



Deciphering the Role of 3D Genome Organization in Breast Cancer Susceptibility

Brittany Baur^{1†}, Da-Inn Lee^{1†}, Jill Haag², Deborah Chasman¹, Michael Gould^{2‡} and Sushmita Roy^{1,3*}

¹Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI, United States, ²McArdle Laboratory for Cancer Research, University of Wisconsin-Madison, Madison, WI, United States, ³Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, United States

OPEN ACCESS

Edited by:

Veniamin Fishman,
Institute of Cytology and Genetics
(RAS), Russia

Reviewed by:

Zhaohui Steve Qin,
Emory University, United States
Audrey Qiuyan Fu,
University of Idaho, United States

*Correspondence:

Sushmita Roy
sroy@biostat.wisc.edu

[†]These authors have contributed
equally to this work and share first
authorship

[‡]This paper is dedicated in memory of
Prof. Michael Gould for his passion for
basic research and its impact on
cancer early detection and treatment

Specialty section:

This article was submitted to
Epigenomics and Epigenetics,
a section of the journal
Frontiers in Genetics

Received: 02 October 2021

Accepted: 21 December 2021

Published: 11 January 2022

Citation:

Baur B,
Lee D-I Haag J, Chasman D, Gould M
and Roy S (2022) Deciphering the Role
of 3D Genome Organization in Breast
Cancer Susceptibility.
Front. Genet. 12:788318.
doi: 10.3389/fgene.2021.788318

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 environmental 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:

$$\min_{U^{(t)}, V^{(t)}, V^{(c)}} \sum_{t=1}^T \|X^{(t)} - U^{(t)}V^{(t)T}\|_F^2 + \alpha \|V^{(t)} - V^{(c)}\|_F^2$$

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, $c_i^{(t)} = \operatorname{argmax}_{j \in \{1, \dots, k\}} U^{(t)}[i, j]$, 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).

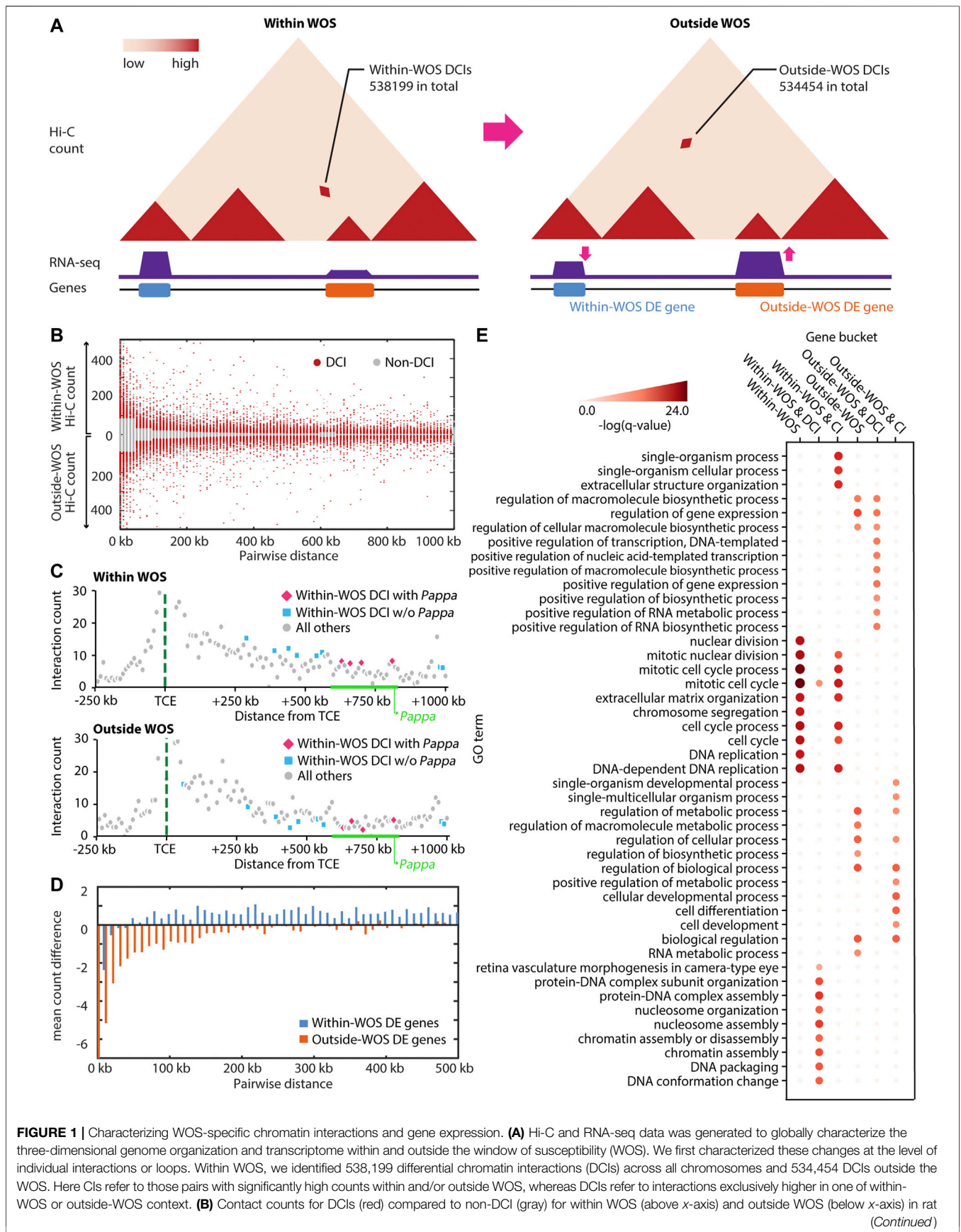


FIGURE 1 | Characterizing WOS-specific chromatin interactions and gene expression. **(A)** Hi-C and RNA-seq data was generated to globally characterize the three-dimensional genome organization and transcriptome within and outside the window of susceptibility (WOS). We first characterized these changes 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)

FIGURE 1 | chromosome 1. **(C)** Visualization of contact counts a -250 kb to +1 Mb around the temporal control element (TCE, green dotted line) within WOS (**top**) and outside WOS (**bottom**). We plot the interaction count between the TCE region and neighboring regions by distance. Blue and pink dots are DCIs that are higher within WOS. Gray dots are all others. Pink dots are additionally associated with the *Pappa* gene. **(D)** Mean count difference (within-WOS count - outside-WOS count) for Fit-Hi-C significant interactions associated with genes that are up-regulated within WOS (i.e., within-WOS DE genes, blue) and up-regulated outside WOS (outside-WOS DE genes, orange). **(E)** Gene Ontology (GO) enrichment of DE genes within/outside WOS, those associated with DCIs, and those associated with significant chromatin interactions (CI). Intensity of red is associated with the $-\log$ (q-value) of the GO enrichment.

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

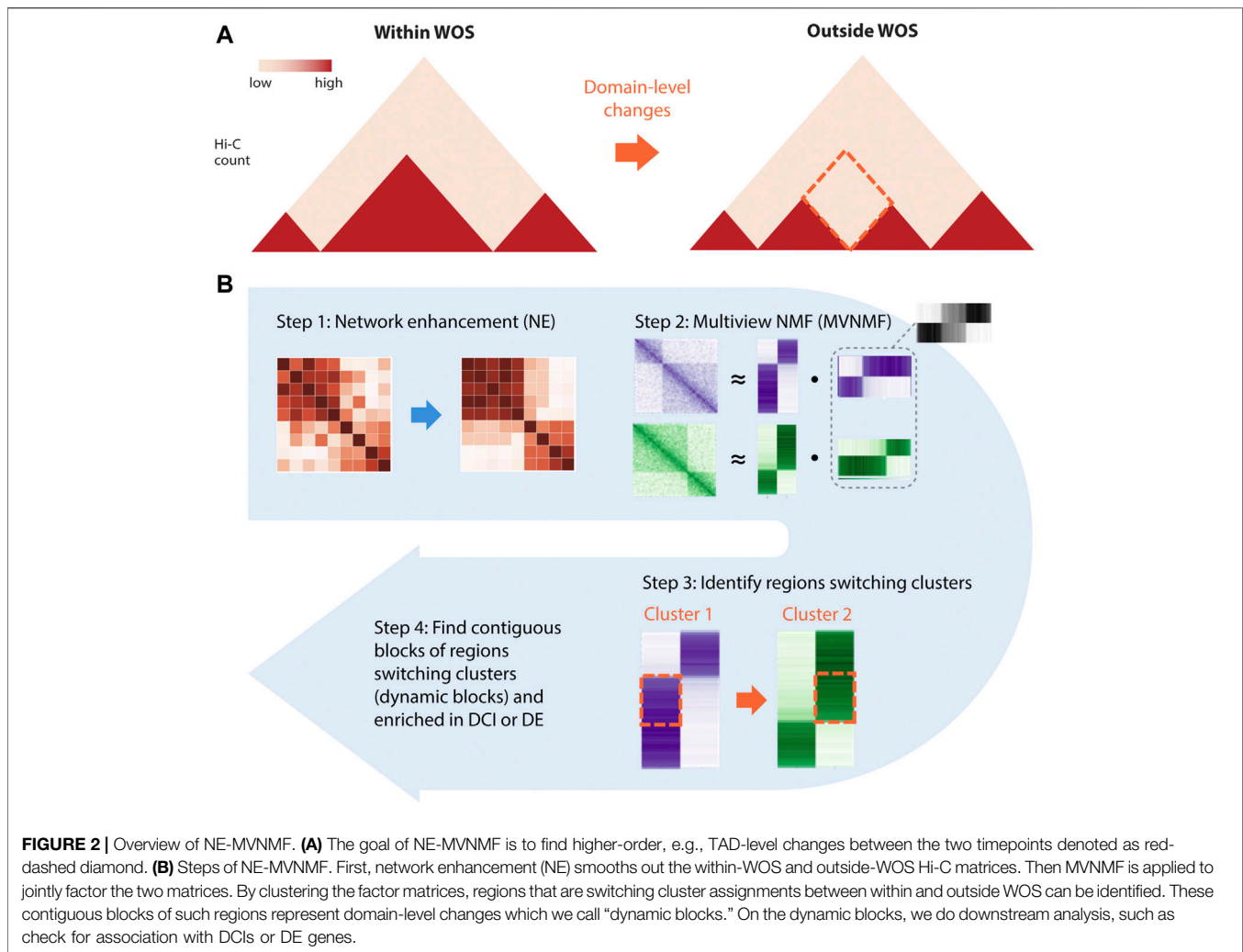
We first identified differential chromatin interactions (DCIs) between the WOS and outside the WOS by taking the union of results from two approaches: Selfish (Ardakany et al., 2019) and the difference in significant chromatin interactions (CIs) identified by applying Fit-Hi-C individually to each sample (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

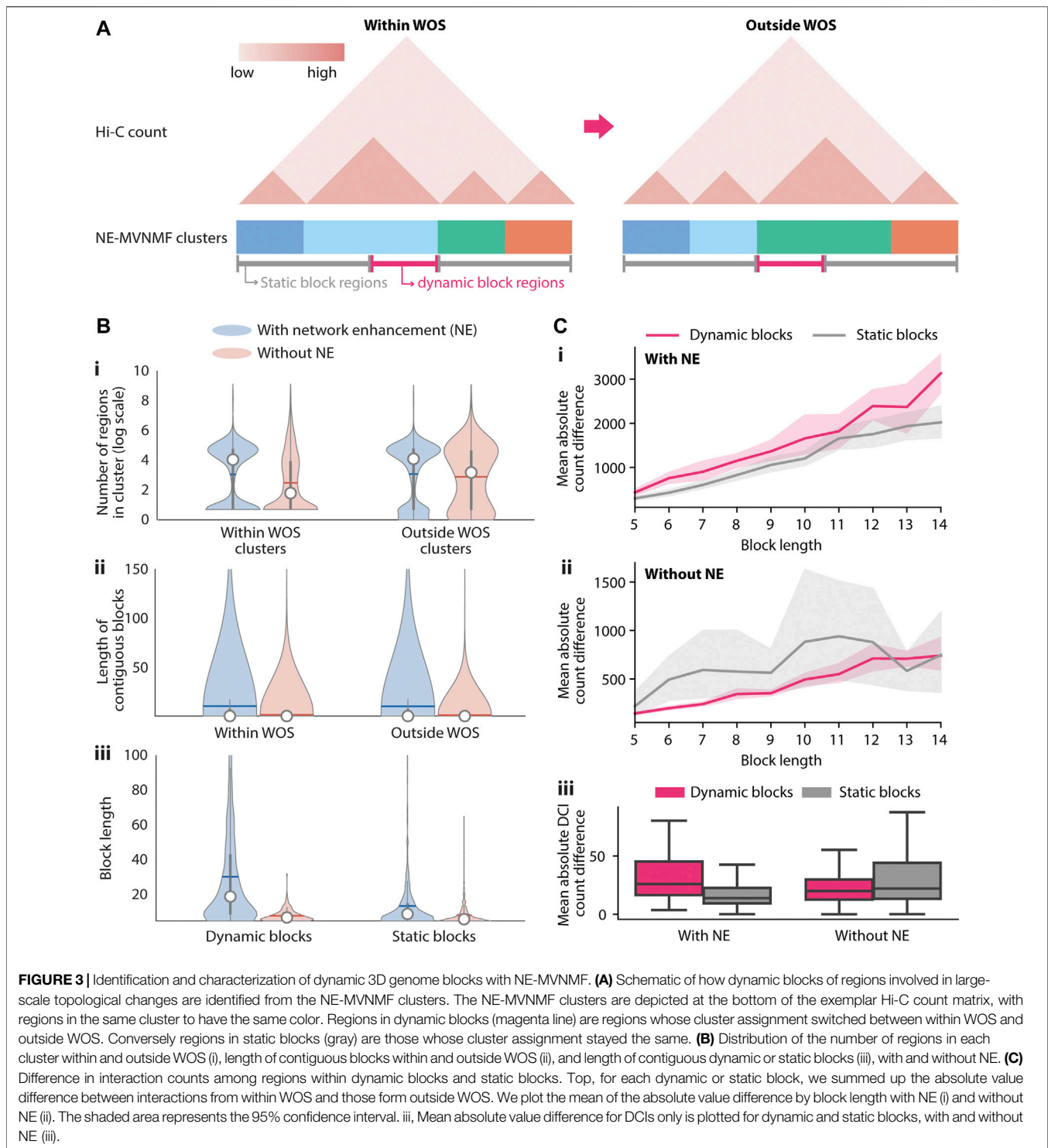


between them can be pinpointed. To address these two challenges, we developed and applied an approach, Network Enhancement with Multi-view Non-negative Matrix Factorization (NE-MVNMf), which first applies network enhancement (NE) to smooth the noisy and sparse Hi-C matrices (Wang et al., 2018) and then 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



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

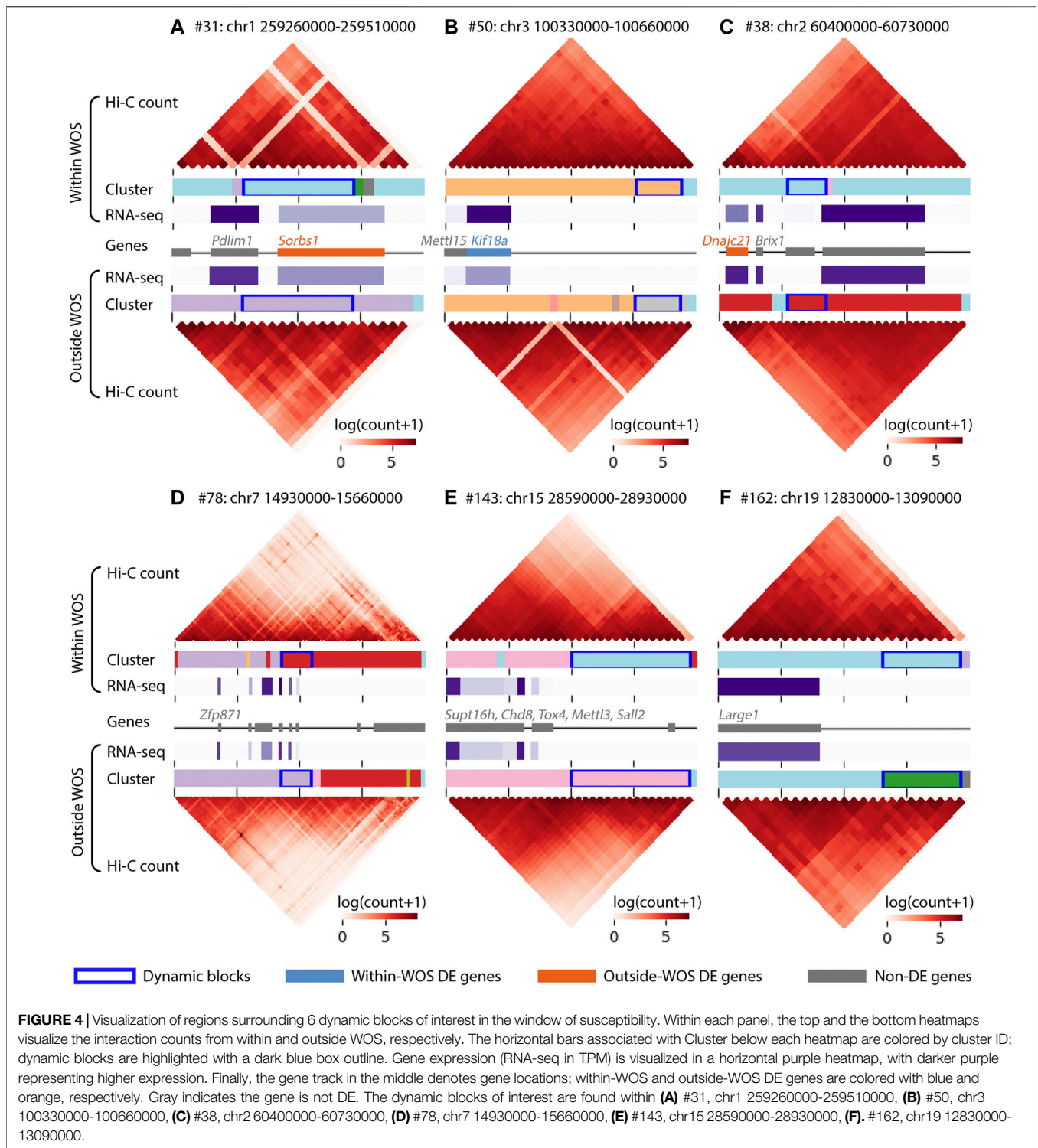
We next examined the dynamic blocks obtained from NE-MVNMF to gain insight into the 3D genome organizational properties within and outside the WOS. There were 168 dynamic blocks in total across all chromosomes, which ranged in size from 50 to 320 kb. We prioritized blocks for interpretation based on 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 (*Large1*, 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



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.,

TABLE 2 | Number of genic and non-genic dynamic blocks. A genic block is one which has genes.

	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

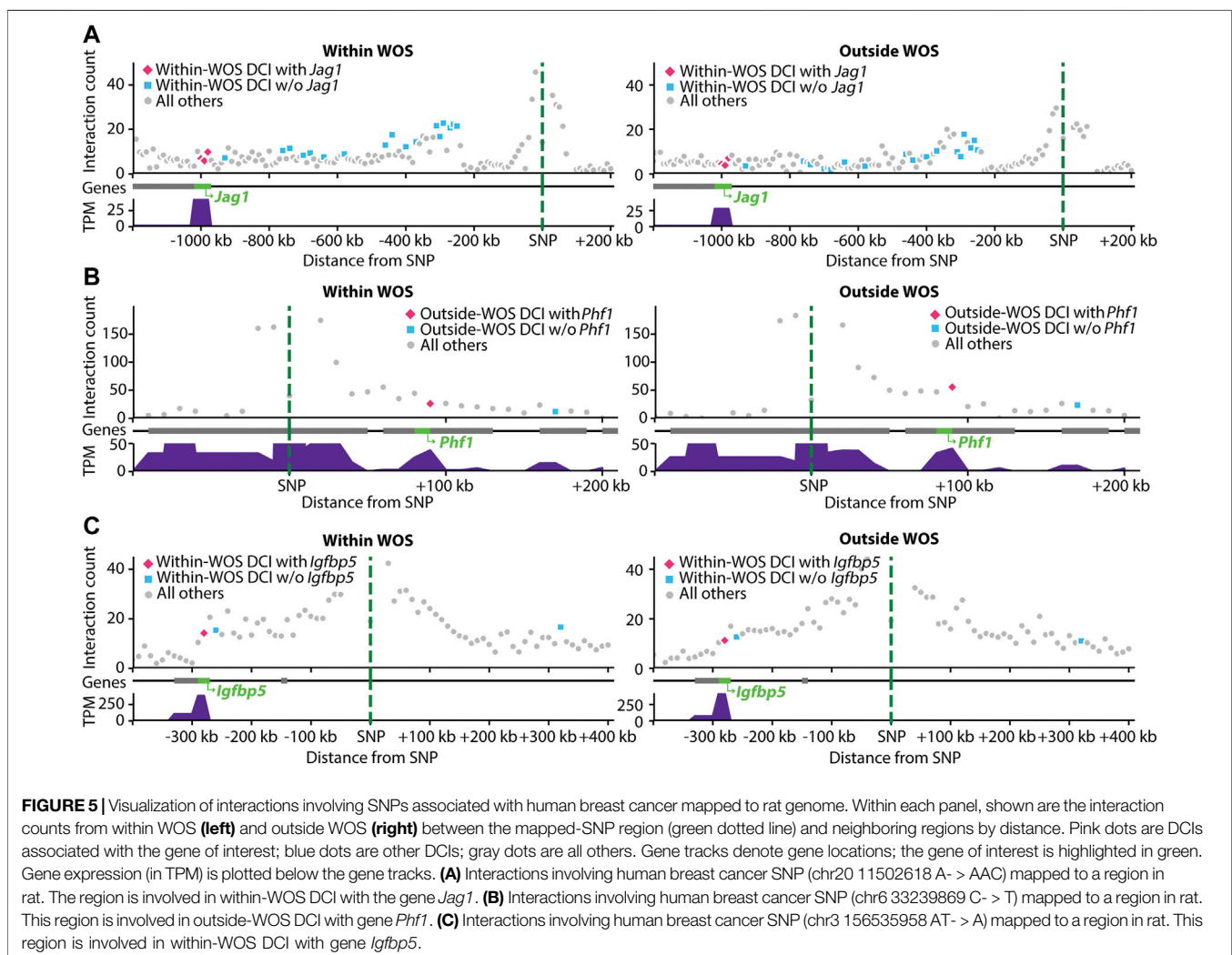


TABLE 3 | Rat loci-gene interactions that recapitulate human breast cancer GWAS SNP-gene interaction from Zhang et al. (2020). Shown is the name of the human SNP, the rat 10 kb bin that has the gene and the variant of the conserved loop. The human SNP is named to show the chromosome ID and the genomic coordinate.

Gene name	Human SNP	Chr	GeneContainingBin	SNPContainingBin	Within or outside WOS
<i>Hnf1a</i>	rs17215231	12	47430000	47420000	Within
<i>Trps1</i>	rs13277568	7	90150000	90310000	Within
<i>Jag1</i>	rs141526427	3	130090000	131090000	Within
<i>Jag1</i>	rs141526427	3	130100000	131090000	Within
<i>Jag1</i>	rs141526427	3	130110000	131090000	Within
<i>Lekr1</i>	rs34052812	2	157570000	157440000	Within
<i>Mrps35</i>	chr12_29140260	4	181430000	182340000	Within
<i>Phf1</i>	rs2464195	20	5530000	5440000	Outside
<i>Cuta</i>	rs2464195	20	5530000	5440000	Outside
<i>Spp13</i>	rs17215231	12	47310000	47420000	Outside
<i>Hnf1a</i>	rs17215231	12	47410000	47420000	Outside
<i>Trps1</i>	rs13277568	7	90190000	90310000	Outside
<i>Trps1</i>	rs13277568	7	90200000	90310000	Outside
<i>Trps1</i>	rs13277568	7	90300000	90310000	Outside
<i>Tiparp</i>	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
<i>Igfbp5</i>	rs13387042	9	80160000	80440000	Within
<i>Zmiz1</i>	rs704010	16	1890000	1760000	Outside
<i>Zmiz1</i>	rs704010	16	1930000	1760000	Outside
<i>Gtpbp3</i>	rs8170	16	19900000	19790000	Outside
<i>Rpl37a</i>	rs16857609	9	80000000	80810000	Outside
<i>Olfml3, Hipk1</i>	rs11552449	2	206220000	206290000	Outside
<i>Ovol1</i>	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

This work was supported by an NHGRI training grant to the Genomics Sciences Training Program at UW-Madison (NHGRI 5T32HG002760) for BB, NHGRI R01 grant R01-HG010045-01 for SR, D-IL and BB, an NLM training grant to the Computation and Informatics in Biology and Medicine Training Program at UW-Madison (NLM 5T15LM007359) for D-IL. This work was also supported by the UW Vilas Foundation.

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., Versteegen, 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/emj.201796717

ACKNOWLEDGMENTS

We would like to thank the Center for High-Throughput Computing (CHTC) and the Bioinformatics Resource Center (BRC) at UW-Madison for providing resources to complete this project. We would also like to thank Sean Mcilwain and Irene Ong for performing the initial analysis of the RNA-seq data.

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>

- Ghousaini, 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., Ragozy, 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/acsmchemlett.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, L., 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

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, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Baur, Lee, Haag, Chasman, Gould and Roy. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.