4See: A Flexible Browser to Explore 4C Data

Yousra Ben Zouari [1,2,3,4], Angeliki Platania [1,2,3,4], Anne M. Molitor [1,2,3,4] and Tom Sexton [1,2,3,4]*

[1] Institute of Genetics and Molecular and Cellular Biology (IGBMC), Illkirch, France, [2] CNRS UMR7104, Illkirch, France, [3] INSERM U1258, Illkirch, France, [4] University of Strasbourg, Illkirch, France

It is established that transcription of many metazoan genes is regulated by distal regulatory sequences beyond the promoter. Enhancers have been identified at up to megabase distances from their regulated genes, and/or proximal to or within the introns of unregulated genes. The unambiguous identification of the target genes of newly identified regulatory elements can thus be challenging. Well-studied enhancers have been found to come into direct physical proximity with regulated genes, presumably by the formation of chromatin loops. Chromosome conformation capture (3C) derivatives that assess the frequency of proximity between different genetic elements is thus a popular method for exploring gene regulation by distal regulatory elements. For studies of chromatin loops and promoter-enhancer communication, 4C (circular chromosome conformation capture) is one of the methods of choice, optimizing cost (required sequencing depth), throughput, and resolution. For ease of visual inspection of 4C data we present 4See, a versatile and user-friendly browser. 4See allows 4C profiles from the same bait to be flexibly plotted together, allowing biological replicates to either be compared, or pooled for comparisons between different cell types or experimental conditions. 4C profiles can be integrated with gene tracks, linear epigenomic profiles, and annotated regions of interest, such as called significant interactions, allowing rapid data exploration with limited computational resources or bioinformatics expertise.

Keywords: 4C, epigenomics, browser, chromatin loops, quantile normalization, biological replicates

INTRODUCTION

Since early transgenic studies it has been clear that promoter sequences are insufficient to regulate the spatiotemporal expression patterns of many developmental genes. "Remote control" is additionally conferred by distal activating sequences, termed enhancers, which have been intensively studied over the last years (Schoenfelder and Fraser, 2019). Genome-wide profiling of histone modifications and protein binding sites by ChIP-seq have uncovered a general chromatin signature of enhancer regions: DNase-hypersensitive, bound by the transcriptional coactivator p300, and marked by the monomethylation of lysine-4 of histone H3 (H3K4me1) (Heintzman et al., 2009). Follow-on studies refined these findings further by identifying chromatin features that were characteristic of different enhancer properties. For example, the strongest-acting enhancers are also accompanied by acetylation of lysine-27 of histone H3 (H3K27ac) (Creyghton et al., 2010;
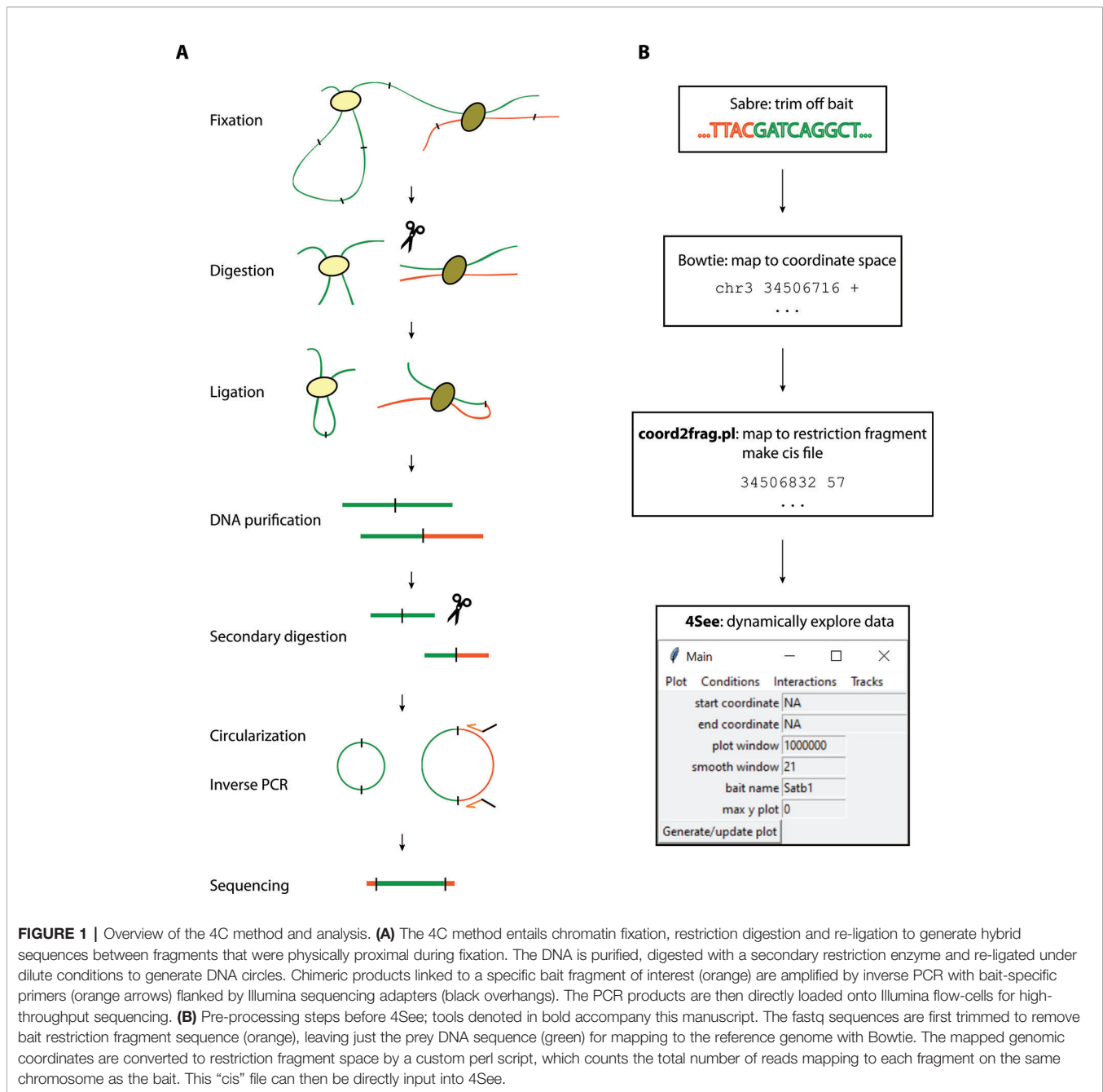
Rada-Iglesias et al., 2011) and/or acetylation on globular histone domains (Pradeepa et al., 2016), recruit RNA polymerase II, and general transcriptional machinery (Koch et al., 2011), and are even transcriptionally active, producing non-coding RNA (eRNAs) (Kim et al., 2010). Enhancers lacking these extra features, and sometimes even encompassing repressive marks, such as H3K27 trimethylation (H3K27me3), are proposed to be "poised" enhancers, which may become activated at later developmental stages. Interestingly, the chromatin states at enhancer sequences vary much more across cell types than those of gene promoters (Roadmap Epigenomics et al., 2015), suggesting that much of the regulatory potential is epigenetically carried by enhancers. However despite advances in identifying enhancers genome-wide, both through epigenomic profiling and high-throughput reporter assays (Arnold et al., 2013; Roadmap Epigenomics et al., 2015), unambiguous identification of their target genes is still a major challenge. Important developmental enhancers have been found at megabase distances from target genes, and/or within the introns of unregulated genes (Lettice et al., 2003; Amano et al., 2009; Herranz et al., 2014); previous studies estimate that up to ~90% of enhancers may indeed skip the closest genes on the linear chromosome fiber (Sanyal et al., 2012; Schoenfelder et al., 2015).

Since the advent of the chromosome conformation capture method (3C) (Dekker et al., 2002) and its variants to measure relative spatial proximity of pairwise genomic regions, many enhancers have been found to physically interact with their target genes, often with "looping out" of the intervening chromatin (Palstra et al., 2003); the resultant "active chromatin hub" has been proposed to provide the permissive regulatory environment for transcription initiation, although the exact mechanism remains unclear. In many studied cases, looping is concomitant with transcriptional induction, whereas in others, the loop is pre-formed to poise the gene for subsequent activation (Schoenfelder and Fraser, 2019). Recent reports using microscopy methods have also been made of enhancers and promoters being well separated on gene activation (Alexander et al., 2019; Benabdallah et al., 2019), although enhancer-promoter interactions were previously reported in the studied loci, raising questions as to whether interactions may completely precede transcription and/or be very transient events. In any case, physical proximity measured by 3C-based methods is becoming a popular means of ascribing target genes to otherwise cryptic distal regulatory elements, or of identifying novel candidate regulatory regions of specific genes of interest. For example, intergenic sequence variants associated with diseases have been better characterized once their target genes were identified by 3C-based approaches (Herranz et al., 2014; Schoenfelder and Fraser, 2019).

With the advent of next-generation sequencing, several higher throughput variants of 3C have been developed to obtain genome-wide chromatin interaction maps. Hi-C is an "all-to-all" method, systematically assessing all pairwise chromatin contacts (Lieberman-Aiden et al., 2009). However, due to the great complexity of the sequenced material, calling specific looping interactions requires prohibitively expensive sequencing depth (Rao et al., 2014; Bonev et al., 2017), and Hi-C loop calling

algorithms have been demonstrated to not be very robust (Forcato et al., 2017). A recent modification, Capture Hi-C, incorporates capture with a pool of thousands of oligonucleotides, allowing the complexity of sequenced Hi-C material to be reduced sufficiently to assess the chromatin looping interactions with all promoters (Hughes et al., 2014; Sahlen et al., 2015; Schoenfelder et al., 2015). However, capture libraries can be expensive, and their design still represents a trade-off between coverage of assessed promoters and resolution of the identified loops. For the highest resolution profiling of smaller numbers of candidate regions, the method of choice is the "one-to-all" 4C (circular chromosome conformation capture), which assesses all the chromatin interactions with one specific bait of interest (Simonis et al., 2006; van de Werken et al., 2012) (**Figure 1A**). In brief, nuclei are fixed in their native topologies with formaldehyde, digested with a restriction enzyme and re-ligated, as for 3C, such that chimeric DNA sequences are generated between restriction fragments which may be unlinked on the linear chromosome fiber but are physically proximal at the time of fixation. The purified DNA is then circularized by digestion with a secondary restriction enzyme and re-ligation under dilute conditions, allowing an inverse PCR strategy to amplify all the chimeric DNA linked to a specific bait restriction fragment of interest. The much reduced complexity of a 4C library, compared to that of Hi-C, means that promoter interactomes can be reliably profiled with just a few million sequence reads, and ~20 baits can readily be multiplexed into a sequencing run, making it a much more cost-effective method (van de Werken et al., 2012). The major limitations of 4C are the relatively small throughput in baits that can be assessed at a time, and that the direct sequencing of PCR products confounds results with large numbers of PCR duplicates that cannot be distinguished from counts of true 3C ligation events. However, *in silico* approaches can minimize the impact of PCR duplicates (de Wit et al., 2015), and "unique molecular identifier" variants of 4C have also been developed (Schwartzman et al., 2016).

Due to the growing popularity of 4C experiments, several algorithms have been developed to call significant interactions (van de Werken et al., 2012; Thongjuea et al., 2013; Williams et al., 2014; Klein et al., 2015; Raviram et al., 2016; Geeven et al., 2018); recent benchmarking shows that all methods work well on simulated data, but no single method is optimum for all experimental setups (Walter et al., 2019). However, whereas most of these methods has an in-built tool to plot the static results after data processing, a simple, flexible browser allowing a user to rapidly visualize their 4C results is currently lacking (see **Figure 2** and summary of the different plotting options currently available in **Table 1**). Moreover, while some methods allowed raw and/or smoothed 4C data to be exported as files that can be opened and visualized alongside epigenomic profiles on genome browsers, they offered no flexibility in plotting the epigenomic profiles directly alongside the 4C plot while different smoothing or peak calling parameters are being trialed. We recently developed ChiCMaxima, a suite of tools to analyze Capture Hi-C data, which includes a GUI (graphical user interface) to flexibly visualize data sets alongside gene annotations and epigenomic profiles (Ben Zouari et al., 2019). Here we report

FIGURE 1 | Overview of the 4C method and analysis. **(A)** The 4C method entails chromatin fixation, restriction digestion and re-ligation to generate hybrid sequences between fragments that were physically proximal during fixation. The DNA is purified, digested with a secondary restriction enzyme and re-ligated under dilute conditions to generate DNA circles. Chimeric products linked to a specific bait fragment of interest (orange) are amplified by inverse PCR with bait-specific primers (orange arrows) flanked by Illumina sequencing adapters (black overhangs). The PCR products are then directly loaded onto Illumina flow-cells for high-throughput sequencing. **(B)** Pre-processing steps before 4See; tools denoted in bold accompany this manuscript. The fastq sequences are first trimmed to remove bait restriction fragment sequence (orange), leaving just the prey DNA sequence (green) for mapping to the reference genome with Bowtie. The mapped genomic coordinates are converted to restriction fragment space by a custom perl script, which counts the total number of reads mapping to each fragment on the same chromosome as the bait. This "cis" file can then be directly input into 4See.

4See, the adaptation of ChiCMaxima tools for the user-friendly integrative exploration of 4C data sets. 4See provides flexibility to compare different replicates side by side, or to average them together when comparing experimental conditions, and to visualize 4C profiles at different smoothing window sizes, necessary to identify putative interactions at different distances from the bait, without the need to reload or re-process the initial data. 4See utilizes quantile normalization to allow different plotted profiles to be fairly compared during the visualization. 4See also allows 4C profiles to be easily plotted alongside gene annotations and linear epigenomic tracks, as well as for specific regions (e.g. interactions called by other algorithms) to be

highlighted. We anticipate that 4See will be a useful tool to the community for quick and easy exploration of 4C data, particularly when used in conjunction with existing interaction calling tools.

MATERIALS AND METHODS

Pre-Processing
4C
J1 mouse embryonic stem (ES) cells were grown on gamma-irradiated mouse embryonic fibroblast cells under standard
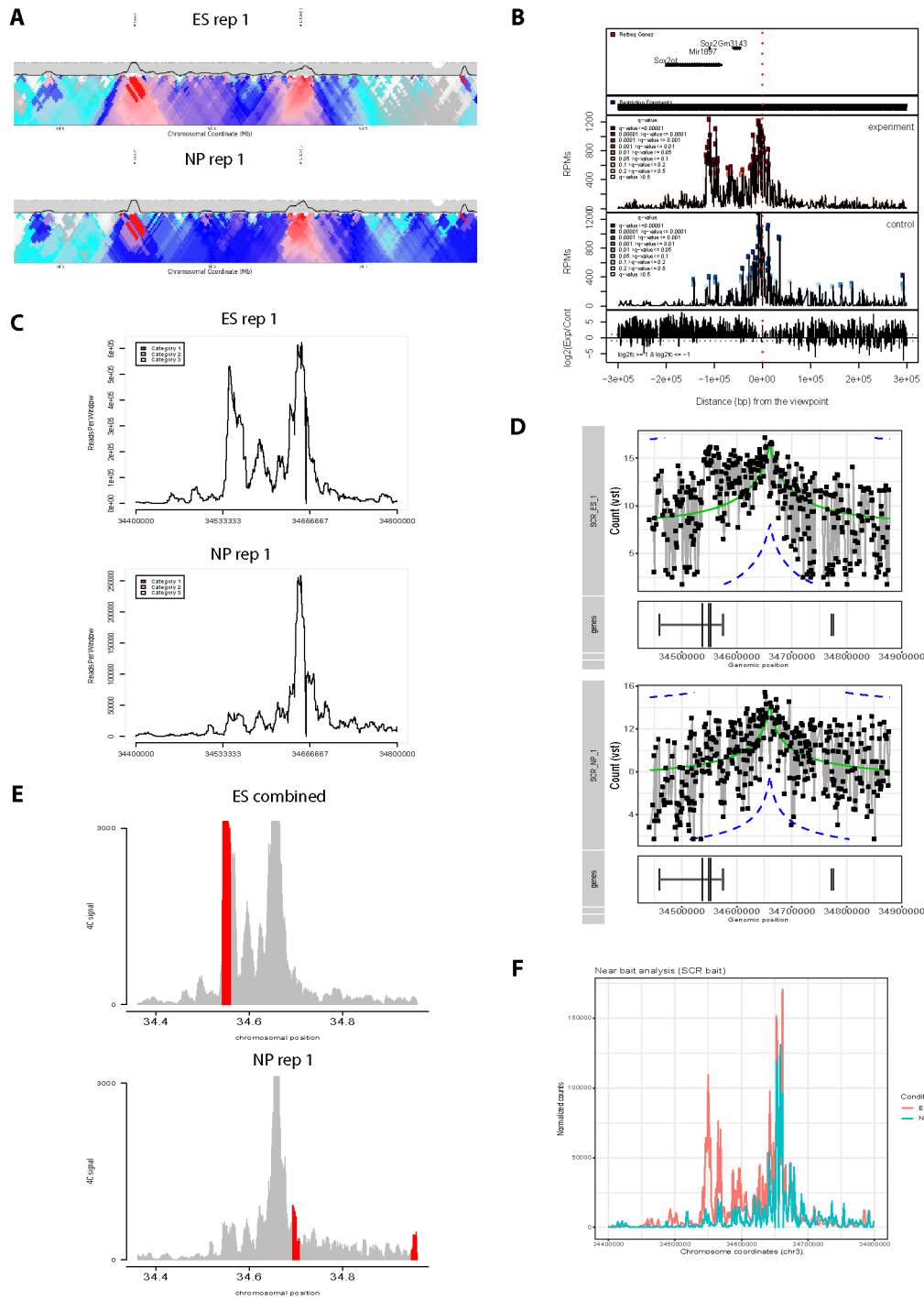
FIGURE 2 | Overview of graphing options from existing 4C analysis methods. All methods have been applied to two ES replicates and one NPC 4C data set for the *Sox2* SCR bait (see also **Figure 3**). **(A)** 4Cseqpipe (van de Werken et al., 2012) results shown independently for one ES replicate and the NPC data set. Running median scores are plotted as a line graph (5 kb resolution), with domainograms plotted underneath as a heat map for median scores at steadily increasing window sizes. Positions of the CTCF site within the SCR and the *Sox2* gene are indicated by arrows. Note that the independent normalization means that the SCR-*Sox2* interaction differences between the two cell types is not evident, compared to other methods. **(B)** r3Cseq (Thongjuea et al., 2013) results for the combined three data sets, showing panels, from top to bottom: positions of Refseq genes; restriction fragment coverage; averaged profile for the "experiment" (ES) condition, with called interactions at different confidence levels highlighted; profile for the "control" (NPC) condition, with called interactions at different confidence levels highlighted; plot of the log2-ratio of experiment vs control 4C signal. The position of the SCR bait is indicated by a red dashed line. Note that r3Cseq appears to call a very large number of

(continued)

conditions (4.5 g/L glucose-DMEN, 15% FCS, 0.1 mM non-essential amino acids, 0.1 mM beta-mercaptoethanol, 1 mM glutamine, 500 U/mL LIF, gentamicin), then passaged onto feeder-free 0.2% gelatin-coated plates for at least two passages to remove feeder cells. For *in vitro* differentiation to neural precursor cells (NPCs), F1 ES cells were cultured in the same medium supplemented with 1 μM PD03259010 and 3 μM CHIR99021 ("2i" conditions) and without feeders. The cells were then cultured for six days with medium without LIF or 2i and with 10% FCS, and with 5 μM retinoic acid for the final four days, to generate embryoid bodies (Bibel et al., 2007). J1/F1 ES or differentiated cells were detached with trypsin, then washed by centrifugation in PBS before fixation. Mouse CD4$^+$ CD8$^+$ double-positive (DP) thymocytes were obtained from 4 week old mouse thymus by FACS with anti-CD4-PE and anti-CD8a-FITC antibodies (eBioScience). Both cell preparations were fixed with 2% formaldehyde in mES culture medium for 10 min at 23°C. The fixation was quenched with cold glycine at a final concentration of 125 mM, then cells were washed with PBS and permeabilized on ice for 1 h with 10 mM Tris-HCl, pH 8, 100 mM NaCl, 0.1% NP-40, and protease inhibitors. Nuclei were resuspended in *Dpn*II restriction buffer at 10 million nuclei/mL concentration, and 5 million nuclei aliquots were further permeabilized by treatment for either 1 h with 0.4% SDS at 37°C (ES cells), or for 20 min with 0.7% SDS at 65°C, then for 40 min at 37°C (DP cells). The SDS was then neutralized by incubating for a further 1 h with either 2.6% (ES) or 3.3% (DP) Triton-X100 at 37°C. Nuclei were digested overnight

with 1000 U *Dpn*II at 37°C, then washed twice by centrifuging and resuspending in T4 DNA ligase buffer. *In situ* ligation was performed in 400 μL T4 DNA ligase buffer with 20,000 U T4 DNA ligase overnight at 16°C. DNA was purified by reverse cross-linking with an overnight incubation at 65°C with proteinase K, followed by RNase A digestion, phenol/chloroform extraction, and isopropanol precipitation. The DNA was digested with 5 U/μg *Csp*6I at 37°C overnight, then re-purified by phenol/chloroform extraction and isopropanol precipitation. The DNA was then circularized by ligation with 200 U/μg T4 DNA ligase under dilute conditions (5 ng/μL DNA), and purified by phenol/ chloroform extraction and isopropanol precipitation. 50 ng aliquots of this DNA were used as template for PCR with bait-specific primers containing Illumina adapter termini (primer sequences and optimal PCR conditions available on request). PCR reactions were pooled, primers removed by washing with 1.8× AMPure XP beads, then quantified on a Bioanalyzer (Agilent) before sequencing with a HiSeq 4000 (Illumina).

Pre-Processing 4C Data for 4See

All bait sequence (including and downstream of the primer sequence, up to but not including the GATC *Dpn*II site) are trimmed by the demultiplexing Sabre tool (https://github.com/najoshi/sabre), allowing two mismatches, before mapping to the mm9 genome with Bowtie (Langmead et al., 2009) (**Figure 1B**). Interaction calling was done with peakC (Geeven et al., 2018) with different window sizes as specified by the parameter wSize.

TABLE 1 | Comparison of graphing options from existing 4C analysis methods.

Method	GUI	Preprocessing	Handles conditions/replicates	Plotting options	Annotations
4Cseqpipe	No	Several custom scripts to convert fastq files to multiple formats and intermediate files. These failed for test data and intermediate files had to be made manually.	No	Can alter trendline resolution and plotting window (coordinate space); domainogram parameters fixed	Limited: manually curated bed file gives arrows on plot
r3Cseq	No	Processes bam files directly	Yes, but restricted to pairwise comparison of "experiment" and "control" conditions	Can only alter plotting window	RefSeq genes
fourSig	No	Custom script converts bam to input format	No	Can only alter plotting window	None
FourCSeq	No	Need to set up metadata table in R, which points to processed bam files	Handled in one combined object, but default is to plot each data set individually, and documentation does not say how to do otherwise.	Can only alter plotting window	Positions of genes from transcriptome (unlabeled)
4C-ker	No	Requires bed file of restriction fragments and bedgraph of 4C coverage per observed fragment. Custom scripts to generate from sam files failed and input files had to be made manually.	Yes	In principle, many settings can be changed in the R command prompt (ggplot2 call settings), but is not documented or user-friendly	None
peakC	No	Essentially the same as this manuscript, but utility scripts not provided	Handles replicates but not different conditions	Can only alter plotting window	None

For previously published 4C results (Narendra et al., 2015), fastq files were downloaded from the Gene Expression Omnibus and processed just like the other data sets.

Analysis and Plotting of 4C Data Sets by Other Methods

Three 4C data sets (two replicates from ES cells; one replicate from *in vitro* differentiation of ES cells towards NPCs) were analyzed and plotted by 4Cseqpipe (van de Werken et al., 2012), r3Cseq (Thongjuea et al., 2013), fourSig (Williams et al., 2014), FourCSeq (Klein et al., 2015), 4C-ker (Raviram et al., 2016), and peakC (Geeven et al., 2018), using the default or recommended parameters given within the documentation accompanying the tools.

4See

System Requirements

4See is a GUI written in R (version > = 3.2), with the following packages (and their dependencies) additionally required, found on Bioconductor and/or CRAN: tcltk2, tkrplot, limma, caTools, rtracklayer. All scripts and test data are available under the terms of the GNU General Public License, version 3, on Github: https://github.com/TomSexton00/4See. The GUI is launched by sourcing the main script, 4See.r, from within an R environment. From then on, all manipulation is performed from a windows interface, and does not require use of command prompts. A full user manual in pdf format is also found with the scripts on Github, and is provided as **Supplementary Data**.

Input Format

4See deals with a simple text format, which we term the "cis" format, entailing a header with three tab-delimited fields (data set name, bait chromosome, bait coordinate) followed by a two-column table, denoting the coordinate of the mid-point of every restriction fragment found on the same chromosome as the bait, and its corresponding number of supporting sequence reads. The cis format is generated by a perl script, coord2frag.pl, provided with 4See, which maps the genomic coordinates of 4C sequencing results into their corresponding restriction fragments and then counts the number of reads for each fragment. The perl script accepts any non-headed text format for sequences, as long as a column for chromosome, coordinate, and strand can be specified. The restriction fragment information is provided by "frag" tables, headed four-column tables, giving a unique integer identifier, the chromosome, coordinate, and fragment length for each restriction fragment. These in turn are generated by a provided perl script, makefrags.pl, requiring a user input for the sequence of the primary restriction enzyme cutting site, and a folder containing the sequences in fasta format for each chromosome of the genome assembly used. The header of the cis file provides the required information on the bait location and 4C data set name, but is also used to ensure that only 4C data sets for the same bait (with identical bait chromosome and coordinate) are treated together.
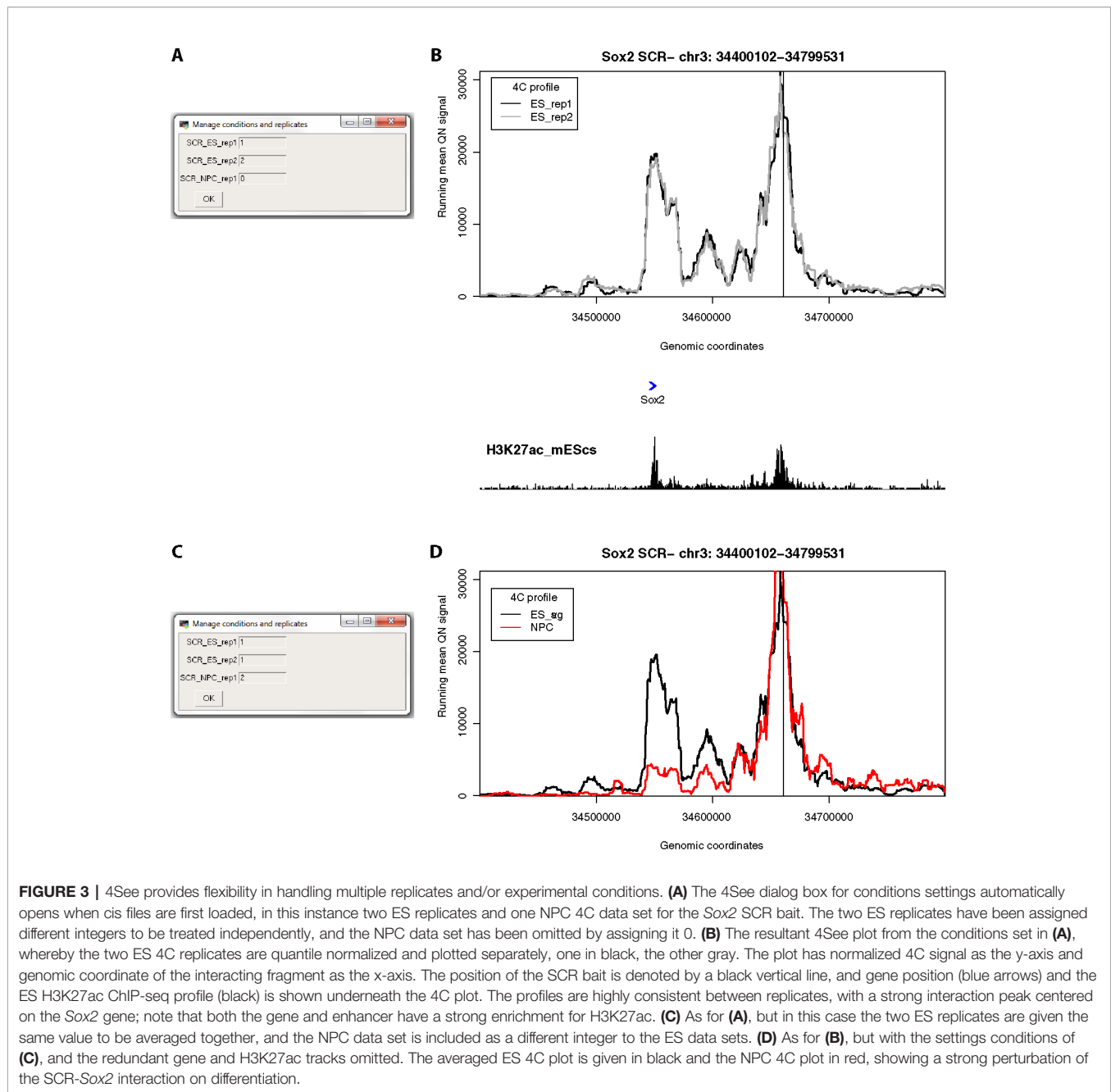
Managing Conditions and Replicates

After loading one or more cis files, 4See opens a dialog box allowing the user to determine how to handle different conditions and replicates by assigning a value to each data set (**Figures 3A, C**). All 4C data sets assigned a non-zero integer are quantile normalized for fairer comparison across data sets (Ritchie et al., 2015). Data sets given the same value are averaged together before plotting; those assigned zero are omitted from normalization and plotting. Additional options allow the plotting color and data label to be specified by the user, and these can be re-run *via* the "Conditions" drop-down menu. Thus a user can rapidly compare different replicates side by side, or average them into one plot for comparison with different cell types or conditions, without needing to reload the data (**Figures 3B, D**).

Plot Settings

4C profiles and the chromatin interactions they uncover differ with bait and experimental condition. In particular, the ease of distinguishing peaks of 4C signal above background depends on the distance of the interaction, since background signal of random chromosome collisions is much higher at shorter ranges (Dekker et al., 2002; Lieberman-Aiden et al., 2009). Other factors, such as whether the interaction is sharp with a single regulatory element, or is broadened across larger regions, such as "super-enhancers", or the extent to which very short-range contacts dominate the plot and hide longer-range loops (which may be a consequence of the 4C digestion efficiency and/or relative compaction of the assessed locus), mean that features of chromatin topology are often overlooked with one fixed plot setting. The control panel of 4See includes options for the user to alter the region plotted, up to ± 1.5 Mb of the bait position, and to set a maximum plotted value on the y-axis (4C signal), to better visualize certain aspects of the 4C profile (**Figure 4A**). However a major confounding factor in visualizing 4C data is the need to smooth the plots, since "spikes" from spurious PCR duplicates make them appear very noisy at single restriction fragment resolution. Most analytical approaches counter this by taking running means (or medians) of sliding windows, but the results can be heavily influenced by the choice of window size. Reflecting this challenge, some 4C analytical tools adopt a "domainogram" approach, whereby averages are taken over many sliding windows of many different sizes, and the results are pooled together in a heat map (de Wit et al., 2008; van de Werken et al., 2012; see also **Figure 2A**), although the visual interpretation of these results is often challenging. To aid user choice in setting appropriate parameters for their particular 4C profile, 4See allows the window size (in numbers of restriction fragments) to be altered, and the appropriate running mean is calculated on the quantile normalized (and averaged, if replicates are pooled) data before plotting. In this manner, different aspects of chromosome topology can be readily explored (**Figures 4B, C**).

Annotations

To put the 4C profiles into a wider biological context, 4See supports the inclusion of three different types of annotations:

FIGURE 3 | 4See provides flexibility in handling multiple replicates and/or experimental conditions. **(A)** The 4See dialog box for conditions settings automatically opens when cis files are first loaded, in this instance two ES replicates and one NPC 4C data set for the *Sox2* SCR bait. The two ES replicates have been assigned different integers to be treated independently, and the NPC data set has been omitted by assigning it 0. **(B)** The resultant 4See plot from the conditions set in **(A)**, whereby the two ES 4C replicates are quantile normalized and plotted separately, one in black, the other gray. The plot has normalized 4C signal as the y-axis and genomic coordinate of the interacting fragment as the x-axis. The position of the SCR bait is denoted by a black vertical line, and gene position (blue arrows) and the ES H3K27ac ChIP-seq profile (black) is shown underneath the 4C plot. The profiles are highly consistent between replicates, with a strong interaction peak centered on the *Sox2* gene; note that both the gene and enhancer have a strong enrichment for H3K27ac. **(C)** As for **(A)**, but in this case the two ES replicates are given the same value to be averaged together, and the NPC data set is included as a different integer to the ES data sets. **(D)** As for **(B)**, but with the settings conditions of **(C)**, and the redundant gene and H3K27ac tracks omitted. The averaged ES 4C plot is given in black and the NPC 4C plot in red, showing a strong perturbation of the SCR-*Sox2* interaction on differentiation.

genes, linear epigenomic profiles (termed "tracks") and called interactions. Gene information is provided as a tab-delimited headed text file with the following fields: Name, Chr (with the prefix "chr"), Start, End, Strand (as "+" or "−"). When selected, the gene track is plotted in blue directly underneath the 4C profile. Only one gene track can be loaded at a time. Management of epigenomic profiles is more flexible. Any format supported by the *import* function of the rtracklayer package can be supported, but for running time efficiency we recommend loading bigWig files. As for the 4C profiles, the color and plotting level for each individual track can be altered by the user in an automatically loaded dialog box. As before, the

plotting levels can be 0 (not plotted), or consecutive, positive integers. When tracks have the same level, their plots are auto-scaled to the maximum value of all of the included data sets within the plotted window. This feature allows fairer comparison for the same epigenetic mark across different conditions/tissue types (**Figure 5**). Technically, the numbers of tracks that can be loaded is only limited by system memory, although the plots become difficult to visually interpret after more than four tracks are loaded at a time.

To better highlight interactions called by existing peak-calling methods, or indeed to test how different methods and/or their parameters perform on specific 4C profiles, interactions (as bed
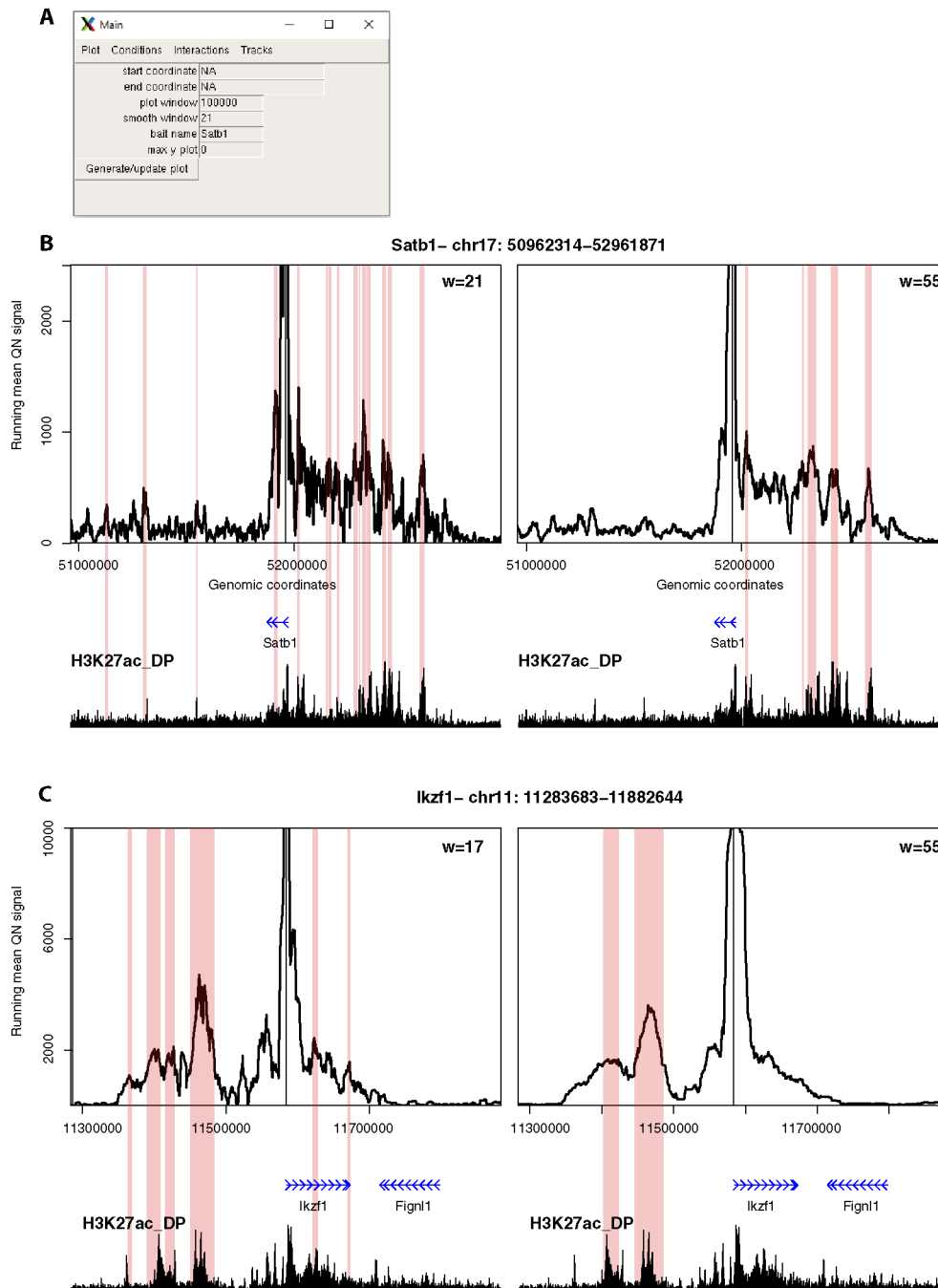
FIGURE 4 | 4See provides flexibility in running mean window sizes. **(A)** The main control panel of 4See, including options to set the x-axis ("start coordinate," "end coordinate," and "plot window") and y-axis ("max y plot") plot limits, to choose a bait name for the plot title ("bait name"), and to set the running mean window size ("smooth window") in number of restriction fragments. **(B)** 4See plots for a 4C data set in DP thymocytes with the *Satb1* gene promoter as bait. Two instances of the same x- and y-axis limits are shown, with a running mean window of 21 (left) or 55 (right) fragments. The position of the *Satb1* bait is denoted by a black vertical line, and gene position (blue arrows) and the DP H3K27ac ChIP-seq profile (black) is shown underneath the 4C plots. Pink rectangles denote regions called as interacting by peakC for the equivalent window size as the plot. For the long-range interaction, the smaller window size appears to have more spurious called interactions, less evidently linked to H3K27ac peaks; the link is better seen with greater smoothing from a larger window size. **(C)** As for **(B)**, but with the *Ikzf1* gene promoter as bait. In this case, the smaller window size (17 fragments) seems to give better resolution of specific interactions with distinct putative enhancers, which are merged into one at larger window sizes.

files or similar, with headed "chr", "start," and "end" columns) can also be loaded, and these are represented as rectangles flanking the relevant region on the 4C profile (**Figures 4B, C**). A dialog box allows the user to alter the color of the annotation and to check whether or not it is plotted. The latter feature is useful since simultaneous plotting of more than one interactions set, which often overlap, can be difficult to visually interpret. Note that whereas these are labeled "Interactions" by 4Sees, any region described by a bed file can be highlighted in this manner. The user can thus use this setting to highlight any feature of interest, such as called differential interactions between two 4C data sets, or the presence of specific sequence motifs (e.g. CTCF) that may be expected to be enriched at interactions.

Exporting 4See Results

Once the user settings have been finalized, a pull-down menu option allows the plot to be saved in.eps format, where it can be further processed in preparation of a figure for publication or presentation. Alternatively, the data that are actually plotted in the current 4See window (one or more quantile-normalized 4C profiles with a running mean of a specified window size applied) can be exported as bedGraph files, ready for integration into other browsers, such as local instances of UCSC (Kent et al., 2002) or IGV (Robinson et al., 2011).

RESULTS

We demonstrate the usefulness of 4See on different original and previously published 4C data sets. First, we investigate the interaction between the mouse *Sox2* gene and an established cluster of enhancers (the "SCR", or *Sox2* control region), which has been shown to be essential for *Sox2* expression in pluripotent cells (Zhou et al., 2014). Using bait primers at the SCR, we generated two biological replicates for ES cells and one after *in vitro* differentiation ("NPC"; **Figure 3**). As expected, we observed a strong interaction with the *Sox2* gene which is greatly reduced on differentiation. After loading the three data sets into the 4See browser, only changing the options within one dialog box is required to switch the view from plotting the two ES biological replicates side by side (omitting the differentiated data set) to confirm that they have consistent profiles, to comparing the averaged ES profile with the differentiated one.

Second, we explored different distance ranges of promoter-enhancer interactions at key developmental genes in mouse CD4$^+$/CD8$^+$ (double-positive, DP) thymocytes, namely the distal (~500 kb) enhancer cluster for *Satb1*, and the shorter-range (~50 kb) enhancer for *Ikzf1* (**Figure 4**). Comparing 4C plots at different running mean window sizes, it is apparent that different insights can be gained, and that no one window size is optimum for all profiles. For *Satb1*, shorter window sizes create what appear to be noisy profiles at the large genomic span assessed, and specific interactions are harder to discern. When the running mean window size is increased, the profile becomes smoother, and apparent peaks line up well with putative enhancers, as denoted by the presence of H3K27ac.

Conversely, at the shorter distances assessed at the *Ikzf1* locus, a smaller window size allows interactions with specific enhancers to be resolved, whereas they merge into one large peak at larger window sizes. In support of this observation, we called interactions using the peakC algorithm (Geeven et al., 2018) at different window sizes, and found a good visual corroboration between discernible peaks and called interactions. 4See allows rapid re-plotting of 4C profiles with different window sizes, and also has the functionalities for adding the epigenomic profile and highlighting called interactions directly on the plot.

Third, we compared the same *Satb1* promoter-enhancer interaction between DP thymocytes, where the gene is highly expressed, and ES cells, where the gene is silent (**Figure 5**). As expected, the gene does not make any specific contacts with the thymocyte enhancer in ES cells. This locus is largely devoid of H3K27ac in ES cells, but a common problem with some browsers is that an automatic scaling creates some apparent peaks from noise on a small range of the y-axis (**Figure 5B**). 4See counters this by providing flexibility with how the epigenomic tracks are handled. By coercing the two tracks to the same scale, the difference between the two tissues is much more evident (**Figure 5C**).

Finally, we used 4See to re-analyze published 4C data, namely comparing profiles from the *Hoxa5* gene between wild-type ES cells and those where one or more key CTCF insulator sites have been deleted (Narendra et al., 2015). In this study, the authors reported that CTCF site loss caused topological defects during differentiation to neurons, with inappropriate spreading of H3K27me3. However, their analyses concluded that the topology of the *Hoxa* locus was largely unchanged in pluripotent cells (**Figure 6A**). Plotting the same data with 4See, it appears that ectopic looping interactions are formed between *Hoxa5* and more caudal regions of the locus (**Figure 6B**). Different loop calling algorithms with different parameter choices were inconsistent in calling these apparent interactions as "significant", and only one biological replicate was available, so the importance of this observation is yet to be confirmed. In any case, the CTCF site deletion did not alter H3K27me3 patterning or *Hoxa* gene expression within undifferentiated ES cells (Narendra et al., 2015), so any potential topological changes do not appear to be borne out in other phenotypes. However, we wish to use this example to highlight how the use of a flexible browser like 4See facilitates exploration of the data, potentially identifying new features that "one size fits all" algorithms may overlook.

DISCUSSION

Using novel and previously published data sets for demonstration, we have shown the flexibility and utility of 4See in exploring 4C data. With limited processing of sequencing results, and one line of code in the R prompt, a user-friendly windows-based interface is available for a broader community to explore chromatin interaction profiles. As a consequence, we envisage that 4See will be of great use to the
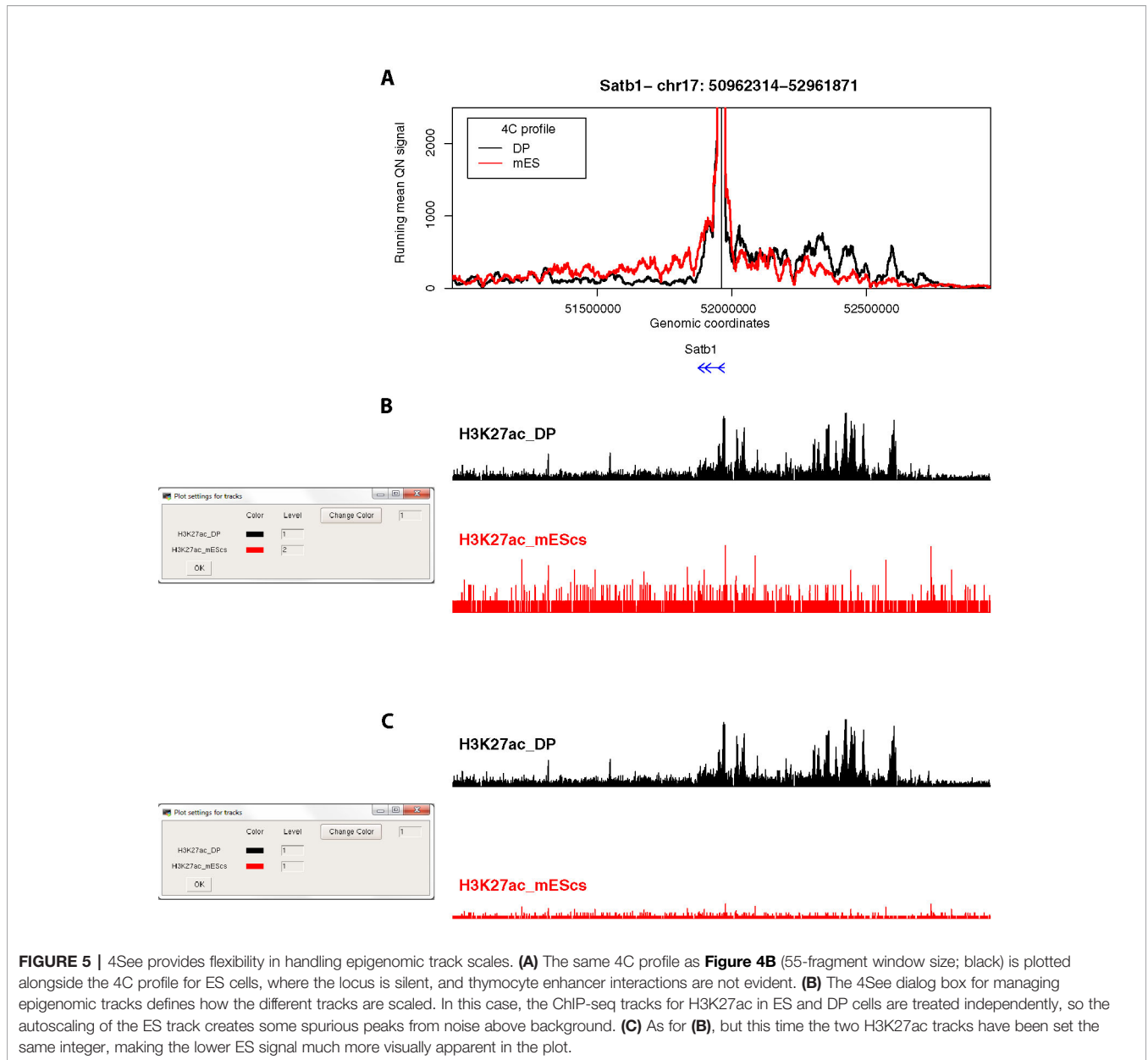
FIGURE 5 | 4See provides flexibility in handling epigenomic track scales. **(A)** The same 4C profile as **Figure 4B** (55-fragment window size; black) is plotted alongside the 4C profile for ES cells, where the locus is silent, and thymocyte enhancer interactions are not evident. **(B)** The 4See dialog box for managing epigenomic tracks defines how the different tracks are scaled. In this case, the ChIP-seq tracks for H3K27ac in ES and DP cells are treated independently, so the autoscaling of the ES track creates some spurious peaks from noise above background. **(C)** As for **(B)**, but this time the two H3K27ac tracks have been set the same integer, making the lower ES signal much more visually apparent in the plot.

chromatin field. The input cis files are not very large (~ 4 MB for mouse or human), so the browser can be run on most desktop computers and laptops. The major systems limitation comes from the importing of epigenomic tracks (which can be >500 MB) with the rtracklayer package, which is the slowest step and may overload some standalone computers if too many tracks are imported at once. If the user is interested in only a specific set of baits, the system load can be reduced by restricting imports to chromosome-specific tracks. Due to the reliance of 4See plotting on quantile normalization, which is confounded by an excessive number of zeros or very small values, 4See is not an appropriate tool for visualizing very long-range (>1.5 Mb) or interchromosomal interactions; although their built-in

graphical capabilities are more limited, the tools linked to algorithms such as fourSig should be used instead (Williams et al., 2014; Walter et al., 2019). It should also be noted that 4See does not replace the existing suite of interaction calling algorithms (Walter et al., 2019). Indeed, due to its capacity to incorporate these algorithms' results into the plots, 4See should be viewed as a complementary tool for comprehensive 4C analysis, whereby the results of the algorithms can be readily visualized and compared to epigenomic tracks for validation and obtaining biological insight. Overall, 4See, in conjunction with other analytical tools, promises to facilitate chromatin interaction exploration, and will thus be of use to the epigenetics community.
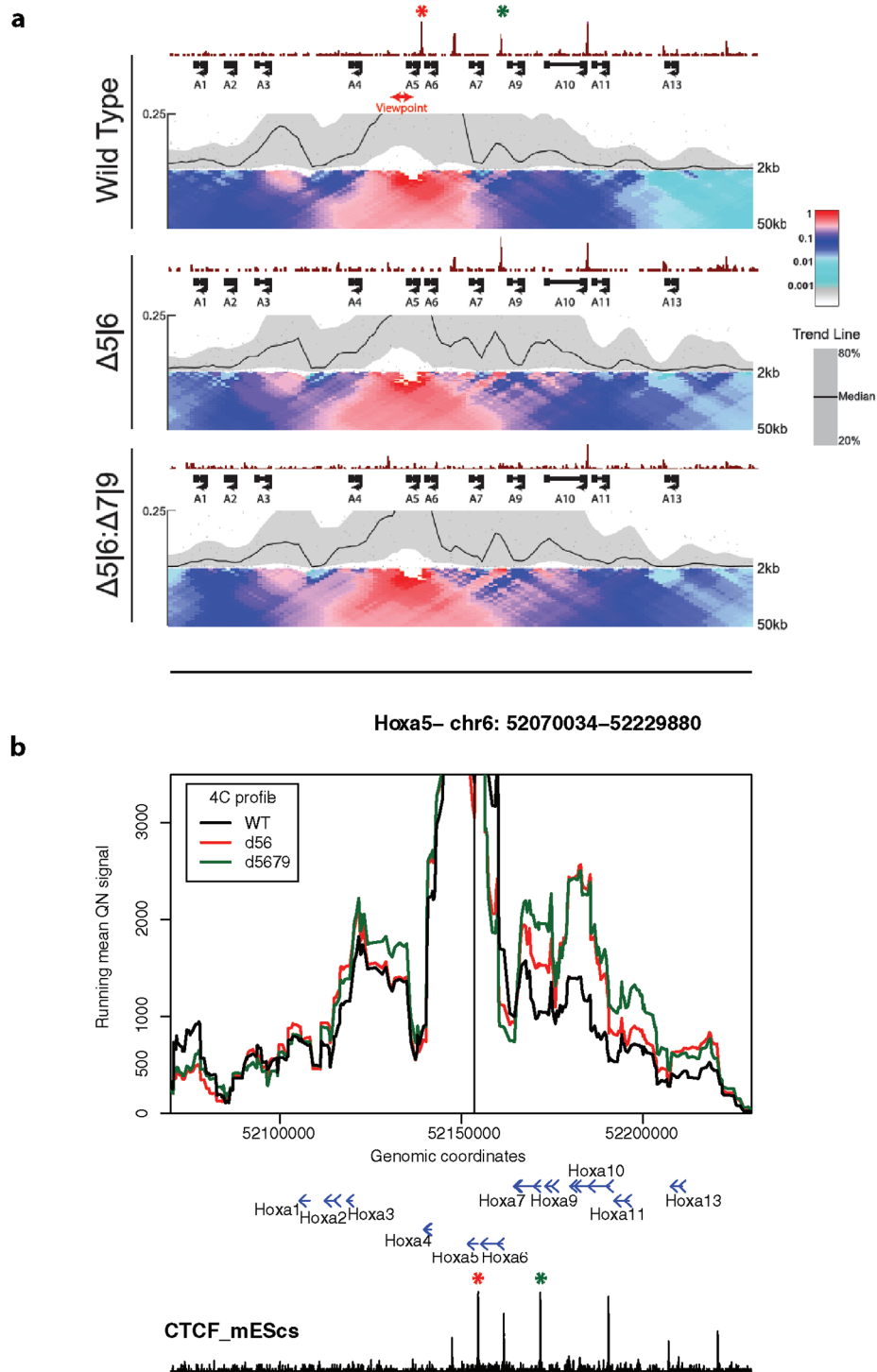
FIGURE 6 | 4See exploration can uncover previously overlooked features of chromatin topology. **(A)** Reproduced Supplemental Fig 7A from Narendra et al. (2015), used with permission. 4C profiles from the *Hoxa5* bait are shown as domainogram heat maps for wild-type ES cells (top), as well as lines that have had deletions of one CTCF site (middle; site of deletion denoted by red asterisk) or two (bottom; additional deletion site denoted by green asterisk). The CTCF ChIP-seq profile is shown above the 4C sets. No chromatin topology differences are apparent between these cell lines, and the original study concluded that spatial phenotypes only occurred on cell differentiation (Narendra et al., 2015). **(B)** The same data, processed and plotted using 4See. The position of the *Hoxa5* bait is denoted by a black vertical line, and gene position (blue arrows) and the ES CTCF ChIP-seq profile (black) is shown underneath the 4C plots. Red and green asterisks denote the positions of the single and double CTCF site deletions, as for **(A)**. In this plot, CTCF-dependent restriction of interactions between *Hoxa5* and more caudal regions (e.g. the gene body of *Hoxa10*) seems apparent.

DATA AVAILABILITY STATEMENT

Raw 4C data and the processed cis files first described here are available on GEO (GSE137417); previously published 4C results (Narendra et al., 2015) are available as data set GSE60240. ChIP-seq tracks were obtained from GEO for mouse ES (Shen et al., 2012) (CTCF and H3K27ac; GSE29218) and DP cells (Vanhille et al., 2015) (H3K27ac; GSE63732).

AUTHOR CONTRIBUTIONS

YBZ and TS designed and wrote the 4See code and accompanying scripts. AP and AM performed 4C experiments and preprocessing. TS wrote the manuscript, with continuous feedback from all authors.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01372/full#supplementary-material

REFERENCES

Alexander, J. M., Guan, J., Li, B., Maliskova, L., Song, M., Shen, Y., et al. (2019). Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *Elife* 8, e41769. doi: 10.7554/eLife.41769

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. *Dev. Cell* 16 (1), 47–57. doi: 10.1016/j.devcel.2008.11.011

Arnold, C. D., Gerlach, D., Stelzer, C., Boryn, L. M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* 339 (6123), 1074–1077. doi: 10.1126/science.1232542

Ben Zouari, Y., 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. doi: 10.1186/s13059-019-1706-3

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. Cell.* 76 (3), 473–484. doi: 10.1016/j.molcel.2019.07.038

Bibel, M., Richter, J., Lacroix, E., and Barde, Y. A. (2007). Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nat. Protoc.* 2 (5), 1034–1043. doi: 10.1038/nprot.2007.147

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 (3), 557–572 e524. doi: 10.1016/j.cell.2017.09.043

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. U.S.A.* 107 (50), 21931–21936. doi: 10.1073/pnas.1016071107

de Wit, E., Braunschweig, U., Greil, F., Bussemaker, H. J., and van Steensel, B. (2008). Global chromatin domain organization of the Drosophila genome. *PloS Genet.* 4 (3), e1000045. doi: 10.1371/journal.pgen.1000045

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 (4), 676–684. doi: 10.1016/j.molcel.2015.09.023

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* 295 (5558), 1306–1311. doi: 10.1126/science.1067799

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

Geeven, G., Teunissen, H., de Laat, W., and de Wit, E. (2018). peakC: a flexible, non-parametric peak calling package for 4C and Capture-C data. *Nucleic Acids Res.* 46 (15), e91. doi: 10.1093/nar/gky443

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459 (7243), 108–112. doi: 10.1038/nature07829

Herranz, D., Ambesi-Impiombato, A., Palomero, T., Schnell, S. A., Belver, L., Wendorff, A. A., et al. (2014). A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. *Nat. Med.* 20 (10), 1130–1137. doi: 10.1038/nm.3665

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 (2), 205–212. doi: 10.1038/ng.2871

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 (6), 996–1006. doi: 10.1101/gr.229102

Kim, T. K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465 (7295), 182–187. doi: 10.1038/nature09033

Klein, F. A., Pakozdi, T., Anders, S., Ghavi-Helm, Y., Furlong, E. E., and Huber, W. (2015). FourCSeq: analysis of 4C sequencing data. *Bioinformatics* 31 (19), 3085–3091. doi: 10.1093/bioinformatics/btv335

Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T. K., Zacarias-Cabeza, J., et al. (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat. Struct. Mol. Biol.* 18 (8), 956–963. doi: 10.1038/nsmb.2085

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10 (3), R25. doi: 10.1186/gb-2009-10-3-r25

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 (14), 1725–1735. doi: 10.1093/hmg/ddg180

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

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 (6225), 1017–1021. doi: 10.1126/science.1262088

Palstra, R. J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., and de Laat, W. (2003). The beta-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* 35 (2), 190–194. doi: 10.1038/ng1244

Pradeepa, M. M., Grimes, G. R., Kumar, Y., Olley, G., Taylor, G. C., Schneider, R., et al. (2016). Histone H3 globular domain acetylation identifies a new class of enhancers. *Nat. Genet.* 48 (6), 681–686. doi: 10.1038/ng.3550

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 (7333), 279–283. doi: 10.1038/nature09692

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 (7), 1665–1680. doi: 10.1016/j.cell.2014.11.021

Raviram, R., Rocha, P. P., Muller, C. L., Miraldi, E. R., Badri, S., Fu, Y., et al. (2016). 4C-ker: A Method to Reproducibly Identify Genome-Wide Interactions Captured by 4C-Seq Experiments. *PloS Comput. Biol.* 12 (3), e1004780. doi: 10.1371/journal.pcbi.1004780

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43 (7), e47. doi: 10.1093/nar/gkv007

Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., et al. (2015). Integrative analysis of 111 reference human epigenomes. *Nature* 518 (7539), 317–330. doi: 10.1038/nature14248

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotech.* 29 (1), 24–26. doi: 10.1038/nbt.1754

Sahlen, P., Abdullayev, I., Ramskold, D., Matskova, L., Rilakovic, N., Lotstedt, B., et al. (2015). Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. *Genome Biol.* 16, 156. doi: 10.1186/s13059-015-0727-9

Sanyal, A., Lajoie, B. R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature* 489 (7414), 109–113. doi: 10.1038/nature11279

Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* 20 (8), 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. (2015). The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* 25 (4), 582–597. doi: 10.1101/gr.185272.114

Schwartzman, O., Mukamel, Z., Oded-Elkayam, N., Olivares-Chauvet, P., Lubling, Y., Landan, G., et al. (2016). UMI-4C for quantitative and targeted chromosomal contact profiling. *Nat. Methods* 13 (8), 685–691. doi: 10.1038/nmeth.3922

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 (7409), 116–120. doi: 10.1038/nature11243

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 (11), 1348–1354. doi: 10.1038/ng1896

Thongjuea, S., Stadhouders, R., Grosveld, F. G., Soler, E., and Lenhard, B. (2013). r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. *Nucleic Acids Res.* 41 (13), e132. doi: 10.1093/nar/gkt373

van de Werken, H. J., Landan, G., Holwerda, S. J., Hoichman, M., Klous, P., Chachik, R., et al. (2012). Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat. Methods* 9 (10), 969–972. doi: 10.1038/nmeth.2173

Vanhille, L., Griffon, A., Maqbool, M. A., Zacarias-Cabeza, J., Dao, L. T., Fernandez, N., et al. (2015). High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nat. Commun.* 6, 6905. doi: 10.1038/ncomms7905

Walter, C., Schuetzmann, D., Rosenbauer, F., and Dugas, M. (2019). Benchmarking of 4C-seq pipelines based on real and simulated data. *Bioinformatics* 35 (23), 4938–4945. doi: 10.1093/bioinformatics/btz426

Williams, R. L.Jr., Starmer, J., Mugford, J. W., Calabrese, J. M., Mieczkowski, P., Yee, D., et al. (2014). fourSig: a method for determining chromosomal interactions in 4C-Seq data. *Nucleic Acids Res.* 42 (8), e68. doi: 10.1093/nar/gku156

Zhou, H. Y., Katsman, Y., Dhaliwal, N. K., Davidson, S., Macpherson, N. N., Sakthidevi, M., et al. (2014). A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. *Genes Dev.* 28 (24), 2699–2711. doi: 10.1101/gad.248526.114