complex Depends on Natural Genomic Context

Our previous data showed that targeted recruitment of the SAYP/PHF10 PBAP signature subunit fused with GAL4 DBD to the *hsp70* promoter [−259/+198, containing GAF and six HSF sites (Ingolia et al., 1980; Steller and Pirrotta, 1984, 1985; Weber et al., 1997; Wu et al., 2001; Kwak et al., 2013)] under the control of UAS strongly induces reporter gene expression in a PBAP-dependent manner (Vorobyeva et al., 2009). These experiments were performed with a transient transfection model and a model of a reporter transgene integrated into the genome in multiple copies in a *Drosophila* S2 cell culture. In our experiments described in this work and in the study by Shidlovskii et al. (2021), the upstream LexA binding sites were

combined with the minimal *hsp70b* core promoter (−45/+207 bp), lacking upstream GAF and HSF binding sites, and this reporter transgene was placed in genomic context in a single copy. Targeted recruitment of the SAYP/BAP170 subunits to the promoter was insufficient for reporter activation, required an active enhancer to occur in the genomic vicinity, and was accompanied by the convergence of the enhancer and the *hsp70* promoter. Although the *hsp70* promoters used in the previous and current models differ in length, our data may point to the importance of the genomic context: a reporter located in a plasmid or in the genome may show different behaviour. Apparently, the chromatin environment and its 3D conformation make a significant contribution to the reporter gene activity.

Previously, the effect of targeted Brahma subunit recruitment on gene activity was tested in yeast and human cells. Recruitment of the RSC/PBAP/PBAF complex subunits Sth1 or Swh3/Rsc8 (MOR/BAF155/170) to the core promoter did not induce expression of the RNAPII promoter (Laurent et al., 1992; Treich and Carlson, 1997), whereas recruitment of the ySWI/SNF/BAP/BAF complex subunits SNF2, SNF5, SNF6, and SNF11 to the same promoter sufficed to induce expression (Laurent et al., 1990; Laurent et al., 1991; Laurent et al., 1992; Treich et al., 1995). Thus, these data may indicate that different roles are played by the ySWI/SNF and RSC complexes in RNAPII promoter regulation in yeast. Subunits of the human BAF complex, hBRM and hDPF3a (D4/TTH/BAF45B/C/D), induced expression of several RNAPII promotes upon their targeted recruitment in human cell cultures (Muchardt and Yaniv, 1993; Trouche et al., 1997; Cui et al., 2016). All of these promoters possess a long enhancer/promoter region, and their BAF-driven activation depends on cell activators. However, the reporter was in a plasmid in all of these cases; therefore, the data cannot provide direct evidence that the test proteins act similarly in the natural genomic context.

Genome-wide data show that BAF occupies enhancer regions, while PBAF is found on both enhancers and promoters (Nakayama et al., 2017; Carcamo et al., 2022). We also found that PBAP occupies both *Dad* enhancer and promoter (Shidlovskii et al., 2021). We hypothesized that, in addition to the catalytic function, PBAP/PBAF mediates interactions between enhancers and promoters in a non-catalytic mode (Kwok et al., 2015; Jordan-Pla et al., 2018).

3.6 Role of SWI/SNF in Establishing Long-Range Contacts Between Enhancers and Promoters

Since the targeted recruitment of the SWI/SNF subunits to the reporter promoter and its activation is accompanied by changes in the conformation of the chromatin fiber, we hypothesized that SWI/SNF may facilitate the organization of long-range regulatory gene contacts in the interphase in *Drosophila*. This mechanism of gene regulation might be basic, more ancient, and more universal than the regulation *via* cohesin-dependent loop extrusion, which is absent in *Drosophila* (Rowley et al., 2017; Eagen, 2018; Gambetta and Furlong, 2018; Matthews and White, 2019;

Bylino et al., 2020; Kaushal et al., 2021). The hypothesis about the role of SWI/SNF in the spatial regulation of gene expression is strongly supported in the literature. For example, loop formation mediated by the SWI/SNF complexes was found in microscopic (Bazett-Jones et al., 1999), real-time *in vitro* (Lia et al., 2006; Zhang et al., 2006; Sirinakis et al., 2011), biochemical, and genetic studies (Ni et al., 2008; Kim et al., 2009a, Kim et al., 2009b; Shi et al., 2013; Jégu et al., 2014). Even in mammals, where cohesin brings together distant regions of immunoglobulin genes, regulating the switching of types of antibodies by the loop extrusion mechanism (Thomas-Claudepierre et al., 2013; Zhang et al., 2019), it has been shown that SWI/SNF assists this process (Kwon et al., 2000; Bossen et al., 2015). The interplay between SWI/SNF and CTCF on their DNA binding sites has also been detected (Euskirchen et al., 2011; Barutcu et al., 2016; Wood et al., 2016); similar data have been reported for condensin II and SWI/SNF (Wu et al., 2019). Thus, in organisms where loop extrusion appears to be the main regulatory mechanism of distant interactions, SWI/SNF also contributes to the organization of this type of regulation. Our data indicate that SWI/SNF is necessary for establishing enhancer–promoter communication in *Drosophila*.

4 CONCLUSION

(i) The 3C procedure optimized using S2 cells as a model can be successfully applied to studying the chromatin conformation and interactions between regulatory elements of developmental genes in living *Drosophila* larvae.

(ii) Whole larvae can be used to study not only ubiquitous intragenomic interactions but also tissue-specific ones. Our results indicate that, even if an interaction is specific for a limited set of cells in larvae, the 3C procedure using whole larvae allows its quantitative measurement and comparison in different lines.

(iii) The genetic background may affect the overall DNA yield and digestion with a restriction enzyme in the 3C procedure.

(iv) The model locus *Dad* exists in different conformations in 3rd instar larvae and cultured S2 embryonic cells. The active conformation correlates with the transcriptional activity of the gene in living larvae.

(v) Targeted recruitment of SAYP and the BAP170 subunits (SWI/SN and PBAP) to a reporter promoter induces the convergence of the promoter and endogenous enhancer. This is accompanied by an increase in reporter gene expression.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

Conceptualization and methodology: YS, OB, EG, and FAD; formal analysis, investigation, and visualization: OB, AI, and YS; writing: OB and YS; supervision, project administration, and funding acquisition: YS and FAD. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.734208/full#supplementary-material

REFERENCES

Akgol Oksuz, B., Yang, L., Abraham, S., Venev, S. V., Krietenstein, N., Parsi, K. M., et al. (2021). Systematic Evaluation of Chromosome Conformation Capture Assays. *Nat. Methods* 18, 1046–1055. doi:10.1038/s41592-021-01248-7

Antequera, F., Boyes, J., and Bird, A. (1990). High Levels of De Novo Methylation and Altered Chromatin Structure at CpG Islands in Cell Lines. *Cell.* 62, 503–514. doi:10.1016/0092-8674(90)90015-7

Bartman, C. R., Hsu, S. C., Hsiung, C. C.-S., Raj, A., and Blobel, G. A. (2016). Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Mol. Cell.* 62, 237–247. doi:10.1016/j.molcel.2016.03.007

Barutcu, A. R., Lajoie, B. R., Fritz, A. J., McCord, R. P., Nickerson, J. A., van Wijnen, A. J., et al. (2016). SMARCA4 Regulates Gene Expression and Higher-Order Chromatin Structure in Proliferating Mammary Epithelial Cells. *Genome Res.* 26, 1188–1201. doi:10.1101/gr.201624.115

Bazett-Jones, D. P., Côté, J., Landel, C. C., Peterson, C. L., and Workman, J. L. (1999). The SWI/SNF Complex Creates Loop Domains in DNA and Polynucleosome Arrays and Can Disrupt DNA-Histone Contacts within These Domains. *Mol. Cell. Biol.* 19, 1470–1478. doi:10.1128/mcb.19.2.1470

Belaghzal, H., Dekker, J., and Gibcus, J. H. (2017). Hi-C 2.0: An Optimized Hi-C Procedure for High-Resolution Genome-wide Mapping of Chromosome Conformation. *Methods* 123, 56–65. doi:10.1016/j.ymeth.2017.04.004

Bernardo, T. J., Dubrovskaya, V. A., Xie, X., and Dubrovsky, E. B. (2014). A View through a Chromatin Loop: Insights into the Ecdysone Activation of Early

Genes in Drosophila. *Nucleic Acids Res.* 42, 10409–10424. doi:10.1093/nar/gku754

Bieli, D., Kanca, O., Requena, D., Hamaratoglu, F., Gohl, D., Schedl, P., et al. (2015). Establishment of a Developmental Compartment Requires Interactions between Three Synergistic Cis-Regulatory Modules. *PLOS Genet.* 11, e1005376. doi:10.1371/journal.pgen.1005376

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G. L., Lubling, Y., et al. (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell.* 171, 557–572. e24. doi:10.1016/j.cell.2017.09.043

Bossen, C., Murre, C. S., Chang, A. N., Mansson, R., Rodewald, H.-R., and Murre, C. (2015). The Chromatin Remodeler Brg1 Activates Enhancer Repertoires to Establish B Cell Identity and Modulate Cell Growth. *Nat. Immunol.* 16, 775–784. doi:10.1038/ni.3170

Bylino, O. V., Ibragimov, A. N., Pravednikova, A. E., and Shidlovskii, Y. V. (2021). Investigation of the Basic Steps in the Chromosome Conformation Capture Procedure. *Front. Genet.* 12, 733937. doi:10.3389/fgene.2021.733937

Bylino, O. V., Ibragimov, A. N., and Shidlovskii, Y. V. (2020). Evolution of Regulated Transcription. *Cells* 9, 1675. *Cells* 9. doi:10.3390/cells9071675

Carcamo, S., Nguyen, C. B., Grossi, E., Filipescu, D., Alpsoy, A., Dhiman, A., et al. (2022). Altered BAF Occupancy and Transcription Factor Dynamics in PBAF-Deficient Melanoma. *Cell. Rep.* 39, 110637. doi:10.1016/j.celrep.2022.110637

Cherbas, L., Willingham, A., Zhang, D., Yang, L., Zou, Y., Eads, B. D., et al. (2011). The Transcriptional Diversity of 25 Drosophila Cell Lines. *Genome Res.* 21, 301–314. doi:10.1101/gr.112961.110

Chintapalli, V. R., Wang, J., and Dow, J. A. T. (2007). Using FlyAtlas to Identify Better *Drosophila melanogaster* Models of Human Disease. *Nat. Genet.* 39, 715–720. doi:10.1038/ng2049

Collombet, S., Ranisavljevic, N., Nagano, T., Varnai, C., Shisode, T., Leung, W., et al. (2020). Parental-to-embryo Switch of Chromosome Organization in Early Embryogenesis. *Nature* 580, 142–146. doi:10.1038/s41586-020-2125-z

Comet, I., Schuettengruber, B., Sexton, T., and Cavalli, G. (2011). A Chromatin Insulator Driving Three-Dimensional Polycomb Response Element (PRE) Contacts and Polycomb Association with the Chromatin Fiber. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2294–2299. doi:10.1073/pnas.1002059108

Comoglio, F., Park, H. J., Schoenfelder, S., Barozzi, I., Bode, D., Fraser, P., et al. (2018). Thrombopoietin Signaling to Chromatin Elicits Rapid and Pervasive Epigenome Remodeling within Poised Chromatin Architectures. *Genome Res.* 28, 295–309. doi:10.1101/gr.227272.117

Cruz-Molina, S., Respuela, P., Tebartz, C., Kolovos, P., Nikolic, M., Fueyo, R., et al. (2017). PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. *Cell. Stem Cell.* 20, 689–705. e9. doi:10.1016/j.stem.2017.02.004

Cubeñas-Potts, C., Rowley, M. J., Lyu, X., Li, G., Lei, E. P., and Corces, V. G. (2017). Different Enhancer Classes in Drosophila Bind Distinct Architectural Proteins and Mediate Unique Chromatin Interactions and 3D Architecture. *Nucleic Acids Res.* 45, 1714–1730. doi:10.1093/nar/gkw1114

Cui, H., Schlesinger, J., Schoenhals, S., Tönjes, M., Dunkel, I., Meierhofer, D., et al. (2016). Phosphorylation of the Chromatin Remodeling Factor DPF3a Induces Cardiac Hypertrophy through Releasing HEY Repressors from DNA. *Nucleic Acids Res.* 44, 2538–2553. doi:10.1093/nar/gkv1244

Davies, J. O. J., Oudelaar, A. M., Higgs, D. R., and Hughes, J. R. (2017). How Best to Identify Chromosomal Interactions: a Comparison of Approaches. *Nat. Methods* 14, 125–134. doi:10.1038/nmeth.4146

Davies, J. O. J., Telenius, J. M., McGowan, S. J., Roberts, N. A., Taylor, S., Higgs, D. R., et al. (2016). Multiplexed Analysis of Chromosome Conformation at Vastly Improved Sensitivity. *Nat. Methods* 13, 74–80. doi:10.1038/nmeth.3664

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing Chromosome Conformation. *Science* 295, 1306–1311. doi:10.1126/science.1067799

Deng, W., Rupon, J. W., Krivega, I., Breda, L., Motta, I., Jahn, K. S., et al. (2014). Reactivation of Developmentally Silenced Globin Genes by Forced Chromatin Looping. *Cell.* 158, 849–860. doi:10.1016/j.cell.2014.05.050

Deshmukh, S., Ponnaluri, V. C., Dai, N., Pradhan, S., and Deobagkar, D. (2018). Levels of DNA Cytosine Methylation in theDrosophilagenome. *PeerJ* 6, e5119. doi:10.7717/peerj.5119

Du, Z., Zheng, H., Huang, B., Ma, R., Wu, J., Zhang, X., et al. (2017). Allelic Reprogramming of 3D Chromatin Architecture during Early Mammalian Development. *Nature* 547, 232–235. doi:10.1038/nature23263

Eagen, K. P., Aiden, E. L., and Kornberg, R. D. (2017). Polycomb-mediated Chromatin Loops Revealed by a Subkilobase-Resolution Chromatin Interaction Map. *Proc. Natl. Acad. Sci. U.S.A.* 114, 8764–8769. doi:10.1073/pnas.1701291114

Eagen, K. P., Hartl, T. A., and Kornberg, R. D. (2015). Stable Chromosome Condensation Revealed by Chromosome Conformation Capture. *Cell.* 163, 934–946. doi:10.1016/j.cell.2015.10.026

Eagen, K. P. (2018). Principles of Chromosome Architecture Revealed by Hi-C. *Trends Biochem. Sci.* 43, 469–478. doi:10.1016/j.tibs.2018.03.006

Edgar, B. A., and Orr-Weaver, T. L. (2001). Endoreplication Cell Cycles. *Cell.* 105, 297–306. doi:10.1016/s0092-8674(01)00334-8

Edgar, B., and Nijhout, H. (2004). "Growth and Cell Cycle Control in Drosophila," in *Cell Growth—Control of Cell Size*. Editor M. N. Hall, et al. (NY: Cold Spring HarborCold Spring Harbor Laboratory Press), 23–83.

El-Sharnouby, S., Fischer, B., Magbanua, J. P., Umans, B., Flower, R., Choo, S. W., et al. (2017). Regions of Very Low H3K27me3 Partition the Drosophila Genome into Topological Domains. *PLOS ONE* 12, e0172725. doi:10.1371/journal.pone.0172725

Erceg, J., AlHaj Abed, J., Goloborodko, A., Lajoie, B. R., Fudenberg, G., Abdennur, N., et al. (2019). The Genome-wide Multi-Layered Architecture of Chromosome Pairing in Early Drosophila Embryos. *Nat. Commun.* 10, 4486. doi:10.1038/s41467-019-12211-8

Euskirchen, G. M., Auerbach, R. K., Davidov, E., Gianoulis, T. A., Zhong, G., Rozowsky, J., et al. (2011). Diverse Roles and Interactions of the SWI/SNF Chromatin Remodeling Complex Revealed Using Global Approaches. *PLOS Genet.* 7, e1002008. doi:10.1371/journal.pgen.1002008

Flyamer, I. M., Gassler, J., Imakaev, M., Brandão, H. B., Ulianov, S. V., Abdennur, N., et al. (2017). Single-nucleus Hi-C Reveals Unique Chromatin Reorganization at Oocyte-To-Zygote Transition. *Nature* 544, 110–114. doi:10.1038/nature21711

Freire-Pritchett, P., Schoenfelder, S., Várnai, C., Wingett, S. W., Cairns, J., Collier, A. J., et al. (2017). Global Reorganisation of Cis-Regulatory Units upon Lineage Commitment of Human Embryonic Stem Cells. *eLife* 6, e21926. doi:10.7554/eLife.21926

Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer Control of Transcriptional Bursting. *Cell.* 166, 358–368. doi:10.1016/j.cell.2016.05.025

Gabdank, I., Ramakrishnan, S., Villeneuve, A. M., and Fire, A. Z. (2016). A Streamlined Tethered Chromosome Conformation Capture Protocol. *BMC Genomics* 17, 274. doi:10.1186/s12864-016-2596-3

Galouzis, C. C., and Furlong, E. E. M. (2022). Regulating Specificity in Enhancer-Promoter Communication. *Curr. Opin. Cell. Biol.* 75, 102065. doi:10.1016/j.ceb.2022.01.010

Gambetta, M. C., and Furlong, E. E. M. (2018). The Insulator Protein CTCF Is Required for Correct Hox Gene Expression, but Not for Embryonic Development in Drosophila. *Genetics* 210, 129–136. doi:10.1534/genetics.118.301350

Gavrilov, A. A., Golov, A. K., and Razin, S. V. (2013). Actual Ligation Frequencies in the Chromosome Conformation Capture Procedure. *PLOS ONE* 8, e60403. doi:10.1371/journal.pone.0060403

Ghavi-Helm, Y., Klein, F. A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., et al. (2014). Enhancer Loops Appear Stable during Development and Are Associated with Paused Polymerase. *Nature* 512, 96–100. doi:10.1038/nature13417

Gilbert, S. (2010). *Developmental Biology*. 9th edition. Sunderland, MA: Sinauer Associates.

Golov, A. K., Abashkin, D. A., Kondratyev, N. V., Razin, S. V., Gavrilov, A. A., and Golimbet, V. E. (2020a). A Modified Protocol of Capture-C Allows Affordable and Flexible High-Resolution Promoter Interactome Analysis. *Sci. Rep.* 10. doi:10.1038/s41598-020-72496-4

Golov, A. K., Ulianov, S. V., Luzhin, A. V., Kalabusheva, E. P., Kantidze, O. L., Flyamer, I. M., et al. (2020b). C-TALE, a New Cost-Effective Method for Targeted Enrichment of Hi-C/3c-Seq Libraries. *Methods* 170, 48–60. doi:10.1016/j.ymeth.2019.06.022

Grob, S., and Cavalli, G. (2018). "Technical Review: A Hitchhiker's Guide to Chromosome Conformation Capture," in Plant Chromatin Dynamics: Methods And Protocols *Methods in Molecular Biology*. Editors M. Bemer and C. Baroux (New York, NY: Springer), 233–246. doi:10.1007/978-1-4939-7318-7_14

Hua, P., Badat, M., Hanssen, L. L. P., Hentges, L. D., Crump, N., Downes, D. J., et al. (2021). Defining Genome Architecture at Base-Pair Resolution. *Nature* 595, 125–129. –5. doi:10.1038/s41586-021-03639-4

Hug, C. B., and Vaquerizas, J. M. (2018). Generation of Genome-wide Chromatin Conformation Capture Libraries from Tightly Staged Early Drosophila Embryos. *JoVE*, e57001. doi:10.3791/57001

Hughes, J. R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., et al. (2014). Analysis of Hundreds of Cis-Regulatory Landscapes at High Resolution in a Single, High-Throughput Experiment. *Nat. Genet.* 46, 205–212. doi:10.1038/ng.2871

Ing-Simmons, E., Vaid, R., Bing, X. Y., Levine, M., Mannervik, M., and Vaquerizas, J. M. (2021). Independence of Chromatin Conformation and Gene Regulation during Drosophila Dorsoventral Patterning. *Nat. Genet.* 53, 487–499. doi:10. 1038/s41588-021-00799-x

Ingolia, T., Craig, E. A., and McCarthy, B. J. (1980). Sequence of Three Copies of the Gene for the Major Drosophila Heat Shock Induced Protein and Their Flanking Regions. *Cell.* 21, 669–679. doi:10.1016/0092-8674(80)90430-4

Jégu, T., Latrasse, D., Delarue, M., Hirt, H., Domenichini, S., Ariel, F., et al. (2014). The BAF60 Subunit of the SWI/SNF Chromatin-Remodeling Complex Directly Controls the Formation of a Gene Loop at FLOWERING LOCUS C in Arabidopsis. *Plant Cell.* 26, 538–551. doi:10.1105/tpc.113.114454

Jin, F., Li, Y., Dixon, J. R., Selvaraj, S., Ye, Z., Lee, A. Y., et al. (2013). A High-Resolution Map of the Three-Dimensional Chromatin Interactome in Human Cells. *Nature* 503, 290–294. doi:10.1038/nature12644

Jordán-Pla, A., Yu, S., Waldholm, J., Källman, T., Östlund Farrants, A.-K., and Visa, N. (2018). SWI/SNF Regulates Half of its Targets without the Need of ATP-Driven Nucleosome Remodeling by Brahma. *BMC Genomics* 19, 367. doi:10.1186/s12864-018-4746-2

Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., and Chen, L. (2012). Genome Architectures Revealed by Tethered Chromosome Conformation Capture and Population-Based Modeling. *Nat. Biotechnol.* 30, 90–98. doi:10.1038/ nbt.2057

Kaushal, A., Mohana, G., Dorier, J., Özdemir, I., Omer, A., Cousin, P., et al. (2021). CTCF Loss Has Limited Effects on Global Genome Architecture in Drosophila Despite Critical Regulatory Functions. *Nat. Commun.* 12, 1011. doi:10.1038/ s41467-021-21366-2

Ke, Y., Xu, Y., Chen, X., Feng, S., Liu, Z., Sun, Y., et al. (2017). 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. *Cell.* 170, 367–381.e20. doi:10.1016/j.cell.2017.06.029

Kim, S.-I., Bresnick, E. H., and Bultman, S. J. (2009a). BRG1 Directly Regulates Nucleosome Structure and Chromatin Looping of the α Globin Locus to Activate Transcription. *Nucleic Acids Res.* 37, 6019–6027. doi:10.1093/nar/gkp677

Kim, S.-I., Bultman, S. J., Kiefer, C. M., Dean, A., and Bresnick, E. H. (2009b). BRG1 Requirement for Long-Range Interaction of a Locus Control Region with a Downstream Promoter. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2259–2264. doi:10. 1073/pnas.0806420106

Kitsis, R. N., and Leinwand, L. A. (1992). Discordance between Gene Regulation *In Vitro* and *In Vivo*. *Gene Expr.* 2, 313–318.

Kunert, N., Marhold, J., Stanke, J., Stach, D., and Lyko, F. (2003). A Dnmt2-like Protein Mediates DNA Methylation in Drosophila. *Dev. Camb. Engl.* 130, 5083–5090. doi:10.1242/dev.00716

Kuter, D., Gminski, D., and Rosenberg, R. (1992). Transforming Growth Factor Beta Inhibits Megakaryocyte Growth and Endomitosis. *Blood* 79, 619–626. doi:10.1182/blood.v79.3.619.bloodjournal793619

Kwak, H., Fuda, N. J., Core, L. J., and Lis, J. T. (2013). Precise Maps of RNA Polymerase Reveal How Promoters Direct Initiation and Pausing. *Science* 339, 950–953. doi:10.1126/science.1229386

Kwok, R. S., Li, Y. H., Lei, A. J., Edery, I., and Chiu, J. C. (2015). The Catalytic and Non-catalytic Functions of the Brahma Chromatin-Remodeling Protein Collaborate to Fine-Tune Circadian Transcription in Drosophila. *PLOS Genet.* 11, e1005307. doi:10.1371/journal.pgen.1005307

Kwon, J., Morshead, K. B., Guyon, J. R., Kingston, R. E., and Oettinger, M. A. (2000). Histone Acetylation and hSWI/SNF Remodeling Act in Concert to Stimulate V(D)J Cleavage of Nucleosomal DNA. *Mol. Cell.* 6, 1037–1048. doi:10.1016/s1097-2765(00)00102-7

Lafontaine, D. L., Yang, L., Dekker, J., and Gibcus, J. H. (2021). Hi-C 3.0: Improved Protocol for Genome-Wide Chromosome Conformation Capture. *Curr. Protoc.* 1, e198. doi:10.1002/cpz1.198

Larkin, A., Marygold, S. J., Antonazzo, G., Attrill, H., dos Santos, G., Garapati, P. V., et al. (2021). FlyBase: Updates to the *Drosophila melanogaster* Knowledge Base. *Nucleic Acids Res.* 49, D899–D907. doi:10.1093/nar/gkaa1026

Laurent, B. C., Treitel, M. A., and Carlson, M. (1990). The SNF5 Protein of Saccharomyces Cerevisiae is a Glutamine- and Proline-Rich Transcriptional Activator That Affects Expression of a Broad Spectrum of Genes. *Mol. Cell. Biol.* 10, 5616–5625. doi:10.1128/mcb.10.11.5616-5625.1990

Laurent, B. C., Treitel, M. A., and Carlson, M. (1991). Functional Interdependence of the Yeast SNF2, SNF5, and SNF6 Proteins in Transcriptional Activation. *Proc. Natl. Acad. Sci.* 88, 2687–2691. doi:10.1073/pnas.88.7.2687

Laurent, B. C., and Carlson, M. (1992). Yeast SNF2/SWI, SNF5, and SNF6 Proteins Function Coordinately With the Gene-Specific Transcriptional Activators GAL4 and Bicoid. *Genes Dev.* 6, 1707–1715. doi:10.1101/gad.6.9. 1707

Lee, H., McManus, C. J., Cho, D.-Y., Eaton, M., Renda, F., Somma, M. P., et al. (2014). DNA Copy Number Evolution in Drosophila Cell Lines. *Genome Biol.* 15. doi:10.1186/gb-2014-15-8-r70

Lee, H. O., Davidson, J. M., and Duronio, R. J. (2009). Endoreplication: Polyploidy with Purpose. *Genes. Dev.* 23, 2461–2477. doi:10.1101/gad.1829209

Li, H.-B. (2016). "Chromosome Conformation Capture in Drosophila," in Polycomb Group Proteins: Methods And Protocols *Methods in Molecular Biology*. Editors C. Lanzuolo and B. Bodega (New York, NY: Springer), 207–212. doi:10.1007/978-1-4939-6380-5_18

Li, L., Lyu, X., Hou, C., Takenaka, N., Nguyen, H. Q., Ong, C.-T., et al. (2015). Widespread Rearrangement of 3D Chromatin Organization Underlies Polycomb-Mediated Stress-Induced Silencing. *Mol. Cell.* 58, 216–231. doi:10. 1016/j.molcel.2015.02.023

Lia, G., Praly, E., Ferreira, H., Stockdale, C., Tse-Dinh, Y. C., Dunlap, D., et al. (2006). Direct Observation of DNA Distortion by the RSC Complex. *Mol. Cell.* 21, 417–425. doi:10.1016/j.molcel.2005.12.013

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* 326, 289–293. doi:10.1126/science.1181369

Lim, B., Heist, T., Levine, M., and Fukaya, T. (2018). Visualization of Transvection in Living Drosophila Embryos. *Mol. Cell.* 70, 287–296.e6. doi:10.1016/j.molcel. 2018.02.029

Loubiere, V., Papadopoulos, G. L., Szabo, Q., Martinez, A.-M., and Cavalli, G. (2020). Widespread Activation of Developmental Gene Expression Characterized by PRC1-dependent Chromatin Looping. *Sci. Adv.* 6, eaax4001. doi:10.1126/sciadv.aax4001

Louwers, M., Splinter, E., van Driel, R., de Laat, W., and Stam, M. (2009). Studying Physical Chromatin Interactions in Plants Using Chromosome Conformation Capture (3C). *Nat. Protoc.* 4, 1216–1229. doi:10.1038/nprot.2009.113

Lozano, E., Sáez, A. G., Flemming, A. J., Cunha, A., and Leroi, A. M. (2006). Regulation of Growth by Ploidy in *Caenorhabditis elegans*. *Curr. Biol.* 16, 493–498. doi:10.1016/j.cub.2006.01.048

Markstein, M., Pitsouli, C., Villalta, C., Celniker, S. E., and Perrimon, N. (2008). Exploiting Position Effects and the Gypsy Retrovirus Insulator to Engineer Precisely Expressed Transgenes. *Nat. Genet.* 40, 476–483. doi:10.1038/ng.101

Marquez, R. M., Singer, M. A., Takaesu, N. T., Waldrip, W. R., Kraytsberg, Y., and Newfeld, S. J. (2001). Transgenic Analysis of the Smad Family of TGF-β Signal Transducers in *Drosophila melanogaster* Suggests New Roles and New Interactions between Family Members. *Genetics* 157, 1639–1648. doi:10. 1093/genetics/157.4.1639

Matthews, N. E., and White, R. (2019). Chromatin Architecture in the Fly: Living without CTCF/Cohesin Loop Extrusion? *BioEssays* 41, 1900048. doi:10.1002/ bies.201900048

Mifsud, B., Tavares-Cadete, F., Young, A. N., Sugar, R., Schoenfelder, S., Ferreira, L., et al. (2015). Mapping Long-Range Promoter Contacts in Human Cells with High-Resolution Capture Hi-C. *Nat. Genet.* 47, 598–606. doi:10.1038/ng.3286

Mitchell, J. A., and Fraser, P. (2008). Transcription Factories Are Nuclear Subcompartments that Remain in the Absence of Transcription. *Genes. Dev.* 22, 20–25. doi:10.1101/gad.454008

Muchardt, C., and Yaniv, M. (1993). A Human Homologue of *Saccharomyces cerevisiae* SNF2/SWI2 and Drosophila Brm Genes Potentiates Transcriptional Activation by the Glucocorticoid Receptor. *EMBO J.* 12, 4279–4290. doi:10. 1002/j.1460-2075.1993.tb06112.x

Nagano, T., Lubling, Y., Stevens, T. J., Schoenfelder, S., Yaffe, E., Dean, W., et al. (2013). Single-cell Hi-C Reveals Cell-To-Cell Variability in Chromosome Structure. *Nature* 502, 59–64. doi:10.1038/nature12593

Nagano, T., Lubling, Y., Várnai, C., Dudley, C., Leung, W., Baran, Y., et al. (2017). Cell-cycle Dynamics of Chromosomal Organization at Single-Cell Resolution. *Nature* 547, 61–67. doi:10.1038/nature23001

Nagano, T., Lubling, Y., Yaffe, E., Wingett, S. W., Dean, W., Tanay, A., et al. (2015a). Single-cell Hi-C for Genome-wide Detection of Chromatin Interactions that Occur Simultaneously in a Single Cell. *Nat. Protoc.* 10, 1986–2003. doi:10.1038/nprot.2015.127

Nagano, T., Várnai, C., Schoenfelder, S., Javierre, B.-M., Wingett, S. W., and Fraser, P. (2015b). Comparison of Hi-C Results Using In-Solution versus In-Nucleus Ligation. *Genome Biol.* 16, 175. doi:10.1186/s13059-015-0753-7

Nakayama, R. T., Pulice, J. L., Valencia, A. M., McBride, M. J., McKenzie, Z. M., Gillespie, M. A., et al. (2017). SMARCB1 Is Required for Widespread BAF Complex-Mediated Activation of Enhancers and Bivalent Promoters. *Nat. Genet.* 49, 1613–1623. doi:10.1038/ng.3958

Neal, S. J., Dolezal, D., Jušić, N., and Pignoni, F. (2019). Drosophila ML-DmD17-C3 Cells Respond Robustly to Dpp and Exhibit Complex Transcriptional Feedback on BMP Signaling Components. *BMC Dev. Biol.* 19. doi:10.1186/s12861-019-0181-0

Ni, Z., Hassan, M. A. E., Xu, Z., Yu, T., and Bremner, R. (2008). The Chromatin-Remodeling Enzyme BRG1 Coordinates CIITA Induction through Many Interdependent Distal Enhancers. *Nat. Immunol.* 9, 785–793. doi:10.1038/ni.1619

Novo, C. L., Javierre, B.-M., Cairns, J., Segonds-Pichon, A., Wingett, S. W., Freire-Pritchett, P., et al. (2018). Long-Range Enhancer Interactions Are Prevalent in Mouse Embryonic Stem Cells and Are Reorganized upon Pluripotent State Transition. *Cell. Rep.* 22, 2615–2627. doi:10.1016/j.celrep.2018.02.040

Nyström, J., Shen, Z.-Z., Aili, M., Flemming, A. J., Leroi, A., and Tuck, S. (2002). Increased or Decreased Levels of *Caenorhabditis elegans* Lon-3, a Gene Encoding a Collagen, Cause Reciprocal Changes in Body Length. *Genetics* 161, 83–97. doi:10.1093/genetics/161.1.83

Ogiyama, Y., Schuettengruber, B., Papadopoulos, G. L., Chang, J.-M., and Cavalli, G. (2018). Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during Drosophila Development. *Mol. Cell.* 71, 73–88.e5. doi:10.1016/j.molcel.2018.05.032

Osborne, C. S., Chakalova, L., Brown, K. E., Carter, D., Horton, A., Debrand, E., et al. (2004). Active Genes Dynamically Colocalize to Shared Sites of Ongoing Transcription. *Nat. Genet.* 36, 1065–1071. doi:10.1038/ng1423

Papantonis, A., Larkin, J. D., Wada, Y., Ohta, Y., Ihara, S., Kodama, T., et al. (2010). Active RNA Polymerases: Mobile or Immobile Molecular Machines? *PLOS Biol.* 8, e1000419. doi:10.1371/journal.pbio.1000419

Peng, Y., and Zhang, Y. (2018). Enhancer and Super-enhancer: Positive Regulators in Gene Transcription. *Anim. Models Exp. Med.* 1, 169–179. doi:10.1002/ame2.12032

Peterson, A. J., and O'Connor, M. B. (2014). Strategies for Exploring TGF-β Signaling in Drosophila. *Methods* 68, 183–193. doi:10.1016/j.ymeth.2014.03.016

Pfeiffer, B. D., Jenett, A., Hammonds, A. S., Ngo, T.-T. B., Misra, S., Murphy, C., et al. (2008). Tools for Neuroanatomy and Neurogenetics in Drosophila. *Proc. Natl. Acad. Sci. U.S.A.* 105, 9715–9720. doi:10.1073/pnas.0803697105

Pfeiffer, B. D., Ngo, T.-T. B., Hibbard, K. L., Murphy, C., Jenett, A., Truman, J. W., et al. (2010). Refinement of Tools for Targeted Gene Expression in Drosophila. *Genetics* 186, 735–755. doi:10.1534/genetics.110.119917

Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y., and Tyagi, S. (2006). Stochastic mRNA Synthesis in Mammalian Cells. *PLoS Biol.* 4, e309. doi:10.1371/journal.pbio.0040309

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell.* 159, 1665–1680. doi:10.1016/j.cell.2014.11.021

Ren, D., Song, J., Ni, M., Kang, L., and Guo, W. (2020). Regulatory Mechanisms of Cell Polyploidy in Insects. *Front. Cell. Dev. Biol.* 8. doi:10.3389/fcell.2020.00361

Rieder, D., Trajanoski, Z., and McNally, J. G. (2012). Transcription Factories. *Front. Gene.* 3. doi:10.3389/fgene.2012.00221

Rowley, M. J., Nichols, M. H., Lyu, X., Ando-Kuri, M., Rivera, I. S. M., Hermetz, K., et al. (2017). Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol. Cell.* 67, 837–852.e7. doi:10.1016/j.molcel.2017.07.022

modENCODE ConsortiumRoy, S., Roy, S., Ernst, J., Kharchenko, P. V., Kheradpour, P., Negre, N., et al. (2010). Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE. *Science* 330, 1787–1797. doi:10.1126/science.1198374

Rubin, A. J., Barajas, B. C., Furlan-Magaril, M., Lopez-Pajares, V., Mumbach, M. R., Howard, I., et al. (2017). Lineage-specific Dynamic and Pre-established Enhancer-Promoter Contacts Cooperate in Terminal Differentiation. *Nat. Genet.* 49, 1522–1528. doi:10.1038/ng.3935

Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B.-M., et al. (2015a). The Pluripotent Regulatory Circuitry Connecting Promoters to Their Long-Range Interacting Elements. *Genome Res.* 25, 582–597. doi:10.1101/gr.185272.114

Schoenfelder, S., Javierre, B.-M., Furlan-Magaril, M., Wingett, S. W., and Fraser, P. (2018a). Promoter Capture Hi-C: High-Resolution, Genome-wide Profiling of Promoter Interactions. *JoVE*, 57320. doi:10.3791/57320

Schoenfelder, S., Mifsud, B., Senner, C. E., Todd, C. D., Chrysanthou, S., Darbo, E., et al. (2018b). Divergent Wiring of Repressive and Active Chromatin Interactions between Mouse Embryonic and Trophoblast Lineages. *Nat. Commun.* 9, 4189. doi:10.1038/s41467-018-06666-4

Schoenfelder, S., Sugar, R., Dimond, A., Javierre, B.-M., Armstrong, H., Mifsud, B., et al. (2015b). Polycomb Repressive Complex PRC1 Spatially Constrains the Mouse Embryonic Stem Cell Genome. *Nat. Genet.* 47, 1179–1186. doi:10.1038/ng.3393

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., et al. (2012). Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell.* 148, 458–472. doi:10.1016/j.cell.2012.01.010

Sharifkhodaei, Z., and Auld, V. J. (2021). Overexpressed Gliotactin Activates BMP Signaling through Interfering with the Tkv-Dad Association. *Genome* 64, 97–108. doi:10.1139/gen-2020-0026

Shi, J., Whyte, W. A., Zepeda-Mendoza, C. J., Milazzo, J. P., Shen, C., Roe, J.-S., et al. (2013). Role of SWI/SNF in Acute Leukemia Maintenance and Enhancer-Mediated Myc Regulation. *Genes. Dev.* 27, 2648–2662. doi:10.1101/gad.232710.113

Shidlovskii, Y. V., Bylino, O. V., Shaposhnikov, A. V., Kachaev, Z. M., Lebedeva, L. A., Kolesnik, V. V., et al. (2021). Subunits of the PBAP Chromatin Remodeler Are Capable of Mediating Enhancer-Driven Transcription in Drosophila. *Ijms* 22, 2856. doi:10.3390/ijms22062856

Sirinakis, G., Clapier, C. R., Gao, Y., Viswanathan, R., Cairns, B. R., and Zhang, Y. (2011). The RSC Chromatin Remodelling ATPase Translocates DNA with High Force and Small Step Size. *EMBO J.* 30, 2364–2372. doi:10.1038/emboj.2011.141

Song, W., Sharan, R., and Ovcharenko, I. (2019). The First Enhancer in an Enhancer Chain Safeguards Subsequent Enhancer-Promoter Contacts from a Distance. *Genome Biol.* 20, 197. doi:10.1186/s13059-019-1808-y

Splinter, E., de Wit, E., van de Werken, H. J. G., Klous, P., and de Laat, W. (2012). Determining Long-Range Chromatin Interactions for Selected Genomic Sites Using 4C-Seq Technology: from Fixation to Computation. *Methods* 58, 221–230. doi:10.1016/j.ymeth.2012.04.009

Stadhouders, R., Kolovos, P., Brouwer, R., Zuin, J., van den Heuvel, A., Kockx, C., et al. (2013). Multiplexed Chromosome Conformation Capture Sequencing for Rapid Genome-Scale High-Resolution Detection of Long-Range Chromatin Interactions. *Nat. Protoc.* 8, 509–524. doi:10.1038/nprot.2013.018

Stadler, M. R., Haines, J. E., and Eisen, M. B. (2017). Convergence of Topological Domain Boundaries, Insulators, and Polytene Interbands Revealed by High-Resolution Mapping of Chromatin Contacts in the Early *Drosophila melanogaster* Embryo. *eLife* 6, e29550. doi:10.7554/eLife.29550

Stavreva, D. A., Garcia, D. A., Fettweis, G., Gudla, P. R., Zaki, G. F., Soni, V., et al. (2019). Transcriptional Bursting and Co-bursting Regulation by Steroid Hormone Release Pattern and Transcription Factor Mobility. *Mol. Cell.* 75, 1161–1177.e11. doi:10.1016/j.molcel.2019.06.042

Steller, H., and Pirrotta, V. (1985). Expression of the Drosophila White Gene under the Control of the Hsp70 Heat Shock Promoter. *EMBO J.* 4, 3765–3772. doi:10.1002/j.1460-2075.1985.tb04146.x

Steller, H., and Pirrotta, V. (1984). Regulated Expression of Genes Injected into Early Drosophila Embryos. *EMBO J.* 3, 165–173. doi:10.1002/j.1460-2075.1984.tb01778.x

Sun, Q., Perez-Rathke, A., Czajkowsky, D. M., Shao, Z., and Liang, J. (2019). High-Resolution Single-Cell 3D-Models of Chromatin Ensembles during Drosophila Embryogenesis. *Nat. Commun.* 12 (1), 205. bioRxiv, 854257. doi:10.1038/s41467-020-20490-9

Sutherland, H., and Bickmore, W. A. (2009). Transcription Factories: Gene Expression in Unions? *Nat. Rev. Genet.* 10, 457–466. doi:10.1038/nrg2592

Thomas-Claudepierre, A.-S., Schiavo, E., Heyer, V., Fournier, M., Page, A., Robert, I., et al. (2013). The Cohesin Complex Regulates Immunoglobulin Class Switch Recombination. *J. Exp. Med.* 210, 2495–2502. doi:10.1084/jem.20130166

Tolhuis, B., Blom, M., Kerkhoven, R. M., Pagie, L., Teunissen, H., Nieuwland, M., et al. (2011). Interactions Among Polycomb Domains Are Guided by Chromosome Architecture. *PLoS Genet.* 7, e1001343. doi:10.1371/journal.pgen.1001343

Tolhuis, B., Palstra, R.-J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and Interaction between Hypersensitive Sites in the Active β-globin Locus. *Mol. Cell.* 10, 1453–1465. doi:10.1016/S1097-2765(02)00781-5

Treich, I., Cairns, B. R., de los Santos, I., Brewster, E., and Carlson, M. (1995). SNF11, A New Component of the Yeast SNF-SWI Complex That Interacts With a Conserved Region of SNF2. *Mol. Cell. Biol.* 15, 4240–4248.

Treich, I., and Carlson, M. (1997). Interaction of a Swi3 Homolog with Sth1 Provides Evidence for a Swi/Snf-Related Complex with an Essential Function in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 17, 1768–1775. doi:10.1128/MCB.17.4.1768

Trouche, D., Le Chalony, C., Muchardt, C., Yaniv, M., and Kouzarides, T. (1997). RB and Hbrm Cooperate to Repress the Activation Functions of E2F1. *Proc. Natl. Acad. Sci. U.S.A.* 94, 11268–11273. doi:10.1073/pnas.94.21.11268

Tsuneizumi, K., Nakayama, T., Kamoshida, Y., Kornberg, T. B., Christian, J. L., and Tabata, T. (1997). Daughters against Dpp Modulates Dpp Organizing Activity in Drosophila Wing Development. *Nature* 389, 627–631. doi:10.1038/39362

Ulianov, S. V., Khrameeva, E. E., Gavrilov, A. A., Flyamer, I. M., Kos, P., Mikhaleva, E. A., et al. (2016). Active Chromatin and Transcription Play a Key Role in Chromosome Partitioning into Topologically Associating Domains. *Genome Res.* 26, 70–84. doi:10.1101/gr.196006.115

Ulianov, S. V., Zakharova, V. V., Galitsyna, A. A., Kos, P. I., Polovnikov, K. E., Flyamer, I. M., et al. (2021). Order and Stochasticity in the Folding of Individual Drosophila Genomes. *Nat. Commun.* 12, 41. doi:10.1038/s41467-020-20292-z

Upadhyay, A., Moss-Taylor, L., Kim, M.-J., Ghosh, A. C., and O'Connor, M. B. (2017). TGF-β Family Signaling in Drosophila. *Cold Spring Harb. Perspect. Biol.* 9, a022152. doi:10.1101/cshperspect.a022152

van de Werken, H. J. G., de Vree, P. J. P., Splinter, E., Holwerda, S. J. B., Klous, P., de Wit, E., et al. (2012). "4C Technology: Protocols and Data Analysis," in *Methods in Enzymology* (Elsevier), 89–112. doi:10.1016/B978-0-12-391938-0.00004-5

Vermeulen, C., Allahyar, A., Bouwman, B. A. M., Krijger, P. H. L., Verstegen, M. J. A. M., Geeven, G., et al. (2020). Multi-contact 4C: Long-Molecule Sequencing of Complex Proximity Ligation Products to Uncover Local Cooperative and Competitive Chromatin Topologies. *Nat. Protoc.* 15, 364–397. doi:10.1038/s41596-019-0242-7

Vizcaya-Molina, E., Klein, C. C., Serras, F., Mishra, R. K., Guigó, R., and Corominas, M. (2018). Damage-responsive Elements in Drosophila Regeneration. *Genome Res.* 28, 1852–1866. doi:10.1101/gr.233098.117

Vorobyeva, N. E., Soshnikova, N. V., Nikolenko, J. V., Kuzmina, J. L., Nabirochkina, E. N., Georgieva, S. G., et al. (2009). Transcription Coactivator SAYP Combines Chromatin Remodeler Brahma and Transcription Initiation Factor TFIID into a Single Supercomplex. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11049–11054. doi:10.1073/pnas.0901801106

Webber, J. L., Zhang, J., Mitchell-Dick, A., and Rebay, I. (2013). 3D Chromatin Interactions Organize Yan Chromatin Occupancy and Repression at the Even-Skipped Locus. *Genes. Dev.* 27, 2293–2298. doi:10.1101/gad.225789.113

Weber, J. A., Taxman, D. J., Lu, Q., and Gilmour, D. S. (1997). Molecular Architecture of the Hsp70 Promoter after Deletion of the TATA Box or the Upstream Regulation Region. *Mol. Cell. Biol.* 17, 3799–3808. doi:10.1128/MCB.17.7.3799

Weiss, A., Charbonnier, E., Ellertsdóttir, E., Tsirigos, A., Wolf, C., Schuh, R., et al. (2010). A Conserved Activation Element in BMP Signaling during Drosophila Development. *Nat. Struct. Mol. Biol.* 17, 69–76. doi:10.1038/nsmb.1715

Wood, C. D., Veenstra, H., Khasnis, S., Gunnell, A., Webb, H. M., Shannon-Lowe, C., et al. (2016). MYC Activation and BCL2L11 Silencing by a Tumour Virus through the Large-Scale Reconfiguration of Enhancer-Promoter Hubs. *eLife* 5, e18270. doi:10.7554/eLife.18270

Wu, C.-H., Madabusi, L., Nishioka, H., Emanuel, P., Sypes, M., Arkhipova, I., et al. (2001). Analysis of Core Promoter Sequences Located Downstream from the TATA Element in the Hsp70 Promoter from *Drosophila melanogaster*. *Mol. Cell. Biol.* 21, 1593–1602. doi:10.1128/MCB.21.5.1593-1602.2001

Wu, S., Fatkhutdinov, N., Rosin, L., Luppino, J. M., Iwasaki, O., Tanizawa, H., et al. (2019). ARID1A Spatially Partitions Interphase Chromosomes. *Sci. Adv.* 5, eaaw5294. doi:10.1126/sciadv.aaw5294

Zaitseva, M., Vollenhoven, B. J., and Rogers, P. A. W. (2006). *In Vitro* culture Significantly Alters Gene Expression Profiles and Reduces Differences between Myometrial and Fibroid Smooth Muscle Cells. *Mol. Hum. Reprod.* 12, 187–207. doi:10.1093/molehr/gal018

Zhang, G., Huang, H., Liu, D., Cheng, Y., Liu, X., Zhang, W., et al. (2015). N6-Methyladenine DNA Modification in Drosophila. *Cell.* 161, 893–906. doi:10.1016/j.cell.2015.04.018

Zhang, X., Zhang, Y., Ba, Z., Kyritsis, N., Casellas, R., and Alt, F. W. (2019). Fundamental Roles of Chromatin Loop Extrusion in Antibody Class Switching. *Nature* 575, 385–389. doi:10.1038/s41586-019-1723-0

Zhang, Y., Malone, J. H., Powell, S. K., Periwal, V., Spana, E., MacAlpine, D. M., et al. (2010). Expression in Aneuploid Drosophila S2 Cells. *PLoS Biol.* 8, e1000320. doi:10.1371/journal.pbio.1000320

Zhang, Y., Smith, C. L., Saha, A., Grill, S. W., Mihardja, S., Smith, S. B., et al. (2006). DNA Translocation and Loop Formation Mechanism of Chromatin Remodeling by SWI/SNF and RSC. *Mol. Cell.* 24, 559–568. doi:10.1016/j.molcel.2006.10.025

Zielke, N., Edgar, B. A., and DePamphilis, M. L. (2013). Endoreplication. *Cold Spring Harb. Perspect. Biol.* 5, a012948. doi:10.1101/cshperspect.a012948

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership