Convergence of topological domain boundaries, insulators, and polytene interbands revealed by high-resolution mapping of chromatin contacts in the early *Drosophila melanogaster* embryo.

Michael R. Stadler[1], Jenna E. Haines[1], and Michael B. Eisen[1,2,3]

1. Department of Molecular and Cell Biology, University of California, Berkeley, CA
2. Department of Integrative Biology, University of California, Berkeley, CA
3. Howard Hughes Medical Institute, University of California, Berkeley, CA

Abstract

Evidence has emerged in recent years linking insulators and the proteins that bind them to the higher order structure of animal chromatin, but the precise nature of this relationship and the manner by which insulators influence chromatin structure have remained elusive. Here we present high-resolution genome-wide chromatin conformation capture (Hi-C) data from early *Drosophila melanogaster* embryos that allow us to map three-dimensional interactions to 500 base pairs. We observe a complex, nested pattern of regions of chromatin self-association, and use a combination of computational and manual annotation to identify boundaries between these topological associated domains (TADs). We demonstrate that, when mapped at high resolution, boundaries resemble classical insulators: short (500 - 1000 bp) genomic regions that are sensitive to DNase digestion and strongly bound by known insulator proteins. Strikingly, we show that for regions where the banding pattern of polytene chromosomes has been mapped to genomic position at comparably high resolution, there is a perfect correspondence between polytene banding and our chromatin conformation maps, with boundary insulators forming the interband regions that separate compacted bands that correspond to TADs. We propose that this precise, high-resolution relationship between insulators and TADs on the one hand and polytene bands and interbands on the other extends across the genome, and suggest a model in which the decompaction of insulator regions drives the organization of interphase chromosomes by creating stable physical separation between adjacent domains.

Introduction

Beginning in the late 19th century, cytological investigations of the polytene chromosomes of insect salivary glands implicated the physical structure of interphase chromosomes in their cellular functions (Balbiani, 1881, 1890; Heitz & Bauer, 1933; King & Beams, 1934; Painter, 1935). Over the next century plus, studies in the model insect species *Drosophila melanogaster* were instrumental in defining structural features of animal chromatin. Optical and electron microscopic analysis of fly chromosomes produced groundbreaking insights into the physical nature of genes, transcription and DNA replication (Belyaeva & Zhimulev, 1994; Benyajati & Worcel, 1976; Laird & Chooi, 1976; Laird, Wilkinson, Foe, & Chooi, 1976; McKnight & Miller, 1976, 1977; Vlassova et al., 1985).

Detailed examination of polytene chromosomes in *Drosophila melanogaster* revealed a stereotyped organization, with compacted, DNA-rich "bands" alternating with extended, DNA-poor "interband" regions (Benyajati & Worcel, 1976; Bridges, 1934; Laird & Chooi, 1976; Lefevre, 1976; Rabinowitz, 1941), and it appears likely that this structure reflects general features of chromatin organization shared by non-polytene chromosomes. While these classical studies offered extensive structural and molecular characterization of chromosomes *in vivo*, the question of what was responsible for organizing chromosome structure remained unanswered.

A critical clue came with the discovery of insulators, DNA elements initially identified based on their ability to block the activity of transcriptional enhancers when located between an enhancer and its targeted promoters (Geyer & Corces, 1992; Holdridge & Dorsett, 1991; Rebecca Kellum & Schedl, 1991; R. Kellum & Schedl, 1992). Subsequent work showed that these elements could also block the spread of silenced chromatin states (Kahn, Schwartz, Dellino, & Pirrotta, 2006; Mallin, Myung, Patton, & Geyer, 1998; Recillas-Targa et al., 2002; Roseman, Pirrotta, & Geyer, 1993; Sigrist & Pirrotta, 1997) and influence the structure of chromatin. Through a combination of genetic screens and biochemical purification, a number of protein factors have been identified that bind to *Drosophila* insulators and modulate their function, including Su(Hw), BEAF-32, mod(mdg4), CP190, dCTCF, GAF, Zw5, and others (Bell, West, & Felsenfeld, 1999; Büchner et al., 2000; Gaszner, Vazquez, & Schedl, 1999; Tatlana I. Gerasimova, Gdula, Gerasimov, Simonova, & Corces, 1995; Lewis, 1981; Lindsley & Grell, 1968; Melnikova et al., 2004; Moon et al., 2005; Pai, Lei, Ghosh, & Corces, 2004; Parkhurst et al., 1988; Parkhurst & Corces, 1985, 1986; Scott, Taubman, & Geyer, 1999; Spana, Harrison, & Corces, 1988; Zhao, Hart, & Laemmli, 1995). Except for CTCF, which is found throughout bilateria, all of these proteins appear to be specific to arthropods (Heger, George, & Wiehe, 2013).

Staining of polytene chromosomes with antibodies against such insulator proteins showed that many of them localize to polytene interbands (Belyaeva & Zhimulev, 1994; Berkaeva, Demakov, Schwartz, & Zhimulev, 2009; Byrd &

Corces, 2003; Eggert, Gortchakov, & Saumweber, 2004; Gilbert, Tan, & Hart, 2006; Gortchakov et al., 2005; Pai et al., 2004; T. Yu Vatolina et al., 2011; Zhao et al., 1995), with some enriched at interband borders. Further, some, though not all, insulator protein mutants disrupt polytene chromosome structure (Roy, Gilbert, & Hart, 2007). Together, these data implicate insulator proteins, and the elements they bind, in the organization of the three-dimensional structure of fly chromosomes.

Several high-throughput methods to probe three-dimensional structure of chromatin have been developed in the last decade (Beagrie et al., 2017; Fullwood et al., 2009; Lieberman-Aiden et al., 2009; Rao et al., 2014). Principle among these are derivatives of the chromosome conformation capture (3C) assay (Dekker, Rippe, Dekker, & Kleckner, 2002), including the genome-wide "Hi-C" (Lieberman-Aiden et al., 2009). Several groups have performed Hi-C on *Drosophila* tissues or cells and have shown that fly chromosomes, like those of other species, are organized into topologically associated domains (TADs), regions within which loci show enriched 3C linkages with each other but depleted linkages with loci outside the domain. Disruption of TAD structures by gene editing in mammalian cells has been shown to disrupt enhancer-promoter interactions and significantly alter transcriptional activity (Guo et al., 2015; Lupiáñez et al., 2015).

Although TADs appear to be a ubiquitous feature of animal genomes, the extent to which TAD structures are a general property of a genome or if they are regulated as a means to control genome function remains unclear, and the question of how TAD structures are established remains open. Previous studies have implicated a number of features in the formation of *Drosophila* TAD boundaries, including transcriptional activity and gene density, and have reached differing conclusions about the role played by insulator protein binding (Hou, Li, Qin, & Corces, 2012; L. Li et al., 2015; Sexton et al., 2012; Ulianov et al., 2015; Van Bortle et al., 2014). Tantalizingly, Eagen et al., using 15 kb resolution Hi-C data from *D. melanogaster* have shown that there is a correspondence between the distribution of large TADs and polytene bands (Eagen, Hartl, & Kornberg, 2015).

We have been studying the formation of chromatin structure in the early *D. melanogaster* embryo because of its potential impact on the establishment of patterned transcription during the initial stages of development. We have previously has shown that regions of "open" chromatin are established at enhancers and promoters prior to the onset of transcriptional activation (Harrison, Li, Kaplan, Botchan, & Eisen, 2011; X.-Y. Li, Harrison, Villalta, Kaplan, & Eisen, 2014) and were interested the extent to which chromatin exhibited associated three-dimensional structures before, during and after transcriptional activation. We were further interested in investigating the role chromatin structure plays in spatial patterning.

We therefore generated high-resolution Hi-C datasets derived from two timepoints during early *Drosophila melanogaster* embryonic development, and from the anterior and posterior halves of hand-dissected embryos. Echoing recent results from (Hug, Grimaldi, Kruse, & Vaquerizas, 2017) we find that extensive chromatin structure emerges around the onset of zygotic genome activation. We show that high-resolution chromatin maps of anterior and posterior halves are nearly identical, suggesting that chromatin structure neither drives nor directly reflects spatially patterned transcriptional activity. However we show that stable long-range contacts evident in our chromatin maps generally involve known patterning genes, implicating chromatin conformation in transcriptional regulation.

To investigate the origins of three-dimensional chromatin structure, we carefully map the locations of the boundaries between topological domains using a combination of manual and computational annotation. We demonstrate that these boundaries resemble classical insulators: short (500 - 1000 bp) genomic regions that are strongly bound by (usually multiple) insulator proteins and are sensitive to DNase digestion. We show that for a region in which the fine polytene banding pattern has been mapped to genomic positions, these boundary insulators correspond perfectly to interband regions that separate compacted bands corresponding to TADs. We propose that this precise, high-resolution relationship between insulators, TADs and polytene bands and interbands extends across the genome, and suggest a model in which the decompaction of insulator regions drives the organization of interphase chromosomes by creating stable physical separation between adjacent domains.

Results

Data quality and general features

We prepared and sequenced *in situ* Hi-C libraries from two biological replicates each of hand-sorted cellular blastoderm (mitotic cycle 14; mid-stage 5) and pre-blastoderm (mitotic cycle 12) embryos using a modestly adapted version of the protocol described in (Rao et al., 2014). To examine possible links between chromatin maps and transcription, we sectioned hand-sorted mitotic cycle 14 embryos along the anteroposterior midline, and generated Hi-C

Figure 1

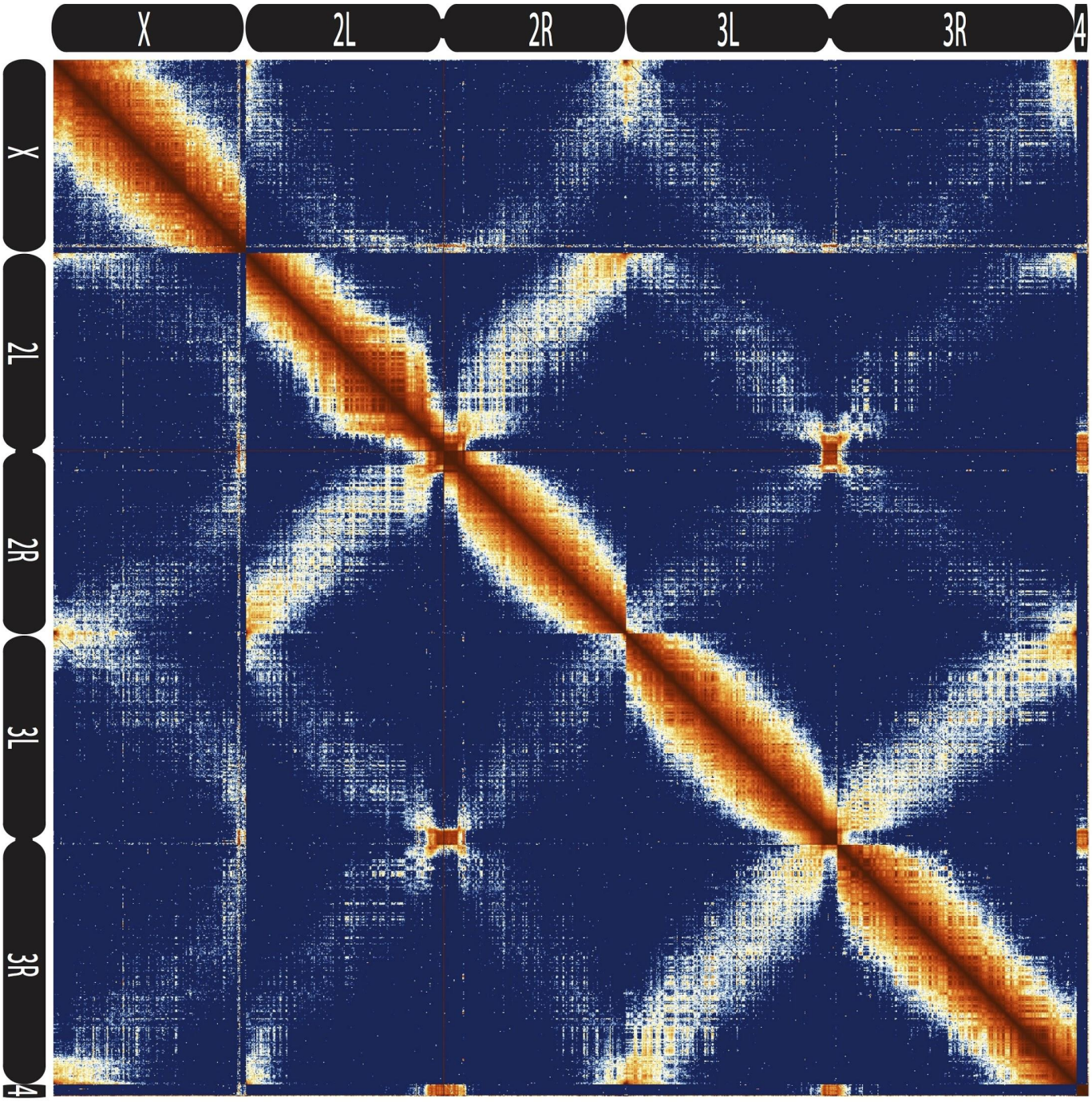


Figure 1: Hi-C map of the stage 5 *Drosophila melanogaster* genome at 100 kb resolution. Data from all nc14 datasets was aggregated and normalized by the "vanilla coverage" method. To enhance contrast, the logarithm of the normalized count values were histogram equalized, and maximum and minimum values were adjusted for optimal display..

data from the anterior and posterior halves separately, also in duplicate. In total, we produced ~288 million informative read pairs (see **Tables S1** and **S2**).

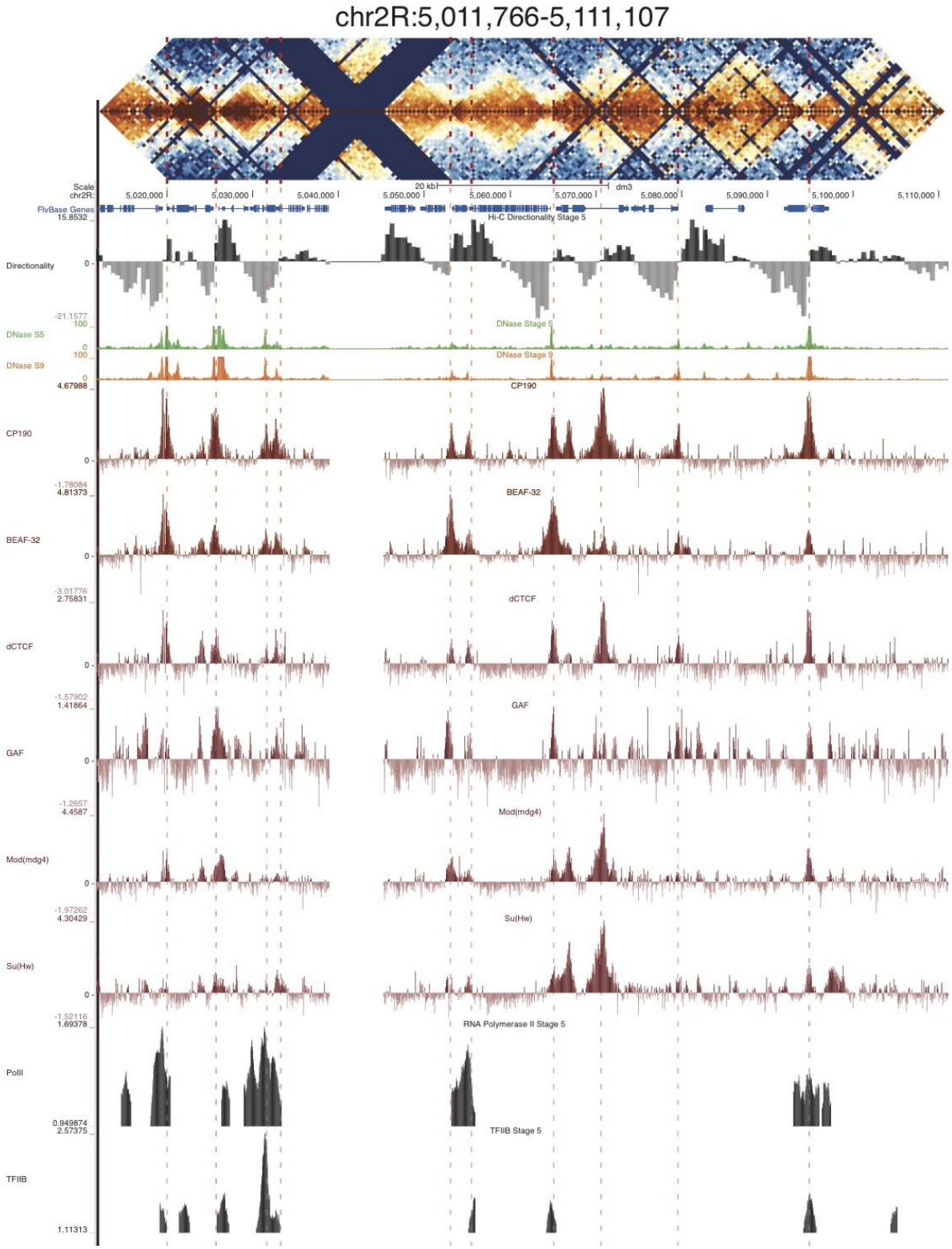
We assessed the quality of these data using metrics similar to those described by (Lieberman-Aiden et al., 2009; Rao et al., 2014). Specifically, the strand orientations of our reads were approximately equal in each sample (as expected from correct Hi-C libraries but not background genomic sequence; see **Table S2**), the signal decay with genomic distance was similar across samples, and, critically, visual inspection of heat maps prepared at a variety of resolutions showed these samples to be very similar both to each other and to previously published data prepared using similar methods (Sexton et al., 2012). We conclude that these Hi-C are of high quality and reproducibility.

We next sought to ascertain the general features of the data at low resolution. We examined heatmaps for all *D. melanogaster* chromosomes together using 100 kb bins, as shown in **Figure 1**. Several features of the data are immediately apparent. The prominent "X" patterns for chromosomes 2 and 3, which indicate an enrichment of linkages between chromosome arms, reflects the known organization of fly chromosomes during early development known as the Rabl configuration (Csink & Henikoff, 1998; Duan et al., 2010; G. Wilkie A Shermoen P. O'Farrell, 1999): telomeres are located on one side of the nucleus, centromeres are located on the opposite side, and chromosome arms are arranged roughly linearly between them. Centromeres and the predominantly heterochromatic chromosome 4 cluster together, as, to a lesser extent, do telomeres, reflecting established cytological features that have been detected by prior Hi-C analysis (Sexton et al., 2012) and fluorescence in situ hybridization (FISH) (Lowenstein, Goddard, & Sedat, 2004). These features were evident in all replicates, further confirming both that these datasets are reproducible and that they capture known features of chromatin topology and nuclear arrangement.

TAD boundaries are short elements bound by insulator proteins

Because we used a 4-cutter restriction enzyme and deep sequencing, and because the fly genome is comparatively small, we were able to resolve features at resolutions below one kilobase. We visually inspected genome-wide maps constructed using bins of 500bp, and were able to see a conspicuous pattern of TADs across a wide range of sizes, some smaller than 5 kb (**Fig. 2, Fig. S1**). When we compared maps for several of these regions with available functional genomic data from embryos, we observed that the boundaries between these domains were formed by short regions of

Figure 2



DNA

Figure 2: Example region of Hi-C data at 500 bp resolution. Heat map of aggregate Hi-C data for all nc14 datasets binned at 500 bp is shown for the region located at chr2R:5,011,766-5,111,107. Raw counts were normalized by the vanilla coverage method, the logarithm was taken, and minimum and maximum values were adjusted for visual contrast. A UCSC browser (Kent et al., 2002) window for the corresponding coordinates is shown with tracks for Hi-C directionality (calculated from the Hi-C data shown in the heatmap), DNase accessibility (X.-Y. Li et al., 2011), RNA polII and TFIIB (X. Y. Li et al., 2008), and the insulator proteins CP190, BEAF-32, dCTCF, GAF, mod(mdg4), and Su(Hw) from (Nègre et al., 2010). Dashed red lines are visual guides and are manually drawn at locations of apparent boundaries; they do not reflect algorithmically or unbiased hand-curated boundary calls. Additional identically prepared example regions are shown in **Figure S1**.

(500-1000 bp) that are nearly always associated with high chromatin accessibility, measured by DNase-seq (X.-Y. Li et al., 2011), and strong occupancy by known insulator proteins as measured by chromatin immunoprecipitation (ChIP) (Nègre et al., 2010) (**Fig. 2, Fig. S1**), all properties of classical insulator elements.

To confirm this visually striking association, we systematically called TAD boundaries by visual inspection of panels of raw Hi-C data covering the entire genome. Critically, these boundary calls were made from Hi-C data alone, and the human caller lacked any information about the regions being examined, including which region (or chromosome) was represented by a given panel. In total, we manually called 3,524 boundaries in the genome for stage 5 embryos. Taking into account the ambiguity associated with intrinsically noisy data, the difficulty of resolving small domains, and the invisibility of sections of the genome due to repeat content or a lack of MboI cut sites, we consider 3,500-5,000 to be a reasonable estimate for the number of boundaries in the genome.

To complement these manual calls, we developed a computational approach for calling boundaries that is similar to methods used by other groups (Crane et al., 2015; Lieberman-Aiden et al., 2009; Rao et al., 2014; Sexton et al., 2012). In brief, we assigned a directionality score to each genomic bin based on the number of Hi-C reads linking the bin to upstream versus downstream regions, and then used a set of heuristics to identify points of transition between regions of upstream and downstream bias. We adjusted the parameters of the directionality score and the boundary calling to account for features of the fly genome, specifically the relatively small size of many topological domains. Of our top 1000 domain boundaries identified by this approach, 937 were matched by a manually-called boundary within 1 kb, which suggested that this approach robustly identified the domain features that are apparent by eye.

Comparing these 937 high-confidence boundaries to other genomic datasets shows a highly stereotyped pattern of associated genomic features. Most strikingly, boundaries are enriched for the binding of the known insulator proteins

CP190, BEAF-32, mod(mdg4), dCTCF, and to a lesser extent GAF and Su(Hw) (**Fig. 3**). CP190 and BEAF-32 show the strongest enrichment, and indeed, virtually all of the examined boundaries feature proximal CP190 peaks (**Fig. S2**). Domains of H3K27 trimethylation, a marker of polycomb silencing, showed a strong tendency to terminate at boundaries, and the enhancer mark H3K4me1 showed an interesting pattern of depletion at boundaries but enrichment immediately adjacent to boundary locations (**Fig. 3**). Boundaries also exhibit peaks of DNase accessibility and nucleosome depletion (**Fig. 3**), as well as marks associated with promoters, including the general transcription factors TFIIB and the histone tail modification H3K4me3. Despite the presence of promoter marks, we find that RNA polII is present at only a subset of stage 5 boundaries (**Fig. 3, Fig. S2**).

It has previously been noted by multiple authors that insulator proteins tend to bind near promoters, specifically between divergent promoters (Nègre et al., 2010; Ramirez et al., 2017). Indeed, we find that boundary elements, as identified from Hi-C, are often found proximal to promoters and show a general enrichment of promoter-associated marks (**Fig. 3**). We wondered whether some of the enrichments we observe were marks of promoters, and were not specific features of topological boundaries. We therefore examined the distributions of the same genomic features around the top 1000 peaks of H3K4me3, a marker of active promoters, in data from stage 5 embryos (**Fig. S3**) (X.-Y. Li et al., 2014). While these sites show enrichments for insulator proteins, these enrichments are substantially weaker than those observed at topological boundaries, while RNA polII enrichment is much stronger at promoters than boundaries. The tendency for polycomb domains to terminate at promoters is also much less pronounced at promoters than boundaries. Additionally, Hug et al. pharmacologically inhibited transcription in early embryos and found that TAD structure was not significantly altered (Hug et al., 2017). While we cannot rule out any role for promoter-bound transcription machinery in the formation of topological boundaries, we think it is unlikely that transcriptional activity is responsible for establishing topological domains.

We next asked whether the presence of these features was predictive of topological boundaries. We plotted the distribution of Hi-C directionality scores around peaks of these features and observed that insulator proteins, DNase accessibility peaks, and H3K4me3 peaks are all clearly associated with directionality changes used to define boundaries (**Fig. S4**). Further, we performed logistic regression individually on each feature to assess how well it predicted the location of boundaries. The results closely matched the enrichment analysis, with CP190, BEAF-32, mod(mdg4), and

dCTCF as the four most predictive features, in order (**Table S3**).

Figure 3

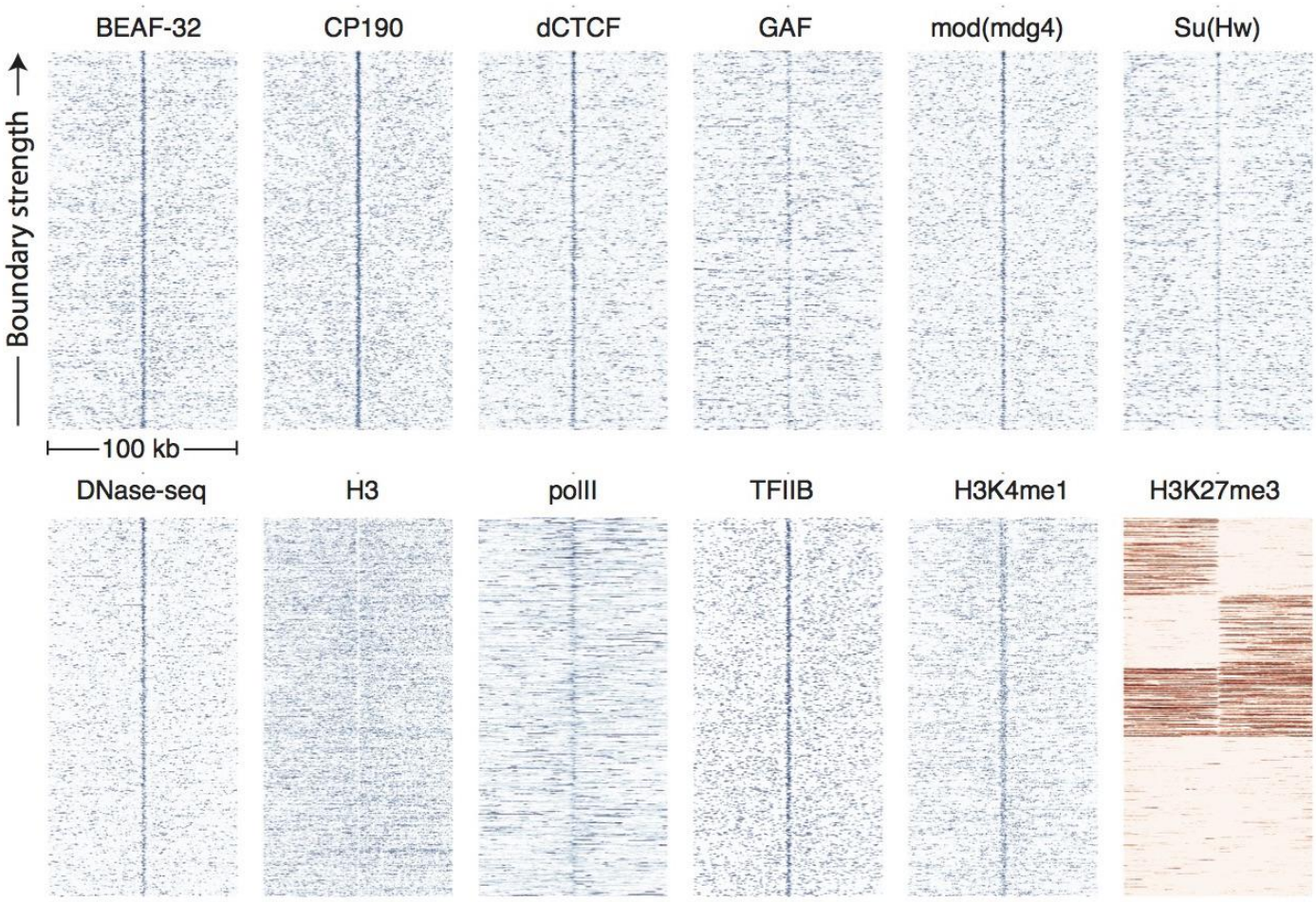


Figure 3: Topological domain boundaries show distinct patterns of associated proteins and genomic features.

Heatmaps showing the distribution of signals from embryonic ChIP and DNase-seq datasets around 937 topological boundaries identified jointly by computational and manual curation. All plots show 500 bp genomic bins in 100 kb windows around boundaries. All plots in blue are sorted by boundary strength, calculated from the difference in upstream and downstream Hi-C directionality scores. The plot for H3K27me3 (in red) is specially sorted to highlight the tendency for enriched domains to terminate at boundaries. Rows for this plot were sorted by calculating the total H3K27me3 signal in the 50 kb windows upstream and downstream of the boundary and then sorting, top to bottom: upstream signal above median and downstream signal below the median, upstream below and downstream above, upstream and downstream both above, upstream and downstream both below the median. For comparison, identically prepared and sorted plots around H3K4me3 peaks are shown in Fig. S3. Data for insulator proteins, DNase accessibility, RNA polII and TFIIB are from the same sources indicated in Fig. 2. ChIP for H3, H3K4me1 are taken from (X.-Y. Li et al., 2014), and H3K27me3 are from modEncode (Contrino et al., 2012).

Finally, we examined the sequence composition of boundary elements by comparing the frequency of DNA words of up to seven base pairs in the set of high confidence boundaries to flanking sequence. The most enriched sequences correspond to the known binding site of BEAF-32 and to a CACA-rich motif previously identified as enriched in regions bound by CP190 (Nègre et al., 2010; Yang & Corces, 2012), both of which show strong association with the set of boundary sequences as a whole (**Fig. S8**).

These analyses indicate that the boundaries between topological domains identified from Hi-C data map to short regions of open chromatin that are strongly bound by insulator proteins, and, conversely, that sites of insulator protein binding are strongly associated with boundaries in Hi-C data. Thus we conclude that the concepts of TAD boundaries and insulators describe a largely overlapping set of genetic elements, and, in general, can be considered equivalent.

Boundary elements correspond to polytene interbands

The identification of these boundary elements led us to consider the physical basis of topological domain separation. Chromosome conformation capture is a complex assay (Gavrilov et al., 2013; Gavrilov, Razin, & Cavalli, 2015), and inferring discrete physical states of the chromatin fiber from Hi-C signals generally requires orthogonal experimental data. To address this problem, we sought to leverage information from polytene chromosomes to draw associations between features of Hi-C data and physical features of chromosomes.

There is surprisingly little data mapping features of polytene structure to specific genomic coordinates at high resolution. Vatolina et al. (Tatyana Yu Vatolina et al., 2011) used exquisitely careful electron microscopy to identify the fine banding pattern of the 65 kb region between polytene bands 10A1-2 and 10B1-2, revealing that this region, which appears as a single interband under a light microscope, actually contains six discrete, faint bands and seven interbands. The region is flanked by two large bands, whose genomic locations has been previously mapped and refined by FISH (Tatyana Yu Vatolina et al., 2011) . Vatolina et al. then used available molecular genomic data to propose a fine mapping of these bands and interbands to genomic coordinates.

Figure 4

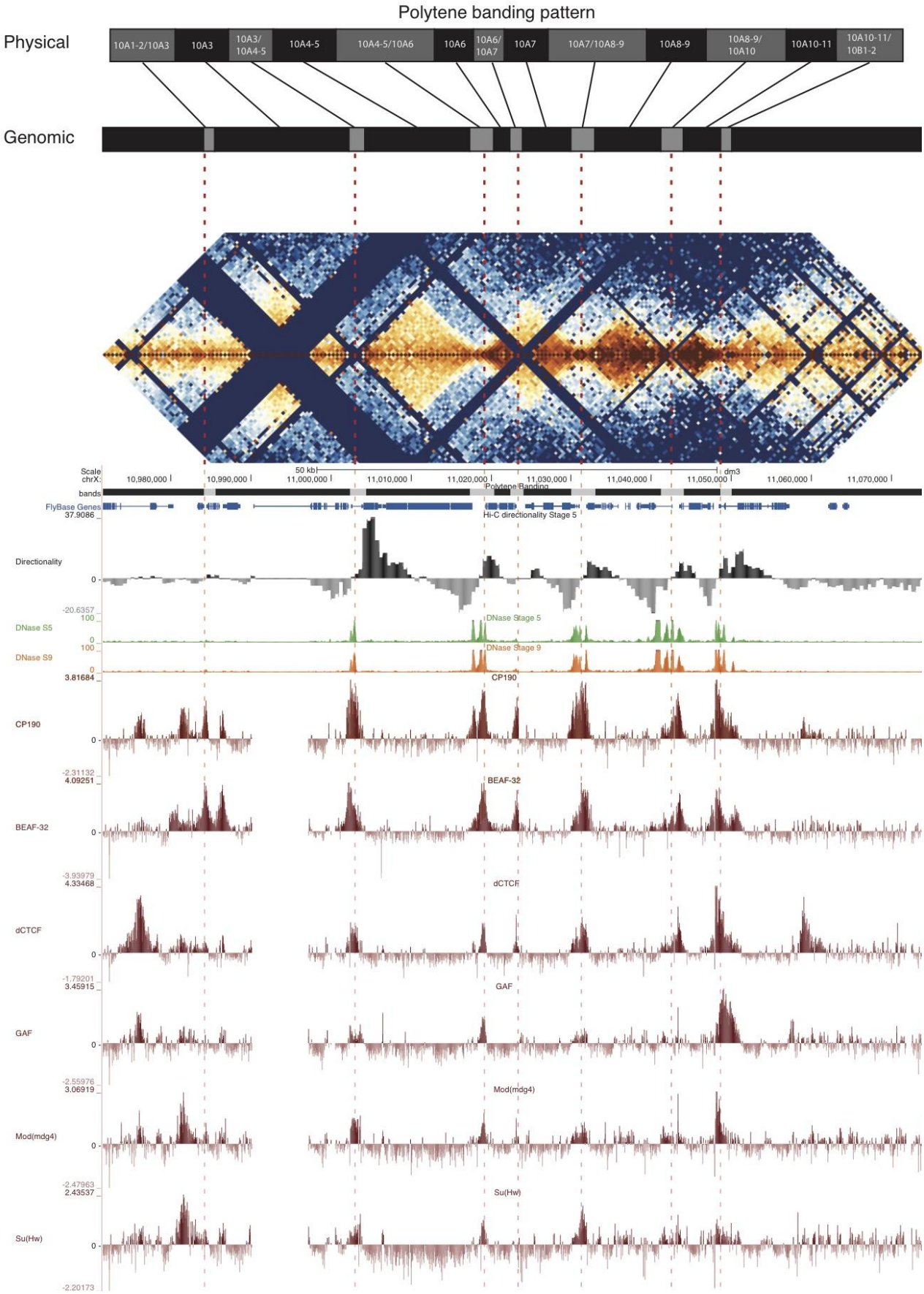


Figure 4: Topological boundary elements correspond to polytene interbands. Heat map of aggregate Hi-C data for all nc14 datasets binned at 500 bp and UCSC browser data shown for the region chrX:10,971,551-11,074,797 for which Vatolina et al. provided fine-mapping of polytene banding structure. Hi-C and browser data were prepared and sourced as indicated in Fig. 2. Dashed red lines are visual guides drawn to highlight coincidence of apparent Hi-C boundaries, insulator protein binding sites, and the interband assignments of Vatolina et al. Top: accurately-scaled representations of the size of the mapped bands and interbands in base pairs ("Genomic") and the corresponding physical distances in polytene chromosomes derived from electron microscopic analysis of polytene chromosomes by Vatolina et al. Increased relative physical size of interband regions demonstrates their lower compaction ratios.

Figure 4 shows the correspondence between Vatolina et al.'s proposed polytene map from this region and our high-resolution Hi-C data, along with measures of early embryonic DNase hypersensitivity from (X.-Y. Li et al., 2011) and the binding of six insulator proteins (Nègre et al., 2010). Strikingly, there is a near-perfect correspondence between the assignments of Vatolina et al. and our Hi-C data: bands correspond to TADs, and interbands correspond to the boundary elements that separate the TADs. To some extent this is unsurprising given that we showed above that inter-TAD regions are strongly bound by insulator proteins, and Vatolina et al. relied on insulator binding in their proposed mapping. However, the overall architecture of both TADs and polytene bands in this region are highly similar, with the same number of TADs and bands, and variation in TAD size corresponding to variation in band intensity.

The 5' region of the *Notch* gene has also been carefully mapped. Rykowski et al. used high-resolution *in situ* hybridization to determine that the coding sequences of *Notch* lies within polytene band 3C7, while the sequences upstream of the transcription start site (TSS) lie in the 3C6-7 interband. Examining the *Notch* locus in our Hi-C data, we see that the gene body is located within an ~20 kb TAD, and the TSS directly abuts a TAD boundary that is strongly bound by CP190 and dCTCF (**Fig. S7**), an arrangement consistent with the correspondence of boundaries and interbands.

Eagen et al. previously identified a broad correspondence between polytene interbands and inter-TAD regions from Hi-C data at 15 kb resolution (Eagen et al., 2015). At sub-kilobase resolution, Hi-C data allows the detection of fine structure within these inter-TAD regions, down to individual boundary elements. The precise correspondence between boundary elements and polytene interbands suggested by a comparison of our data and the proposed polytene map of Vatolina et al has not, to our knowledge, been described previously.

The association between boundary elements and interbands suggests a simple model for insulator function. A key feature that distinguishes polytene interbands from bands is their low compaction ratio: they span a larger physical

distance per base pair. The association between insulator binding and genomic regions with low compaction ratios suggests insulators may function by simply increasing the physical distance between adjacent domains via the unpacking and extension of intervening chromatin. **Figure 4** (top) shows a representation of the conversion of genomic distance to physical distance for the 10A region, as measured by Vatolina et al. Any model for insulator function must explain several features of insulator function, including the ability to organize chromatin into physical domains, block interactions between enhancers and promoters exclusively when inserted between them, and protect transgenes from position effect variegation and block the spread of chromatin silencing states. This chromatin extension model for insulator function can explain these defining characteristics via simple physical separation.

Topological boundaries are identical in anterior and posterior sections of the embryo

We next asked whether the boundaries we identified as boundary elements represent constitutive features of chromatin organization or whether their function might be regulated in a cell-type specific or developmental manner. We reasoned that, since different sets of patterning genes are transcribed in the anterior and posterior portions of the pre-gastrula *D. melanogaster* embryos, a comparison of chromatin interaction maps between anterior and posterior regions would reveal whether context, especially transcriptional state, affects the TAD/boundary structure of the genome. To this end, we performed two separate biological replicates of an experiment in which we sectioned several hundred mid stage 5 embryos along the anteroposterior midline, and produced deep-sequenced Hi-C libraries from the anterior and posterior halves in parallel.

Resulting Hi-C signals at boundaries are virtually identical in the two halves, despite substantially different gene expression profiles in these two embryonic regions (**Fig. 5A**). Indeed, overall Hi-C signals are remarkably similar, with anterior and posterior samples correlating as strongly as replicates. Examination of individual loci at high resolution reveal consistent profiles and boundaries, notably including genes expressed differentially in the anterior or posterior (**Fig. 5B**).

The correspondence of insulator boundary elements and interbands, and the chromatin extension model, implies that the chromatin accessibility of insulator regions will be a useful proxy for their functionality in structurally organizing the genome. Intriguingly, (Van Bortle et al., 2014) found that DNase accessibility of insulator protein-bound regions

Figure 5

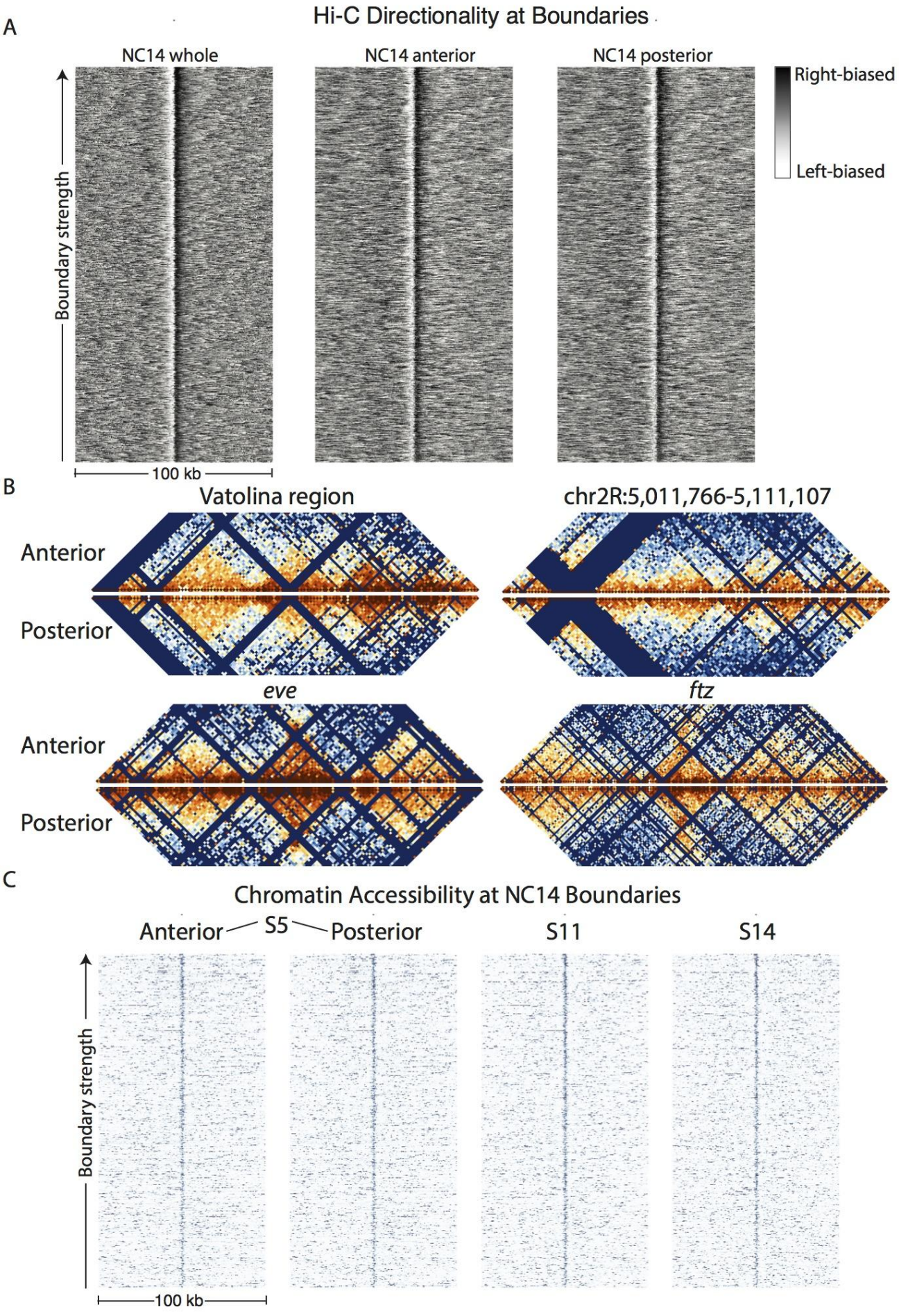


Figure 5: Hi-C signals from anterior and posterior halves of stage 5 embryos reveal highly similar chromatin topologies. (A) The distribution of Hi-C directionality scores in whole embryos, anterior, and posterior halves is shown around 937 topological boundaries identified jointly by computational and manual curation. **(B)** Heat maps of Hi-C data at 500 bp resolution at four example regions in anterior and posterior embryo halves. Plots represent the aggregate data of two biological and technical replicates each for anterior and posterior samples, and were prepared as in Fig. 2. The regions shown are the region mapped by Vatolina et al. (chrX:10,971,551-11,074,797), the example region from Fig. 2 (chr2R:5,011,766-5,111,107) , and the genomic regions surrounding the *eve* (chr2R:5,790,565-5,944,464) and *ftz* (chr3R:2,594,956-2,787,055) loci. **(C)** Chromatin accessibility around topological boundaries as measured by ATAC-seq in anterior and posterior nc14 (S5) embryos and by DNase-seq on stage 11 and 14 embryos (X.-Y. Li et al., 2011).

tracked with the ability of these sequences to block enhancer-promoter interactions in a cell-culture assay. We again sectioned embryos into anterior and posterior halves and performed ATAC-seq (Buenrostro, Giresi, Zaba, Chang, & Greenleaf, 2013) on pools of 20 embryo halves. ATAC-seq is a technique in which intact chromatin is treated with Tn5 transposase loaded with designed DNA sequences which are preferentially inserted into open, accessible chromatin regions. These insertions can be used to generate high-throughput sequencing libraries, producing data that is largely analogous to DNase-seq data. Analysis of ATAC-seq signal at insulator boundary elements in anterior and posterior halves showed that these elements have nearly identical accessibility in these two samples (**Fig. 5C**). Additionally, DNase-seq data from later embryonic stages that feature substantial tissue differentiation, transcription, and chromatin changes show highly consistent profiles at boundaries (**Fig. 5C, Fig. S5)**. These results are consistent with a model in which insulator-mediated chromatin organization is a constitutive feature of interphase chromatin of *D. melanogaster* embryos.

Topological boundaries begin to emerge before zygotic genome activation

The chromatin landscape of the early fly embryo is known to change dramatically upon the onset of zygotic genome activation (ZGA), the period approximately two hours into development in which the zygotic genome switches from a largely silent state to actively transcribing thousands of genes (Foe & Alberts, 1983; Hug et al., 2017; X.-Y. Li et al., 2014; McKnight & Miller, 1976; Newport & Kirschner, 1982; Tadros & Lipshitz, 2009). We asked whether the boundary effects we observed in stage 5 embryos were present in nuclei before ZGA. To address this, we prepared Hi-C libraries from two biological replicates of whole embryos hand-staged at nuclear cycle 12. Hug et al. (Hug et al., 2017)

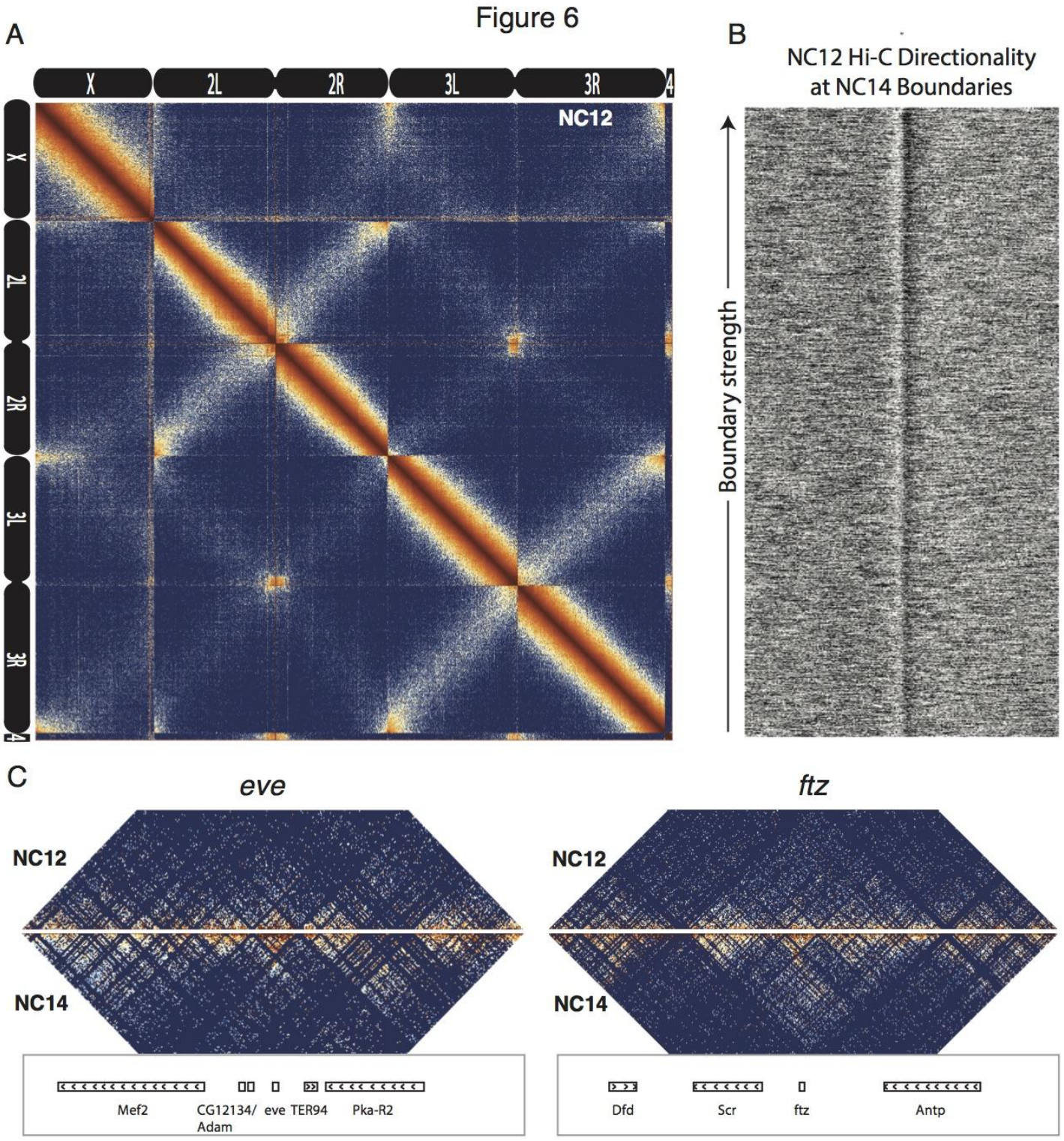


Figure 6: Pre-ZGA embryos show weaker boundary definition compared to post-ZGA embryos. (A) Heat map at 100 kb resolution of all autosomes and the X chromosome for aggregate data of two replicates of Hi-C from whole nc12 embryos. Broad features (Rabl chromosome configuration, interaction of centromeres and telomeres) are consistent with nc14, but nc12 data shows significantly less fine structure compared to nc14. **(B)** The distribution of Hi-C directionality signal from nc12 embryos around 937 topological boundaries defined from nc14 embryo data. **(C)** Hi-C maps comparing signal around two representative loci in nc12 data

with data from nc14 whole embryos matched for read depth to nc12 samples. Regions surround key developmental genes *eve* (chr2R:5,790,565-5,944,464) and *ftz* (chr3R:2,594,956-2,787,055).

recently used a similar Hi-C dataset to show that chromatin structure is largely featureless prior to ZGA and adopts complex structure only after ZGA onset. We observed a similar transition between nc12 and stage 5 (**Fig. 6A**, compare with **Fig. 1**). At individual insulator elements, we see that transitions in the directionality of Hi-C data is present but substantially weaker than in stage 5 embryos (**Fig. 6B**). A simple explanation for this data is that the establishment of decompacted boundary domains occurs on a time scale such that they are present in only a fraction of nuclei in our nc12 samples, which are not sorted to isolate a specific time points but rather are distributed randomly through the 15 minute nc12 interphase. Blythe et al. (Blythe & Wieschaus, 2016) performed ATAC-seq on early embryos staged precisely at three minute intervals and reported that chromatin containing insulators generally only opens late in the nuclear cycles. It is therefore likely that the components necessary for insulator boundary establishment are present and active in the early embryo, but are prevented by the rapid nuclear cycles of cleavage stage fly embryos from generating decompacted domains until the end of the longer interphase periods of the later nuclear cycles.

Distal chromatin contacts in the early fly embryo

Many models of insulator function invoke physical contact between insulators to form "looped" chromatin domains (Fujioka, Wu, & Jaynes, 2009; Kravchenko et al., 2005; Kyrchanova & Georgiev, 2014; Yang & Corces, 2012), and a substantial literature exists demonstrating that many insulator proteins are able to interact with each other and to self-associate (Blanton, Gaszner, & Schedl, 2003; Büchner et al., 2000; Gause, Morcillo, & Dorsett, 2001; Ghosh, Gerasimova, & Corces, 2001; Golovnin et al., 2007; Mohan et al., 2007; Pai et al., 2004; Vogelmann et al., 2014). In general, we do not observe looping interactions between domain boundaries in our Hi-C data. However, visual inspection of heat maps of Hi-C data for the entire genome identified 36 examples of interactions between non-adjacent domains (**Fig. S6, Table S4**), in addition to the previously noted clustering of PcG-regulated Hox gene clusters (Sexton et al., 2012).

The most visually striking locus, which we emphasize was identified in an unbiased manner without knowing its identify, is the locus containing the *Scr*, *ftz*, and *Antp* genes (**Fig. 7A**). This locus has been extensively studied, and a number of regulatory elements have been identified that reside between the *ftz* and *Antp* genes but "skip" the *ftz* promoter

to regulate *Scr* (Calhoun & Levine, 2003; Calhoun, Stathopoulos, & Levine, 2002). Consistent with this, we observe enriched contacts between the region containing the *Scr* promoter and a domain on the other side of *ftz* that contains the known *Scr*-targeting cis regulatory elements, while the *ftz*-containing domain makes minimal contact with its neighboring domains. Critically, we observe hot spots of apparent interaction between two sets of boundary elements (**Fig. 7A**: 1 and 4, 2 and 3), suggesting that physical association of boundary elements (or their associated proteins) may play a role in this interaction.

Curiously, we detected a similar situation on the other side of *Scr*, where a domain containing the hox gene *Dfd* is "skipped" over by the *Ama* locus to interact with a short element 3' of the *Scr* transcription unit (**Fig. S6**). We also observe a similar arrangement near the *eve* locus (**Fig. S6**). In these cases, a plausible topology is that the skipped domain is "looped out", preventing interaction with neighbors, while the adjacent domains are brought into proximity.

In addition to these domain-skipping events, we observe a small number of looping interactions, where two distal loci show high levels of interaction, without the associated enriched interactions between the domains flanking the loop. In every case we observe, the loop forms between two domain boundaries. As shown in **Figure 7B**, one of these loops brings together the promoters of *kni* and the related *knrl* genes. Other loops connect the *achaete* and *scute* genes, *slp1* and *slp2*, and the promoter of *Ubx* with an element in its first intron (**Fig. S6**).

These loci demonstrate that looping and domain-skipping events can be detected in our Hi-C data, but it appears that such interactions are rare and that looping does not occur between the overwhelming majority of insulator boundary elements. Nevertheless, it is striking that of the limited number of distal interactions we observed, many of them involve genes that are transcriptionally active during stage 5 of embryogenesis. This raises the possibility that these interactions may be stage or tissue-specific regulatory phenomena, and that more may be present in other tissues, developmental time points, or conditions.

Figure 7

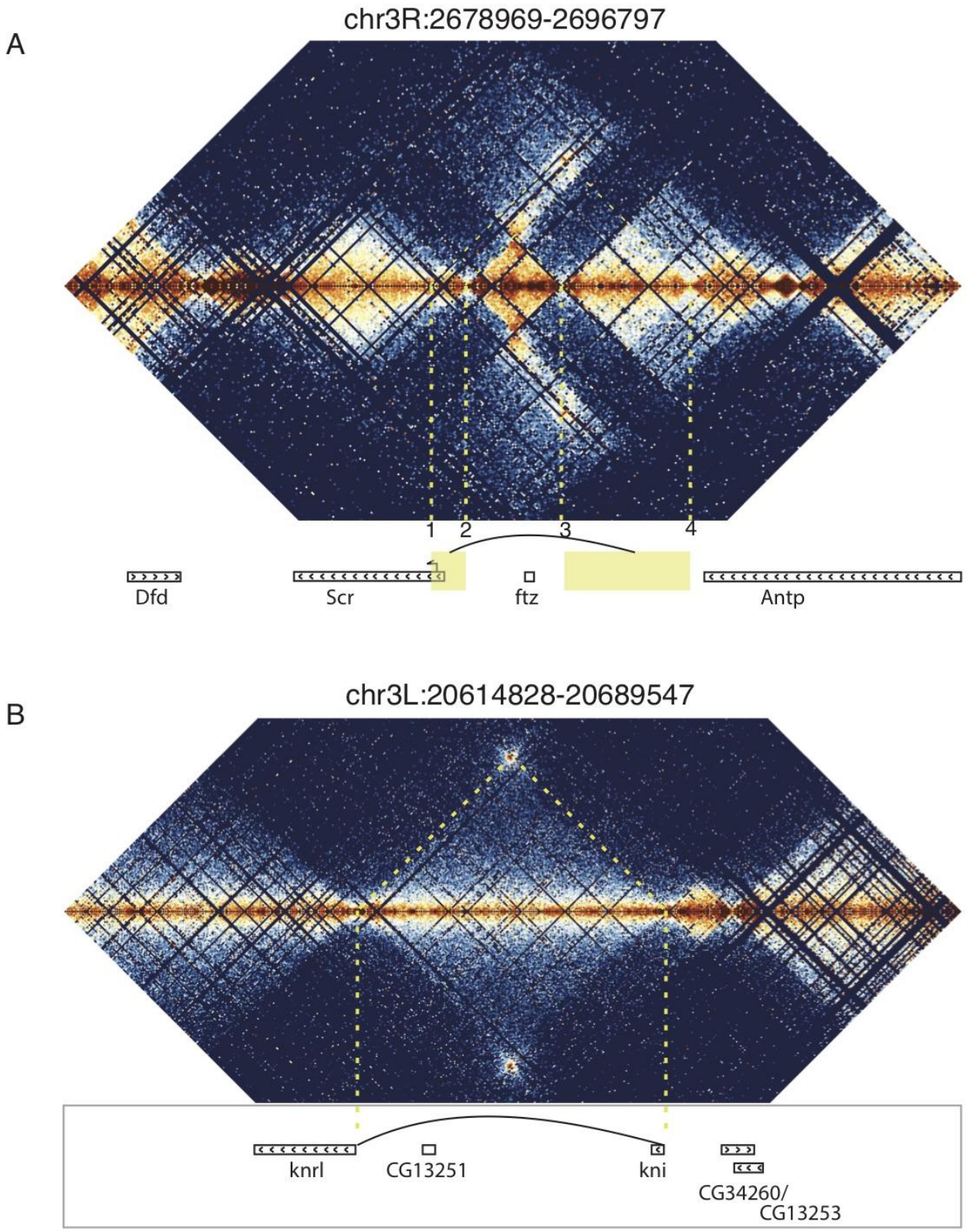


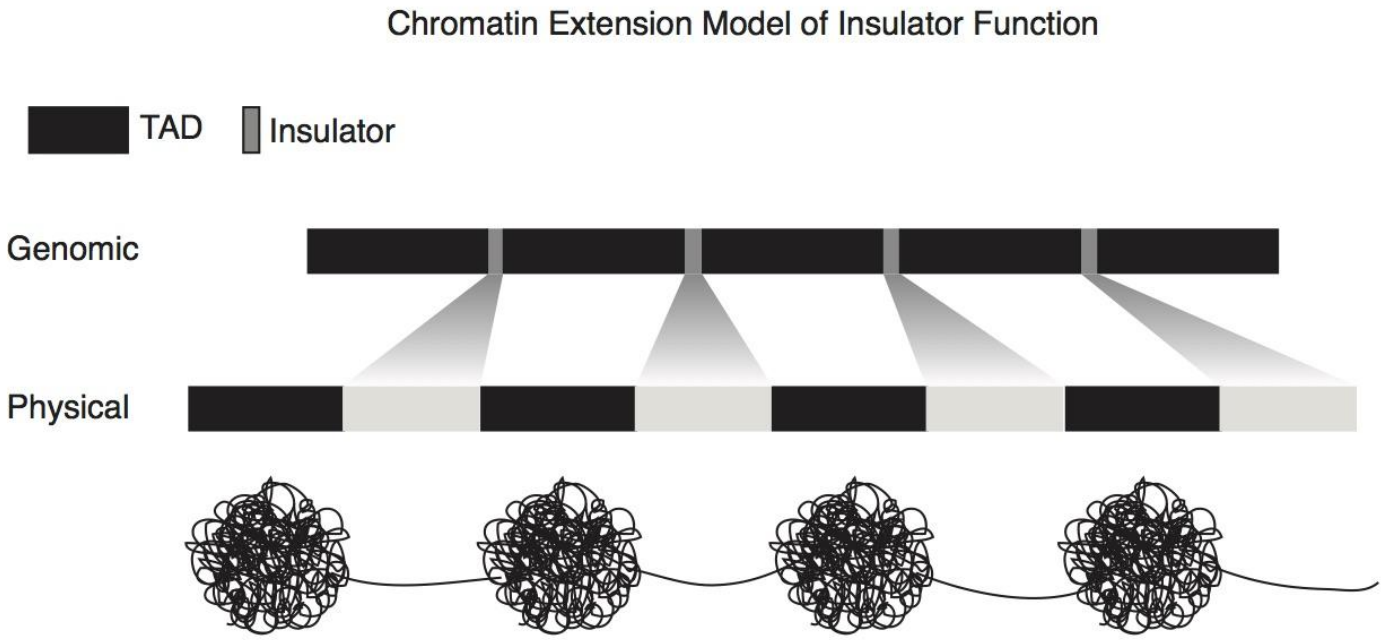
Figure 7: Looping and domain-skipping activity observed in nc14 chromatin. (A) An example of domain-skipping and looping at the *Scr-ftz-Antp* locus. *ftz* is contained within a domain that shows enriched Hi-C interactions between its boundaries, indicative of the formation of a looped domain. Adjacent domains show depleted interaction with the *ftz* domain and enriched interaction with each other, with especially strong contacts between the region containing the *Scr* promoter and characterized *Scr* regulatory elements 3' of the *Antp* locus (Calhoun & Levine, 2003; Calhoun et al., 2002). Dotted lines connect features in the Hi-C map to the genomic locations of genes in this region. **(B)** A strong looping interaction between the *kni* locus and the 5' end of the related *knrl* (*kni*-like) gene. *kni* and *knrl* are known to have identical expression patterns and partially redundant, though distinct domains of biochemical activity (González-Gaitán, Rothe, Wimmer, Taubert, & Jäckle, 1994).

Discussion

While several Hi-C studies in flies have identified enrichments of insulator proteins at TAD boundaries (Eagen et al., 2015; Mourad & Cuvier, 2016; Sexton et al., 2012; Ulianov et al., 2015), none explicitly mapped boundaries to the discrete insulator elements we have identified. This is likely explained by limitations in resolution due to sequencing depth, restriction enzyme cut site frequency, or analysis routines. A common feature that we observe in our Hi-C maps is the presence of large (up to hundreds of kb) TADs characterized by inaccessible chromatin and the presence of few genes. When viewed at lower resolution, these would be the only TADs detected, with intervening regions characterized by clusters of insulator protein binding, transcription, and open chromatin. This closely matches prior descriptions of *Drosophila* chromatin topology. By examining these "inter-TAD" regions at exceptionally high-resolution, we were able to show that they in fact consist of series of smaller TADs, and that the boundaries of both large and small TADs are defined by a common class of insulator elements.

Our most intriguing finding is the association of TAD boundaries with polytene interbands. The implication that these elements are decompacted, extended chromatin regions provides an attractive model in which simple physical separation explains multiple activities associated with insulators, including the ability to block enhancer-promoter interactions, prevent the spread of silenced chromatin, and organize chromatin structure.

Figure 8



400

Figure 8: A chromatin extension model of insulator function. We propose a model in which insulators achieve domain separation by lowering the compaction ratio of bound chromatin, thereby converting the short lengths of insulator DNA (measured in base pairs) into large relative physical distances. By increasing the distance between domains, this model plausibly explains how insulators can achieve their diverse effects, including organizing chromatin structure, blocking

405 enhancer-promoter interactions, and limiting the spread of chromatin silencing states.

A number of prior observations are consistent with the identity of insulators/boundaries as interbands. First, estimates suggest that there are ~5000 interbands constituting 5% of genomic DNA, with an average length of 2 kb (Tatyana Yu Vatolina et al., 2011; Zhimulev, 1996), numbers that are in line with our estimates of boundary element length and number. Second, interbands are associated with insulator proteins, with CP190 appearing to be a constitutive feature of all or nearly all interbands (Tatiana I. Gerasimova, Lei, Bushey, & Corces, 2007; Pai et al., 2004), which is precisely what we observe for boundary elements. Third, interbands and boundary elements are highly sensitive to DNase digestion (T. Yu Vatolina et al., 2011). Fourth, interbands have been shown to contain the promoters and 5' ends of genes (Jamrich, Greenleaf, & Bautz, 1977; Rykowski, Parmelee, Agard, & Sedat, 1988; H. Sass, 1982; H. Sass & Bautz, 1982; Heinz Sass & Bautz, 1982), and we see a strong enrichment for promoters oriented to transcribe away from boundaries, which would place upstream regulatory elements within or near the interband. Finally, deletion of both isoforms of BEAF-32, the second-most highly enriched insulator protein at boundary elements, results in polytene X chromosomes that exhibit loss of banding and are wider and shorter than wild type, consistent with a loss of decompacted BEAF-32-bound regions (Roy et al., 2007). It is possible that interbands in polytene chromosomes result from multiple underlying molecular phenomena, but we believe it is likely that decompacted insulator elements constitute a significant fraction of these structures.

While we and others have not observed frequent looping of insulators in Hi-C data from fly tissue, our model of chromatin compaction at insulators is not mutually exclusive with a role for looping in the function of some insulators. Indeed, we have observed a limited set of cases in which interactions between boundaries seem to organize special genome structures with, at least in the case of the *Scr* locus, clear functional implications. It is likely that additional boundary-associated distal interactions will be found in other tissues and stages of fly development. However, we emphasize that these interactions are exceedingly rare and do not appear to be general features of the function of boundary elements.

Conclusions

The data presented here offer a picture of the structure of the interphase chromatin of *Drosophila* that unifies years of studies of polytene chromosomes with modern genomic methods. In this picture, interphase chromatin consists of alternating stretches of compacted, folded chromatin domains separated by regions of decompacted, stretched regions. The compacted regions vary in size from a few to hundreds of kilobases and correspond to both polytene band regions and

TADs in Hi-C data. Decompacted regions that separate these domains are short DNA elements that are defined by the strong binding of insulator proteins and correspond to polytene interbands and TAD boundaries (insulators). An intuitive view of this structure in a non-polytene context might resemble the well-worn "beads on a string", in which insulator/interband regions are the string and bands/TADs form beads of various sizes. Future work, including experimental manipulation of the sequences underlying these structures, will focus on validating and refining this model and understanding its implications for genome function.

Materials and methods

Embryo collection, sorting, and sectioning

OregonR strain *D. melanogaster* embryos were collected on molasses plates seeded with fresh yeast paste from a population cage and aged to appropriate developmental stages, all at 25 C. Embryos were washed into nitex meshes, dechorionated by treatment with dilute bleach for 2 minutes, dipped briefly (15-20 s) in isopropanol, and gently rocked in fixative solution of (76.5% hexanes, 5% formaldehyde in 1x PBS) for 28-30 minutes. Embryos were then thoroughly washed in PBS with 0.5% triton and stored for no more than 3 days at 4 C. For sample HiC-2/4, embryos were inspected under a light microscope to confirm that the vast majority corresponded to early cellularized blastoderm, and approximately 4000 embryos were used in the Hi-C protocol. For samples HiC-8, 10, 11, 12, 13-16, fixed embryos were hand-sorted under a light microscope as described in (Harrison et al., 2011), using morphological markers to identify early cellularized embryos (nc14, stage 5) or nc12 embryos. For whole embryo experiments, sorted embryos were placed directly into the Hi-C protocol, with no more than 3 days having elapsed since fixation.

For sectioned embryos, hand-sorted embryos of precise developmental stages were first arranged in rows on a block of 1% agarose with bromophenol blue in a shared anterior-posterior orientation, with between 20-40 embryos per block. Aligned embryos were then transferred to the bottom of a plastic embedding mold (Sigma Aldrich E6032), the bottom of which had previously been coated with hexane glue, carefully keeping track of the anterior-posterior orientation of embryos by marking the cup with marker. Embryos were covered with clear frozen section compound (VWR 95057-838) and frozen at -80C for up to two months. Frozen blocks were retrieved from the freezer and embryos rapidly sliced at approximately the mid-point by hand using a standard razor blade under a dissecting microscope. Anterior and posterior halves were separately transferred to microcentrifuge tubes containing ~200 µL PBS with 0.5% triton using an embryo

pick (a tool of mysterious provenance that appears to be a clay sculpting tool). Successful transfer was confirmed visually by the presence of blue embryos which had absorbed bromophenol blue from the agarose block. Between transferring anterior and posterior halves, the pick was washed thoroughly with water and ethanol, and rubbed vigorously with kimwipes. We note that anterior and posterior half samples are precisely matched: samples HiC-13 and 14 contain the anterior and posterior halves (respectively) of the same embryos, and the same is true for HiC-15 and 16.

Hi-C

Experimental procedure: Hi-C experiments were conducted as described in Rao (Rao et al., 2014), with slight modifications. For completeness, we describe the detailed protocol: Embryos (or halves) were suspended in 1X NEB2 buffer (NEB B7002) and homogenized on ice by douncing for several minutes each with the loose and tight dounces. Insoluble material (including nuclei) was pelleted by spinning for 5 minutes at 4500 x g in microcentrifuge cooled to 4 C (all wash steps used these conditions for pelleting). Nuclei were washed twice with 500 µL of 1x NEB2 buffer and then suspended in 125 µL of the same. 42.5 µL of 2% SDS was added and tubes are placed at 65 C for 10 minutes, then returned to ice, followed by addition of 275 µL of 1x NEB2 buffer and 22 µL of 20% Triton X-100, then incubated at room temperature for 5 minutes. Samples were digested overnight with 1500 units of MboI by shaking at 37 C. The next day, samples were washed twice with 1X NEB2, resuspended in 100 µL 1X NEB2, and 15 µL of fill-in mix (1.5 µL 10x NEB2, 0.4 µL each of 10 mM dATP, dGTP, dTTP, 9 µL 0.4 mM biotin-14-dCTP, 2.5 µL 5 U/µL Klenow (NEB M0210), 1 µL water) was added, followed by 1.5 hours at 37 C. Samples were then washed twice with 500 µL 1X ligation buffer (10X: 0.5 M Tris-HCl pH7.4, 0.1M MgCl2, 0.1M DTT), resuspended in 135 µL of the same, then supplemented with 250 µL of ligation mix (25 µL 10x ligation buffer, 25 µL 10% Triton X-100, 2.6 µL 10 mg/ml BSA, 2.6 µL 100 mM ATP, 196 µL water) and 2000 units of T4 DNA ligase (NEB M0202T) and incubated for 2 hours (or overnight) at room temperature. An additional 2000 units of ligase were added, followed by another 2 hours at room temperature. Cross-link reversal was carried out by adding 50 µL of 20 mg/mL proteinase K and incubating overnight at 65 C. An additional 50 µL proteinase K was then added followed by a 2 hour 65 C incubation. 0.1 volumes of 3M NaCl and 2 µL of glycoblue (Thermo Fisher AM9515) were added, then samples were extracted once with one volume of phenol pH 7.9, once with phenol-chloroform pH7.9, then precipitated with 3 volumes of EtOH. Washed pellets were resuspended in 130 µL water and treated with 1 µL of RNase A for 15 minutes at 37 C. DNA was fragmented using the Covaris instrument (Covaris,

Woburn, MA) with peak power 140.0, duty factor 10.0, cycles/burst 200 for 80 seconds. Samples are brought to 300 µL total volume with water.

75 µL of Dynabeads MyOne Streptavidin C1 beads (Thermo Fisher 65001) were washed twice with 400 µL of tween wash buffer (TWB) (2X binding buffer [BB]: 100 µL of 1M Tris-HCl pH8, 20 µL 0.5 M EDTA, 4 mL of 5M NaCl, 5.88 mL water; TWB: 5 ml 2X binding buffer, 50 µL 10% Tween, 4.95 µL water), resuspended in 300 µL 2X BB, then added to 300 µL DNA. Samples were rocked at room temperature for 15 minutes, then washed once with TWB, twice with 1X BB, reclaimed on magnetic stand and resuspended in 100 µL 1X T4 DNA ligase buffer. Samples were then supplemented with end-repair mix (78 µL water, 10 µL 10X T4 DNA ligase buffer with ATP, 2 µL 25 mM dNTPs, 1 µL 10U/µL T4 PNK (Thermo Fisher EK0031), 2 µL 5U/µL Klenow, 3 µL 3U/µL T4 DNA polymerase (Thermo Fisher EP0061)), incubated 30 minutes at room temp, washed as before, washed once with 100 µL 1X NEB2, and resuspended in 90 µL 1X NEB2. dA overhangs were added by adding 2 µL 10mM dATP and 1 µL Klenow exo minus (NEB M0212S), incubating at 37 C for 30 minutes. Beads were washed as before, washed once with 100 µL 1X Quickligase (NEB M2200S) buffer, resuspended in 50 µL 1X Quickligase buffer, then supplemented with 3 µL Illumina adaptors and 1 µL Quickligase. Samples were incubated 15 minutes at room temperature, then were twice with TWB, twice with 1X BB, twice with 200 µL TLE, and resuspend in 50 µL TLE. Beads are stable at 4 C, but were always amplified quickly. 100 µL (or more) of phusion PCR reaction was prepared (50 µL 2X Phusion master mix, 1 µL 100 µM forward primer [5-AATGATACGGCGACCACCGAG-3], 1 µL 100 µM reverse primer [5-CAAGCAGAAGACGGCATACGAG-3], 10 µL of beads with Hi-C library attached, 38 µL water). Reaction was mixed well and split into separate 12 µL reactions. Thermocycler conditions were 16 cycles of 98 C for 30 s, 63 C for 30 s, 72 C for 2 m. Reactions were pooled and loaded on a 2% agarose gel. Fragments corresponding to an insert size of ~300 bp (amplicon size of 421 bp) were excised from the gel, purified with the Zymo Gel DNA Recovery Kit (D4001T, Zymo), and submitted for sequencing at the Vincent J. Coates Genomic Sequencing Laboratory (Berkeley, CA).

Read processing and mapping: Our analysis routine was adapted by examining the approaches of multiple groups (Crane et al., 2015; Lieberman-Aiden et al., 2009; Rao et al., 2014; Sexton et al., 2012) in addition to procedures we developed independently. All analysis was performed with custom Python, R, and Perl scripts except where noted. Single-ends of demultiplexed reads were separately mapped using Bowtie (B. Langmead, Trapnell, Pop, & Salzberg, 2009) (parameters: -m1 --best --strata) to the *D. melanogaster* genome dm3 R5_22 downloaded from flybase on June 11, 2014. Due to the

formation of chimeric reads intrinsic to the Hi-C procedure, reads can fail to properly map if the ligation junction lies within the 100 bp read. To address this, we used an iterative mapping procedure, in which we began by mapping the first 20 nt of the reads (using Bowtie's --trim3 feature). Unique mappings were kept, reads that failed to map were stored, and the procedure was repeated on the multiply-mapping reads, incrementing the length of sequence to map by 7 nt each round (attempt to uniquely map using first 20, first 27, first 34...). We found that this method gave 5-10% increases in yield of mapped reads over a procedure in which we attempted to explicitly detect and trim ligation junctions from reads. Uniquely mapping reads from all iterations were collated as a single file.

Uniquely-mapping single-ends were paired based on read identity, and only pairs with two uniquely-mapping ends were retained. Duplicate reads that shared identical left and right mapping positions were removed. Resulting paired, collapsed, uniquely mapping reads were then inspected for quality. Primary indicators of successful Hi-C libraries were the distance distribution of mapped pairs and the relative frequencies of reads in the four orientations described by (Rao et al., 2014), in-in, in-out, out-in, and out-out. In all of our libraries, we detect some ~3-15% reads that appear to be simple genomic sequence, not the result of a Hi-C ligation event. These reads are readily detected by examining the size distributions of in-out reads (the orientation expected from standard genomic sequence) compared with the other three orientations. The in-out reads have a unique hump of reads showing a distance distribution of ~150-500 bp, varying slightly from sample to sample. In-out reads pairs spanning less than 500 bp were removed from further analysis.

Topological boundary detection: We explored a number of ways of identifying boundaries from directionality data. In the end, the most robust was to use a simple heuristic that at a boundary, by definition, regions to the left show left-bias and regions to the right show right bias. While attempts to derive a boundary score from a comparison of directionality scores upstream and downstream showed susceptibility to noise and artifacts, requiring expected upstream and downstream behavior allowed robust detection of sets of boundary elements. We describe the complete procedure below.

Read counts were assigned to 500 bp bins for all genomic bin combinations within 500 kb of the diagonal. Local directionality scores were calculated for each bin by summing the counts linking the bin to regions in a window encompassing the genomic regions between 1 and 15 kb from the bin (skipping the two proximal 500 bp bins, summing the next 28) upstream and downstream, then taking the log (10) ratio of downstream to upstream. These parameters were determined by visually comparing local directionality scores from a range of inputs to Hi-C heat maps for a number of genomic regions, identifying parameters in which directionality transitions reflected boundaries evident in the heat maps.

We observed high levels of noise in the directionality metric in regions of low read coverage. To suppress these noisy signals, we devised a weighted local directionality score to weight these scores based on the total number of reads used in the calculation. We experimented with a variety of scaling factors a such that w = [read count]^a and found that a weighting of a=0.5 worked well to reduce signal from low-read regions. From these directional scores, sites were first selected for which the mean directionality score of the 5 adjacent upstream bins was less than -2, and the mean for the 5 adjacent downstream bins was greater than 2. Boundary scores were assigned to resulting bins by subtracting the sum of the directionality scores for the 5 adjacent upstream bins from the 5 adjacent downstream bins. An issue with this scoring system is that bins that lack MboI sites can cause inflated directionality scores in adjacent regions. To address this, we simply assigned a boundary score of 0 to any bin with more than 1 such bin in its radius. The resulting distribution of boundary scores is dominated by series of consecutive bins with large boundary score maximums, which is uninformative since these scores are essentially derived from the same data (window shifted by one bin). We therefore merged adjacent bins that passed the cutoff and selected only the bin with the maximum boundary score within a contiguous block. By sorting the resulting table on the boundary score, we were able to select sets of candidate boundaries of various strengths for analysis.

In additional to these computationally-identified boundary locations, we manually called boundaries for the entire genome. An R script serially displayed Hi-C heat maps of 250 kb genomic windows and recorded the genomic coordinates of mouse clicks made at visually-identified boundaries. The human caller was unaware of any features of the regions examined other than the Hi-C maps, and was unaware of the locations being displayed in a given plot.

Sequence analysis: We used simple custom Python scripts to count the occurrences of all words of length 4, 5, 6 and 7 in 500 bp windows from 10,000 bp upstream to 10,000 bp downstream of the 500 bp window identified as a boundary. We then computed a simple enrichment score for each unique word equal to the counts of that word and its reverse complement in the boundary divided by the mean counts for the word and its reverse complement in the remaining windows. We noticed that many of the words identified as enriched in this analysis were also enriched in the 500 bp bins immediately flanking the boundary. We therefore updated our enrichment score for each word to be the mean of the counts of the word and its reverse complement in the boundary and the 500 bp bins immediately adjacent to it (three bins in total) divided by the mean counts of the word and its reverse complement in the remaining 38 bins. Counts and scores for all words are provided in the supplemental materials.

ATAC-seq

Experimental procedure: Early nc14 embryos were placed in ATAC-seq lysis buffer (Buenrostro et al., 2013) without detergent, with 5% glycerol added. Embryos were then taken out of the freezing solution and placed onto a glass slide which was then put on dry ice for 2 minutes. Once embryos were completely frozen, the glass slide was removed and embryos were sliced with a razor blade chilled in dry ice. Once sliced embryo halves were moved to tubes containing ATAC-seq lysis buffer with 0.15mM spermine added to help stabilize chromatin. Embryo halves were then homogenized using single use plastic pestles. IGEPal CA-630 was added to a final concentration of 0.1%. After a 10 minute incubation nuclei were spun down and resuspended in water. Twenty halves were added to the transposition reaction containing 25 µl of 2x TD buffer (Illumina), and 2.5ul of Tn5 enzyme (Illumina) and the reaction was incubated at 37°C for 30 minutes as in (Buenrostro et al., 2013). Transposed DNA was purified using Qiagen Minelute kit. Libraries were then amplified using phusion 2x master mix (NEB) and indexed primers from Illumina. Libraries were then purified with Ampure Beads and sequenced on the Hiseq4000 using 100 bp paired end reads.

Analysis: Fastq files were aligned to Drosophila Dm3 genome with Bowtie2 (Ben Langmead & Salzberg, 2012) using the following parameters: -5 5 -3 5 -N 1 -X 2000 --local --very-sensitive-local. Sam files were then sorted and converted to Bam files using Samtools (H. Li et al., 2009), only keeping mapped, properly paired reads with a MAPq score of 30 or higher using -q 30. Bams were then converted to Bed files with bedtools and shifted using a custom shell script to reflect a 4bp increase on the plus strand and a 5bp decrease on the minus strand as recommended by Buenrostro et al. 2013. Finally shifted bed files were converted into wig files using custom scripts and wig files which were uploaded to the genome browser. Wig files were normalized to reflect 10 million mapped reads.

Acknowledgements

We are especially thankful to Emily Brown for her assistance in adapting Hi-C to fly embryos, to Xiao-Yong Li for help with embryo sorting and with optimizing the fixation and chromatin isolation protocols, and to Steven Kuntz for assistance with developing embryo sectioning protocols. We thank Mustafa Mir, Xavier Darzacq and members of the

Eisen and Darzacq labs for critical discussions and advice supplied throughout the work. MS was supported by an American Cancer Society postdoctoral fellowship (126730-PF-14-256-01-DDC), JH was supported by the National Science Foundation Graduate Research Fellows Program, and the work was supported by an HHMI investigator award to ME.

Competing Interests

The authors declare that they have no competing interests.

References

Balbiani, E. G. (1881). Sur la structure du noyau des cellules salivares chez les larves de Chironomus. *Zoologischer Anzeiger*, *4*, 637–641, 662–666.

Balbiani, E. G. (1890). Sur la structure intime du noyau du Loxophyllum meleagris. *Zoologischer Anzeiger*, *13*, 110–115, 132–136.

Beagrie, R. A., Scialdone, A., Schueler, M., Kraemer, D. C. A., Chotalia, M., Xie, S. Q., … Pombo, A. (2017). Complex multi-enhancer contacts captured by genome architecture mapping. *Nature*. https://doi.org/10.1038/nature21411

Bell, A. C., West, A. G., & Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*, *98*(3), 387–396. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/10458613

Belyaeva, E. S., & Zhimulev, I. F. (1994). Molecular and Cytogenetical Characterization of the 1OA1-2 Band and Adjoining. *Genetics*, *136*(1), 1063–1073.

Benyajati, C., & Worcel, A. (1976). Isolation, characterization, and structure of the folded interphase genome of Drosophila melanogaster. *Cell*, *9*(3), 393–407. https://doi.org/10.1016/0092-8674(76)90084-2

Berkaeva, M., Demakov, S., Schwartz, Y. B., & Zhimulev, I. (2009). Functional analysis of Drosophila polytene chromosomes decompacted unit: The interband. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, *17*(6), 745–754. https://doi.org/10.1007/s10577-009-9065-7

Blanton, J., Gaszner, M., & Schedl, P. (2003). Protein:protein interactions and the pairing of boundary elements in vivo.

Genes & Development, *17*(5), 664–675. https://doi.org/10.1101/gad.1052003

Blythe, S. A., & Wieschaus, E. F. (2016). Establishment and maintenance of heritable chromatin structure during early Drosophila embryogenesis. *eLife*, *5*. https://doi.org/10.7554/eLife.20148

Bridges, C. B. (1934). Salivary chromosome maps with a key to the banding of the chromosomes of Drosophila melanogaster. *The Journal of Heredity*, 60–64.

Büchner, K., Roth, P., Schotta, G., Krauss, V., Saumweber, H., Reuter, G., & Dorn, R. (2000). Genetic and molecular complexity of the position effect variegation modifier mod(mdg4) in Drosophila. *Genetics*, *155*(1), 141–157. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/10790390

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*, *10*(12), 1213–1218. https://doi.org/10.1038/nmeth.2688

Byrd, K., & Corces, V. G. (2003). Visualization of chromatin domains created by the gypsy insulator of Drosophila. *The Journal of Cell Biology*, *162*(4), 565–574. https://doi.org/10.1083/jcb.200305013

Calhoun, V. C., & Levine, M. (2003). Long-range enhancer-promoter interactions in the Scr-Antp interval of the Drosophila Antennapedia complex. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(17), 9878–9883. https://doi.org/10.1073/pnas.1233791100

Calhoun, V. C., Stathopoulos, A., & Levine, M. (2002). Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(14), 9243–9247. https://doi.org/10.1073/pnas.142291299

Contrino, S., Smith, R. N., Butano, D., Carr, A., Hu, F., Lyne, R., … Micklem, G. (2012). modMine: Flexible access to modENCODE data. *Nucleic Acids Research*, *40*(D1), 1082–1088. https://doi.org/10.1093/nar/gkr921

Crane, E., Bian, Q., McCord, R. P., Lajoie, B. R., Wheeler, B. S., Ralston, E. J., … Meyer, B. J. (2015). Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature*, *523*(7559), 240–244. https://doi.org/10.1038/nature14450

Csink, a. K., & Henikoff, S. (1998). Large-scale chromosomal movements during interphase progression in Drosophila. *The Journal of Cell Biology*, *143*(1), 13–22. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2132807&tool=pmcentrez&rendertype=abstract

Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. *Science*, *295*(5558),

1306–1311. https://doi.org/10.1126/science.1067799

Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y. J., Lee, C., … Noble, W. S. (2010). A three-dimensional model of the yeast genome. *Nature*, *465*(7296), 363–367. https://doi.org/10.1038/nature08973

Eagen, K. P., Hartl, T. A., & Kornberg, R. D. (2015). Stable Chromosome Condensation Revealed by Chromosome Conformation Capture. *Cell*, *163*(4), 934–946. https://doi.org/10.1016/j.cell.2015.10.026

Eggert, H., Gortchakov, A., & Saumweber, H. (2004). Identification of the Drosophila interband-specific protein Z4 as a DNA-binding zinc-finger protein determining chromosomal structure. *Journal of Cell Science*, *117*(Pt 18), 4253–4264. https://doi.org/10.1242/jcs.01292

Foe, V. E., & Alberts, B. M. (1983). Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in Drosophila embryogenesis. *Journal of Cell Science*, *61*, 31–70. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/6411748

Fujioka, M., Wu, X., & Jaynes, J. B. (2009). A chromatin insulator mediates transgene homing and very long-range enhancer-promoter communication. *Development* , *136*(18), 3077–3087. https://doi.org/10.1242/dev.036467

Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. B., … Ruan, Y. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*, *462*(7269), 58–64. https://doi.org/10.1038/nature08497

Gaszner, M., Vazquez, J., & Schedl, P. (1999). The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interaction. *Genes & Development*, *13*(16), 2098–2107. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/10465787

Gause, M., Morcillo, P., & Dorsett, D. (2001). Insulation of Enhancer-Promoter Communication by a Gypsy Transposon Insert in the Drosophila cut Gene: Cooperation between Suppressor of Hairy-wing and Modifier of mdg4 Proteins. *Molecular and Cellular Biology*, *21*(14), 4807–4817. https://doi.org/10.1128/MCB.21.14.4807-4817.2001

Gavrilov, A. a., Gushchanskaya, E. S., Strelkova, O., Zhironkina, O., Kireev, I. I., Iarovaia, O. V., & Razin, S. V. (2013). Disclosure of a structural milieu for the proximity ligation reveals the elusive nature of an active chromatin hub. *Nucleic Acids Research*, *41*(6), 3563–3575. https://doi.org/10.1093/nar/gkt067

Gavrilov, A., Razin, S. V., & Cavalli, G. (2015). In vivo formaldehyde cross-linking: It is time for black box analysis. *Briefings in Functional Genomics*, *14*(2), 163–165. https://doi.org/10.1093/bfgp/elu037

Gerasimova, T. I., Gdula, D. A., Gerasimov, D. V., Simonova, O., & Corces, V. G. (1995). A drosophila protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation. *Cell*, *82*(4), 587–597.

https://doi.org/10.1016/0092-8674(95)90031-4

Gerasimova, T. I., Lei, E. P., Bushey, A. M., & Corces, V. G. (2007). Coordinated control of dCTCF and gypsy chromatin insulators in Drosophila. *Molecular Cell*, *28*(5), 761–772. https://doi.org/10.1016/j.molcel.2007.09.024

Geyer, P. K., & Corces, V. G. (1992). DNA position-specific repression of transcription by a Drosophila zinc finger protein. *Genes and Development*, *6*(10), 1865–1873. https://doi.org/10.1101/gad.6.10.1865

Ghosh, D., Gerasimova, T. I., & Corces, V. G. (2001). Interactions between the Su(Hw) and Mod(mdg4) proteins required for gypsy insulator function. *The EMBO Journal*, *20*(10), 2518–2527. https://doi.org/10.1093/emboj/20.10.2518

Gilbert, M. K., Tan, Y. Y., & Hart, C. M. (2006). The drosophila boundary element-associated factors BEAF-32A and BEAF-32B affect chromatin structure. *Genetics*, *173*(3), 1365–1375. https://doi.org/10.1534/genetics.106.056002

Golovnin, A., Mazur, A., Kopantseva, M., Kurshakova, M., Gulak, P. V., Gilmore, B., … Georgiev, P. (2007). Integrity of the Mod(mdg4)-67.2 BTB Domain Is Critical to Insulator Function in Drosophila melanogaster. *Molecular and Cellular Biology*, *27*(3), 963–974. https://doi.org/10.1128/MCB.00795-06

González-Gaitán, M., Rothe, M., Wimmer, E. A., Taubert, H., & Jäckle, H. (1994). Redundant functions of the genes knirps and knirps-related for the establishment of anterior Drosophila head structures. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(18), 8567–8571. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8078924

Gortchakov, A. A., Eggert, H., Gan, M., Mattow, J., Zhimulev, I. F., & Saumweber, H. (2005). Chriz, a chromodomain protein specific for the interbands of Drosophila melanogaster polytene chromosomes. *Chromosoma*, *114*(1), 54–66. https://doi.org/10.1007/s00412-005-0339-3

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D. U., … Wu, Q. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*, *162*(4), 900–910. https://doi.org/10.1016/j.cell.2015.07.038

G. Wilkie A Shermoen P. O'Farrell, I. D. (1999). Transcribed genes are localized according to chromosomal position within polarized Drosophila embryonic nuclei. *Current Biology: CB*, *9*(21), 1263–1266. https://doi.org/10.1016/j.mib.2008.09.004.Bacteriophage

Harrison, M. M., Li, X.-Y., Kaplan, T., Botchan, M. R., & Eisen, M. B. (2011). Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genetics*, *7*(10), e1002266. https://doi.org/10.1371/journal.pgen.1002266

Heger, P., George, R., & Wiehe, T. (2013). Successive gain of insulator proteins in arthropod evolution. *Evolution; International Journal of Organic Evolution*, *67*(10), 2945–2956. https://doi.org/10.1111/evo.12155

Heitz, E., & Bauer, H. (1933). Beweise fur die Chromosomennatur der Kernschleifen in den Knauelkernen von Bibio hortulanus. *Zeitschrift Fur Zellforschung*, *17*, 67–82.

Holdridge, C., & Dorsett, D. (1991). Repression of hsp70 heat shock gene transcription by the suppressor of hairy-wing protein of Drosophila melanogaster. *Molecular and Cellular Biology*, *11*(4), 1894–1900. https://doi.org/10.1128/MCB.11.4.1894.Updated

Hou, C., Li, L., Qin, Z. S., & Corces, V. G. (2012). Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Molecular Cell*, *48*(3), 471–484. https://doi.org/10.1016/j.molcel.2012.08.031

Hug, C. B., Grimaldi, A. G., Kruse, K., & Vaquerizas, J. M. (2017). Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell*, *169*(2), 216–228.e19. https://doi.org/10.1016/j.cell.2017.03.024

Jamrich, M., Greenleaf, A. L., & Bautz, E. K. (1977). Localization of RNA polymerase in polytene chromosomes of Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(5), 2079–2083. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/405671

Kahn, T. G., Schwartz, Y. B., Dellino, G. I., & Pirrotta, V. (2006). Polycomb complexes and the propagation of the methylation mark at the drosophila Ubx gene. *The Journal of Biological Chemistry*, *281*(39), 29064–29075. https://doi.org/10.1074/jbc.M605430200

Kellum, R., & Schedl, P. (1991). A position-effect assay for boundaries of higher order chromosomal domains. *Cell*, *64*(5), 941–950. https://doi.org/10.1016/0092-8674(91)90318-S

Kellum, R., & Schedl, P. (1992). A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Molecular and Cellular Biology*, *12*(5), 2424–2431. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1569958

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human genome browser at UCSC. *Genome Research*, *12*(6), 996–1006. https://doi.org/10.1101/gr.229102. Article published online before print in May 2002

King, R. L., & Beams, H. W. (1934). Somatic synapsis in Chironomus, with special reference to the individuality of chromosomes. *Journal of Morphology*, *5*(3), 577–591.

Kravchenko, E., Savitskaya, E., Kravchuk, O., Parshikov, A., Georgiev, P., & Savitsky, M. (2005). Pairing between gypsy Insulators Facilitates the Enhancer Action in trans throughout the Drosophila Genome. *Molecular and Cellular Biology*, *25*(21), 9283–9291. https://doi.org/10.1128/MCB.25.21.9283-9291.2005

Kyrchanova, O., & Georgiev, P. (2014). Chromatin insulators and long-distance interactions in Drosophila. *FEBS Letters*, *588*(1), 8–14. https://doi.org/10.1016/j.febslet.2013.10.039

Laird, C. D., & Chooi, W. Y. (1976). Morphology of transcription units in Drosophila melanogaster. *Chromosoma*, *58*(2), 193–218. https://doi.org/10.1007/BF00701359

Laird, C. D., Wilkinson, L. E., Foe, V. E., & Chooi, W. Y. (1976). Analysis of chromatin-associated fiber arrays. *Chromosoma*, *58*(2), 169–190. https://doi.org/10.1007/BF00701357

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

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

Lefevre, G. (1976). A photographic representation and interpretation of the polytene chromosomes of Drosophila melanogaster salivary glands. *The Genetics and Biology of Drosophila*.

Lewis, E. (1981). Developmental genetics of the bithorax complex in Drosophila. *Developmental Biology Using Purified Genes (ed. D.D. Brown and C.F. Fox)*, 189–208.

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., … Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, *326*(5950), 289–293. https://doi.org/10.1126/science.1181369

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, L., Lyu, X., Hou, C., Takenaka, N., Nguyen, H. Q., Ong, C. T., … Corces, V. G. (2015). Widespread Rearrangement of 3D Chromatin Organization Underlies Polycomb-Mediated Stress-Induced Silencing. *Molecular Cell*, *58*(2), 216–231. https://doi.org/10.1016/j.molcel.2015.02.023

Lindsley, D. L., & Grell, E. H. (1968). Genetic variations of Drosophila melanogaster. *Carnegie Inst. Wash. Publication*. Retrieved from http://garfield.library.upenn.edu/classics1983/A1983QS51100002.pdf

Li, X.-Y., Harrison, M. M., Villalta, J. E., Kaplan, T., & Eisen, M. B. (2014). Establishment of regions of genomic activity during the Drosophila maternal to zygotic transition. *eLife*, *3*, e03737. https://doi.org/10.7554/eLife.03737

Li, X. Y., MacArthur, S., Bourgon, R., Nix, D., Pollard, D. A., Iyer, V. N., … Biggin, M. D. (2008). Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. *PLoS Biology*, *6*(2), 0365–0388. https://doi.org/10.1371/journal.pbio.0060027

Li, X.-Y., Thomas, S., Sabo, P. J., Eisen, M. B., Stamatoyannopoulos, J. a., & Biggin, M. D. (2011). The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. *Genome Biology*, *12*(4), R34. https://doi.org/10.1186/gb-2011-12-4-r34

Lowenstein, M. G., Goddard, T. D., & Sedat, J. W. (2004). Long-range interphase chromosome organization in Drosophila: a study using color barcoded fluorescence in situ hybridization and structural clustering analysis. *Molecular Biology of the Cell*, *15*(12), 5678–5692. https://doi.org/10.1091/mbc.E04-04-0289

Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., … Mundlos, S. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, *161*(5), 1012–1025. https://doi.org/10.1016/j.cell.2015.04.004

Mallin, D. R., Myung, J. S., Patton, J. S., & Geyer, P. K. (1998). Polycomb group repression is blocked by the Drosophila suppressor of Hairy-wing [SU(HW)] insulator. *Genetics*, *148*(1), 331–339. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9475743

McKnight, S. L., & Miller, O. L. (1976). Ultrastructural patterns of RNA synthesis during early embryogenesis of Drosophila melanogaster. *Cell*, *8*(2), 305–319. https://doi.org/10.1016/0092-8674(76)90014-3

McKnight, S. L., & Miller, O. L. (1977). Electron microscopic analysis of chromatin replication in the cellular blastoderm drosophila melanogaster embryo. *Cell*, *12*(3), 795–804. https://doi.org/10.1016/0092-8674(77)90278-1

Melnikova, L., Juge, F., Gruzdeva, N., Mazur, A., Cavalli, G., & Georgiev, P. (2004). Interaction between the GAGA factor and Mod(mdg4) proteins promotes insulator bypass in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(41), 14806–14811. https://doi.org/10.1073/pnas.0403959101

Mohan, M., Bartkuhn, M., Herold, M., Philippen, A., Heinl, N., Bardenhagen, I., … Renkawitz, R. (2007). The Drosophila insulator proteins CTCF and CP190 link enhancer blocking to body patterning. *The EMBO Journal*, *26*(19), 4203–4214. https://doi.org/10.1038/sj.emboj.7601851

Moon, H., Filippova, G., Loukinov, D., Pugacheva, E., Chen, Q., Smith, S. T., … Lobanenkov, V. (2005). CTCF is

conserved from Drosophila to humans and confers enhancer blocking of the Fab-8 insulator. *EMBO Reports*, *6*(2), 165–170. https://doi.org/10.1038/sj.embor.7400334

Mourad, R., & Cuvier, O. (2016). Computational Identification of Genomic Features That Influence 3D Chromatin Domain Formation. *PLoS Computational Biology*, *12*(5), e1004908. https://doi.org/10.1371/journal.pcbi.1004908

Nègre, N., Brown, C. D., Shah, P. K., Kheradpour, P., Morrison, C. a., Henikoff, J. G., … White, K. P. (2010). A comprehensive map of insulator elements for the Drosophila genome. *PLoS Genetics*, *6*(1), e1000814. https://doi.org/10.1371/journal.pgen.1000814

Newport, J., & Kirschner, M. (1982). A major developmental transition in early Xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. *Cell*, *30*(3), 675–686. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/6183003

Pai, C. Y., Lei, E. P., Ghosh, D., & Corces, V. G. (2004). The centrosomal protein CP190 is a component of the gypsy chromatin insulator. *Molecular Cell*, *16*(5), 737–748. https://doi.org/10.1016/j.molcel.2004.11.004

Painter, T. S. (1935). The morphology of the third chromosome in the salivary gland of rosophila melanogaster and a new cytological map of this element. *Genetics*.

Parkhurst, S. M., & Corces, V. G. (1985). Forked, gypsys, and suppressors in Drosophila. *Cell*, *41*(2), 429–437. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/2985277

Parkhurst, S. M., & Corces, V. G. (1986). Interactions among the gypsy transposable element and the yellow and the suppressor of hairy-wing loci in Drosophila melanogaster. *Molecular and Cellular Biology*, *6*(1), 47–53. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/3023836

Parkhurst, S. M., Harrison, D. A., Remington, M. P., Spana, C., Kelley, R. L., Coyne, R. S., & Corces, V. G. (1988). The Drosophila su(Hw) gene, which controls the phenotypic effect of the gypsy transposable element, encodes a putative DNA-binding protein. *Genes & Development*, *2*(10), 1205–1215. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/2462523

Rabinowitz, M. (1941). Studies on the cytology and early embryology of the egg of Drosophila melanogaster. *Journal of Morphology*, *69*. https://doi.org/10.1002/jmor.1050690102

Ramirez, F., Bhardwaj, V., Villaveces, J., Arrigoni, L., Gruening, B. A., Lam, K. C., … Manke, T. (2017). High-resolution TADs reveal DNA sequences underlying genome organization in flies. *bioRxiv*, *49*(0), 115063. https://doi.org/10.1101/115063

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., … Aiden, E. L. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*, 1–16. https://doi.org/10.1016/j.cell.2014.11.021

Recillas-Targa, F., Pikaart, M. J., Burgess-Beusse, B., Bell, A. C., Litt, M. D., West, A. G., … Felsenfeld, G. (2002). Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(10), 6883–6888. https://doi.org/10.1073/pnas.102179399

Roseman, R. R., Pirrotta, V., & Geyer, P. K. (1993). The su(Hw) protein insulates expression of the Drosophila melanogaster white gene from chromosomal position-effects. *The EMBO Journal*, *12*(2), 435–442. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=413226&tool=pmcentrez&rendertype=abstract

Roy, S., Gilbert, M. K., & Hart, C. M. (2007). Characterization of BEAF mutations isolated by homologous recombination in Drosophila. *Genetics*, *176*(2), 801–813. https://doi.org/10.1534/genetics.106.068056

Rykowski, M. C., Parmelee, S. J., Agard, D. A., & Sedat, J. W. (1988). Precise determination of the molecular limits of a polytene chromosome band: regulatory sequences for the Notch gene are in the interband. *Cell*, *54*(4), 461–472. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/3135939

Sass, H. (1982). RNA polymerase B in polytene chromosomes: immunofluorescent and autoradiographic analysis during stimulated and repressed RNA synthesis. *Cell*, *28*(2), 269–278. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/7037199

Sass, H., & Bautz, E. K. (1982). Interbands of polytene chromosomes: binding sites and start points for RNA polymerase B (II). *Chromosoma*, *86*(1), 77–93. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/6756817

Sass, H., & Bautz, E. K. F. (1982). Immunoelectron microscopic localization of RNA polymerase B on isolated polytene chromosomes of Chironomus tentans. *Chromosoma*, *85*(5), 633–642. https://doi.org/10.1007/BF00330777

Scott, K. C., Taubman, A. D., & Geyer, P. K. (1999). Enhancer blocking by the Drosophila gypsy insulator depends upon insulator anatomy and enhancer strength. *Genetics*, *153*(2), 787–798. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/10511558

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., … Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, *148*(3), 458–472. https://doi.org/10.1016/j.cell.2012.01.010

Sigrist, C. J. A., & Pirrotta, V. (1997). Chromatin insulator elements block the silencing of a target gene by the Drosophila polycomb response element (PRE) but allow trans interactions between PREs on different chromosomes. *Genetics*, *147*(1), 209–221. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9286681

Spana, C., Harrison, D. A., & Corces, V. G. (1988). The Drosophila melanogaster suppressor of Hairy-wing protein binds to specific sequences of the gypsy retrotransposon. *Genes & Development*, *2*(11), 1414–1423. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/2850261

Tadros, W., & Lipshitz, H. D. (2009). The maternal-to-zygotic transition: a play in two acts. *Development* , *136*(18), 3033–3042. https://doi.org/10.1242/dev.033183

Ulianov, S. V., Khrameeva, E. E., Gavrilov, A. a., Flyamer, I. M., Kos, P., Mikhaleva, E. a., … Razin, S. V. (2015). Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Research*, *26*(1), 70–84. https://doi.org/10.1101/gr.196006.115

Van Bortle, K., Nichols, M. H., Li, L., Ong, C.-T., Takenaka, N., Qin, Z. S., & Corces, V. G. (2014). Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biology*, *15*(5), R82. https://doi.org/10.1186/gb-2014-15-5-r82

Vatolina, T. Y., Boldyreva, L. V., Demakova, O. V., Demakov, S. A., Kokoza, E. B., Semeshin, V. F., … Zhimulev, I. F. (2011). Identical functional organization of nonpolytene and polytene chromosomes in Drosophila melanogaster. *PloS One*, *6*(10). https://doi.org/10.1371/journal.pone.0025960

Vatolina, T. Y., Demakov, S. A., Semeshin, V. F., Makunin, I. V., Babenko, V. N., Belyaeva, E. S., & Zhimulev, I. F. (2011). Identification and molecular genetic characterization of the polytene chromosome interbands in Drosophila melanogaster. *Russian Journal of Genetics*, *47*(5), 521–532. https://doi.org/10.1134/S1022795411040144

Vlassova, I. E., Umbetova, G. H., Zimmermann, V. H., Alonso, C., Belyaeva, E. S., & Zhimulev, I. F. (1985). Immunofluorescence localization of DNA:RNA hybrids in Drosophila melanogaster polytene chromosomes. *Chromosoma*, *91*(3-4), 251–258. https://doi.org/10.1007/BF00328220

Vogelmann, J., Le Gall, A., Dejardin, S., Allemand, F., Gamot, A., Labesse, G., … Nöllmann, M. (2014). Chromatin Insulator Factors Involved in Long-Range DNA Interactions and Their Role in the Folding of the Drosophila Genome. *PLoS Genetics*, *10*(8), e1004544. https://doi.org/10.1371/journal.pgen.1004544

Yang, J., & Corces, V. G. (2012). Insulators, long-range interactions, and genome function. *Current Opinion in Genetics & Development*, *22*(2), 86–92. https://doi.org/10.1016/j.gde.2011.12.007

Zhao, K., Hart, C. M., & Laemmli, U. K. (1995). Visualization of chromosomal domains with boundary element-associated factor BEAF-32. *Cell*, *81*(6), 879–889. https://doi.org/10.1016/0092-8674(95)90008-X

Zhimulev, I. F. (1996). Morphology and Structure of Polytene Chromosomes. *Advances in Genetics*, *34*, 1–490. https://doi.org/10.1016/S0065-2660(08)60533-7

885

Supplemental Material

Supplemental figures S1-S8:

S1: High resolution Hi-C maps of additional example genomic regions from stage 5 *Drosophila melanogaster* embryos.

S2: Genomic signals around topological boundaries, self-sorted.

S3: Genomic signals around H3K4me3 peaks

S4: Directionality around peaks of genomic features

S5: Developmental time series of DNase accessibility at TAD boundaries

S6: Distal chromatin contacts in stage 5 embryos

S7: TAD structure corresponds to mapped polytene structure at the *Notch* locus

S8: Sequence features of TAD boundary elements

Supplemental Tables S1-S7:

S1: List of samples used to generate Hi-C libraries

S2: Sequencing and processing statistics for Hi-C libraries

S3: Results of logistic regression analysis of high confidence boundaries with various genomic datasets

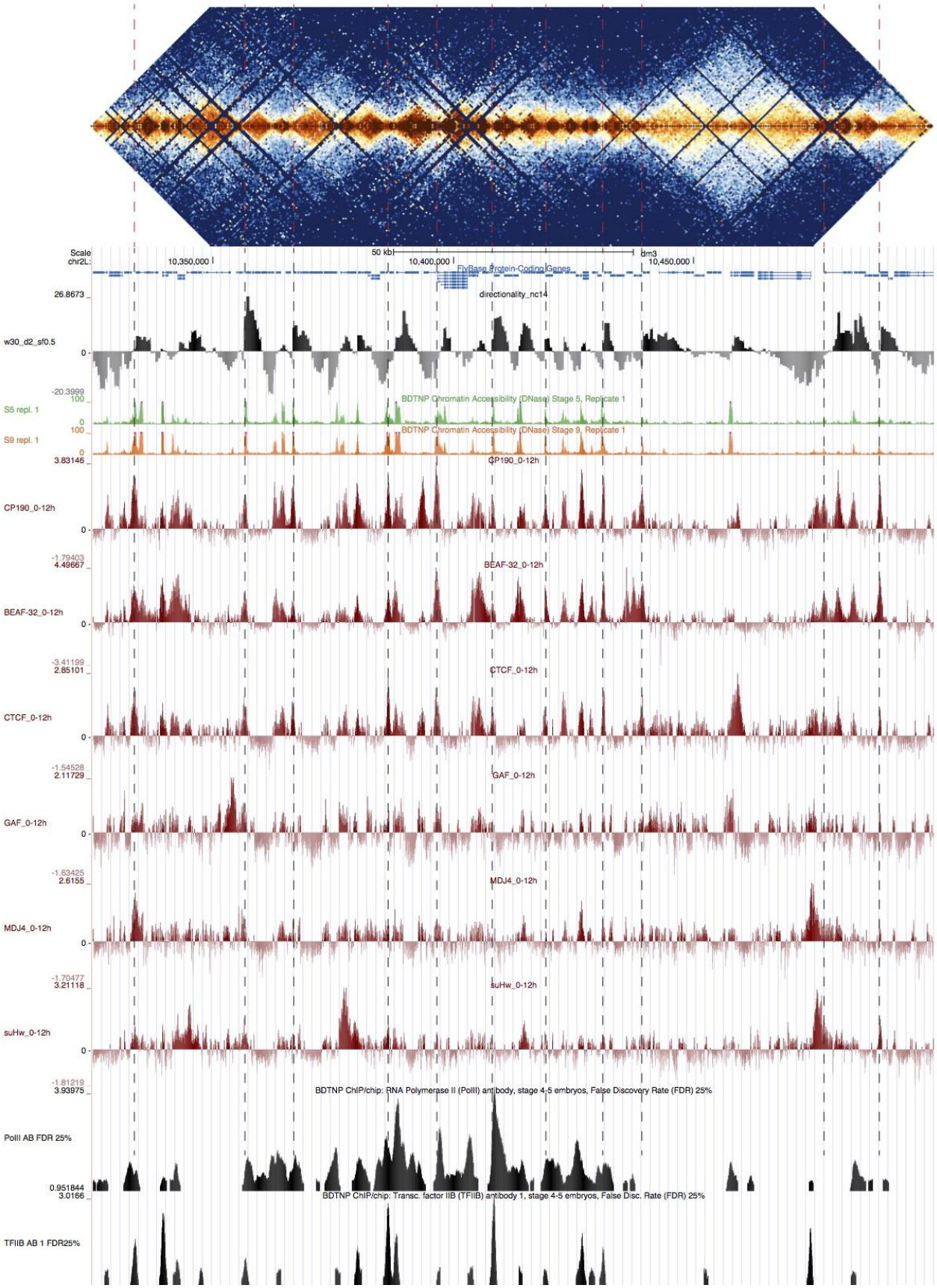
S4: Location of loops and domain-skipping identified in the stage 5 genome

S5: Manually called boundaries

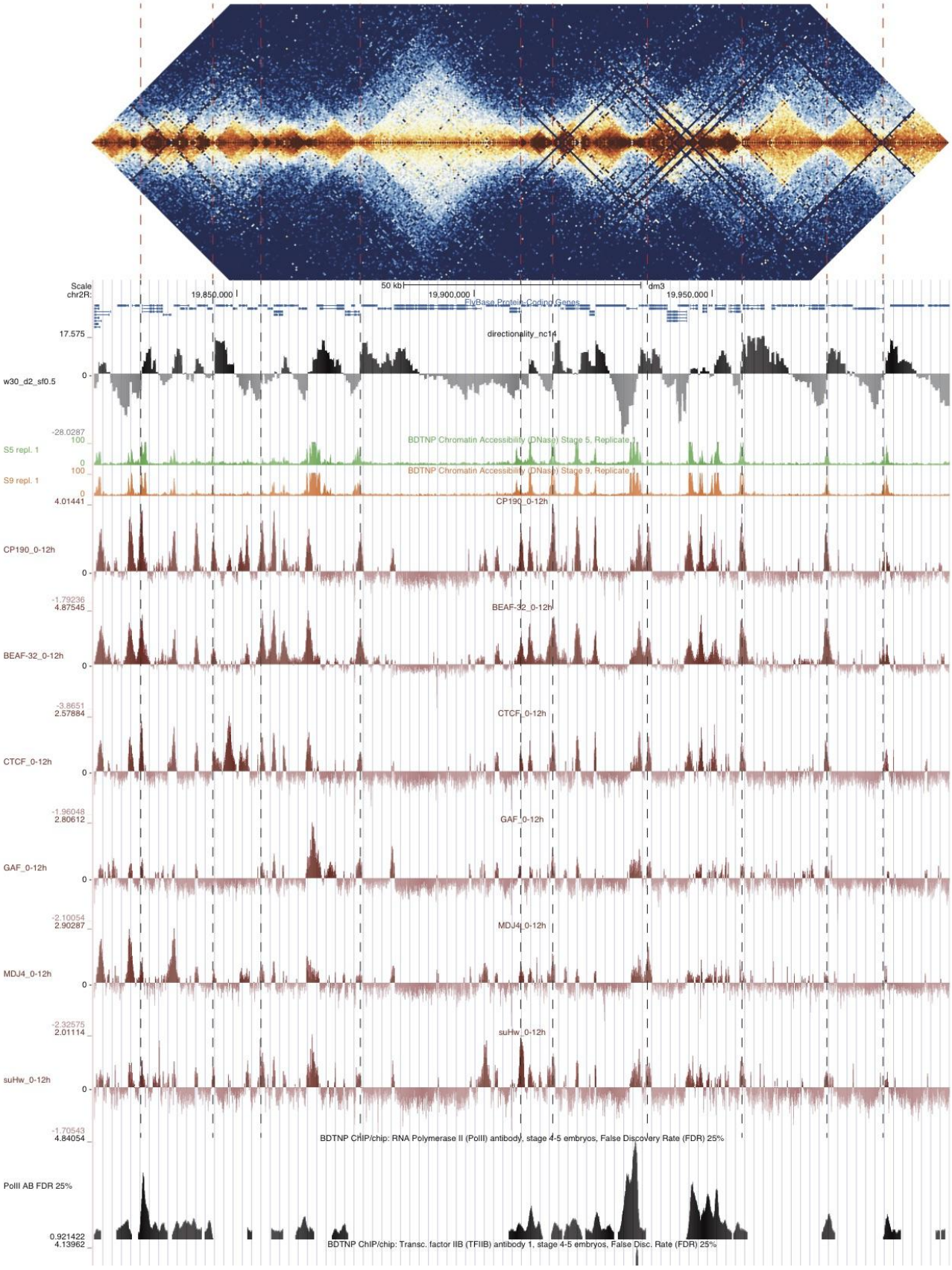
S6: Computationally identified boundaries

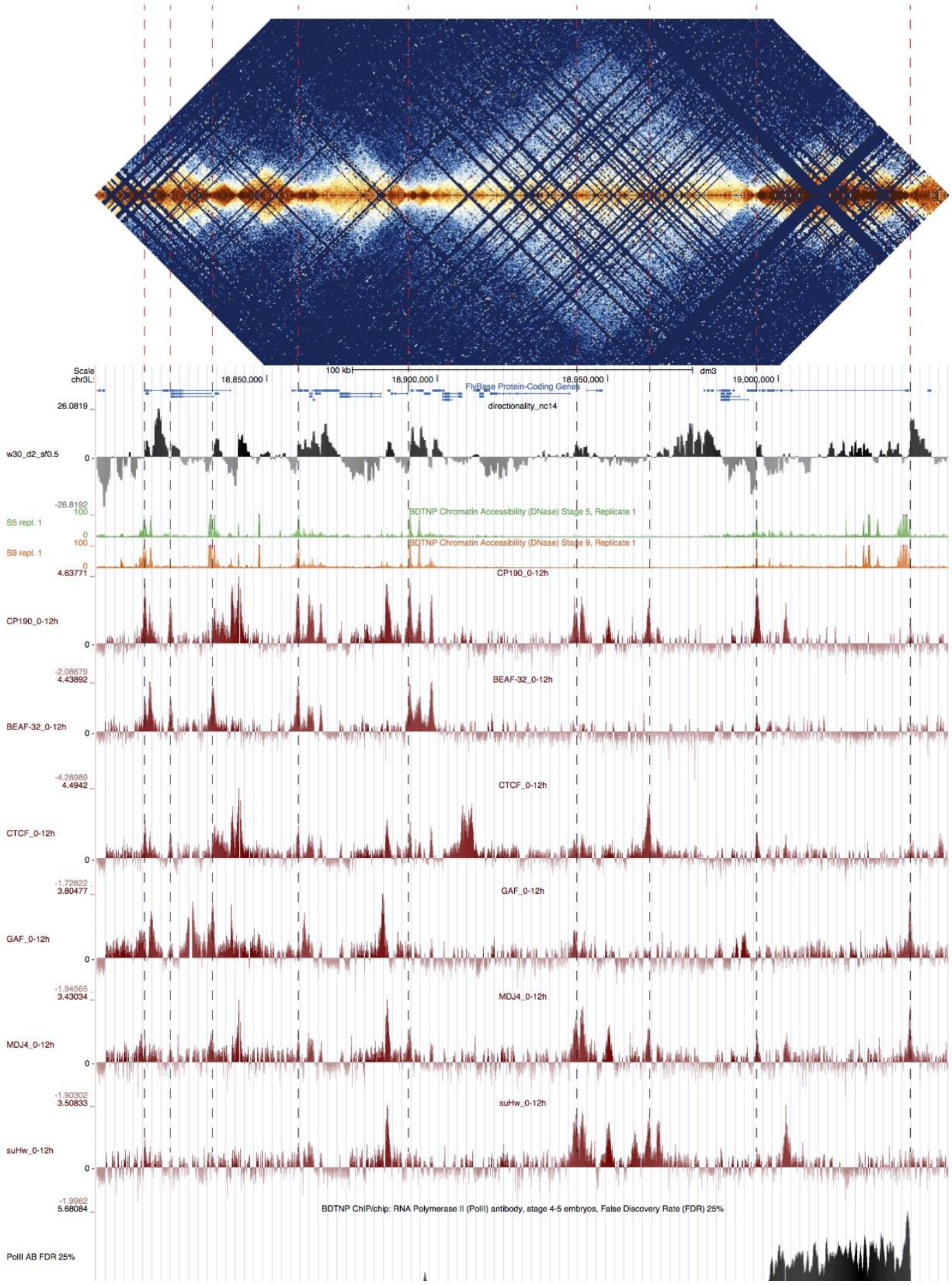
S7: High confidence boundaries, merge of manual and computational curations

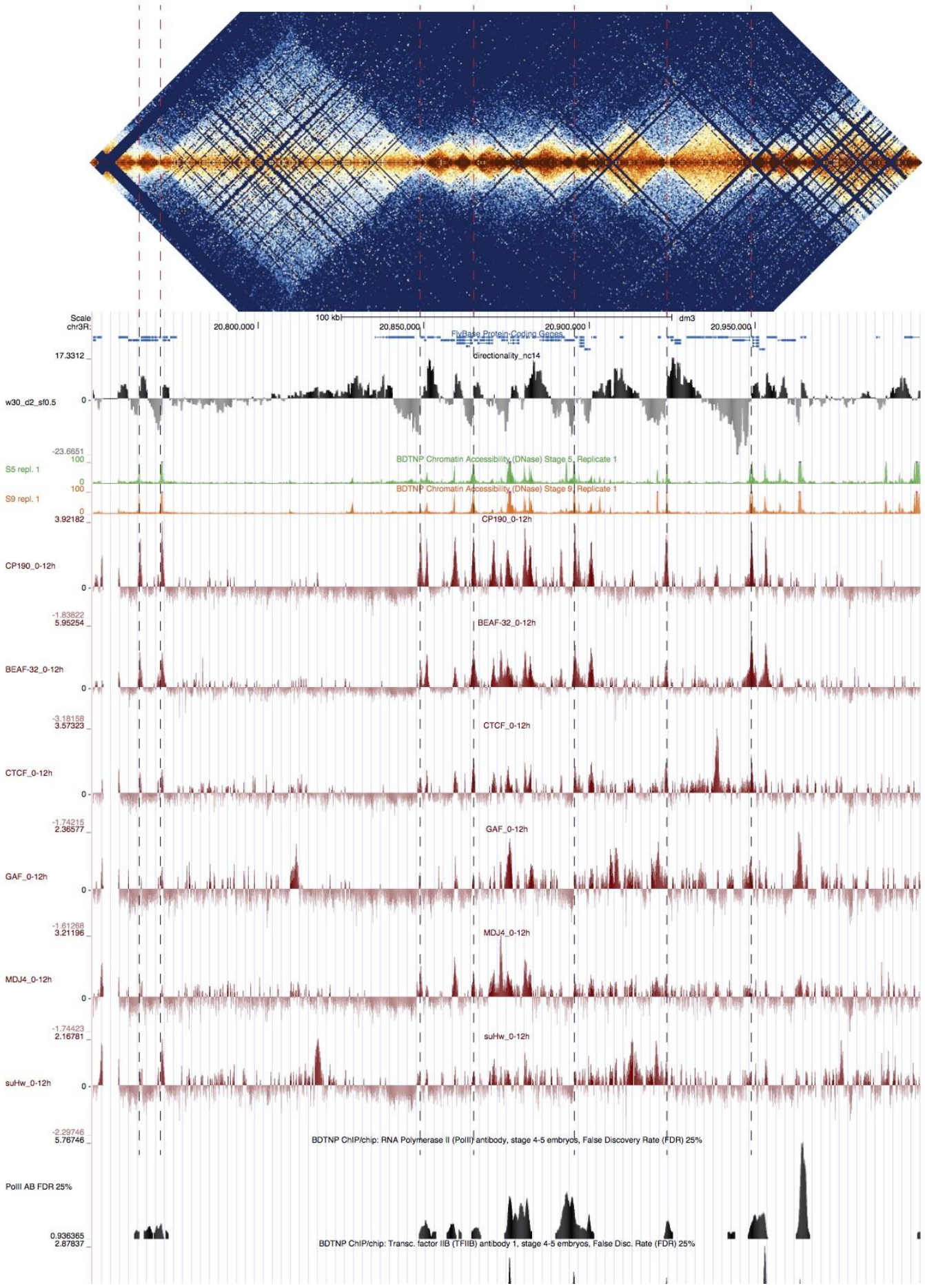
Figure S1



910







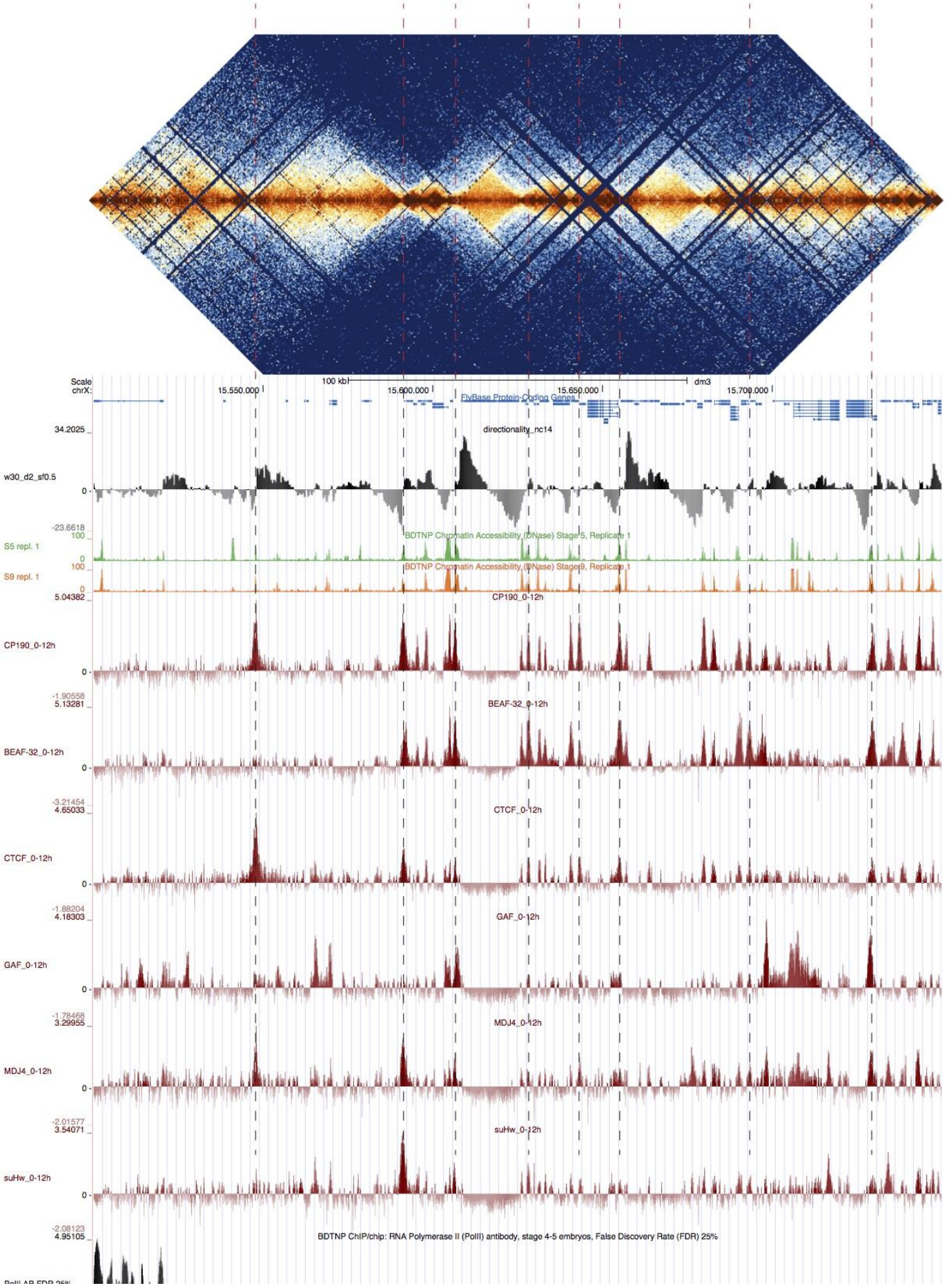


Figure S1: High resolution Hi-C maps of additional example genomic regions from stage 5 *Drosophila melanogaster* embryos. Data were prepared identically to Figure 2, and all data is from the same sources. Regions shown are 2L:10325000-10499999, 2R:19820000-19999999, 3L:18800000-19050000, 3R:20750000-20999999, X:15500000-15749999 (dm3, R5_22).

Figure S2

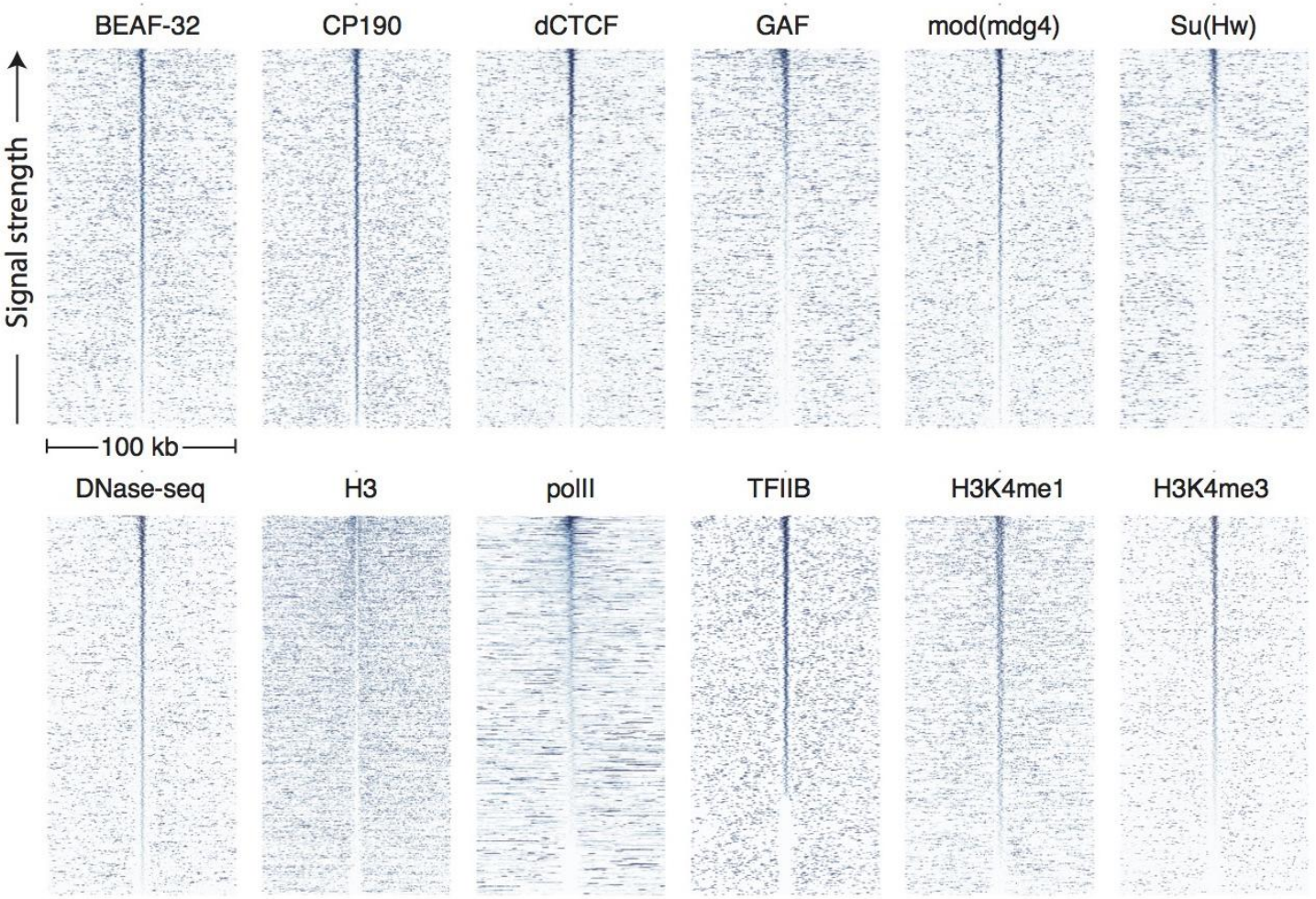


Figure S2: Genomic signals around topological boundaries, self-sorted. Data were prepared and displayed identically to Figure 3 except that each plot is sorted high-to-low by the sum of its own signal in the middle 10 bins (5 kb).

925

Figure S3

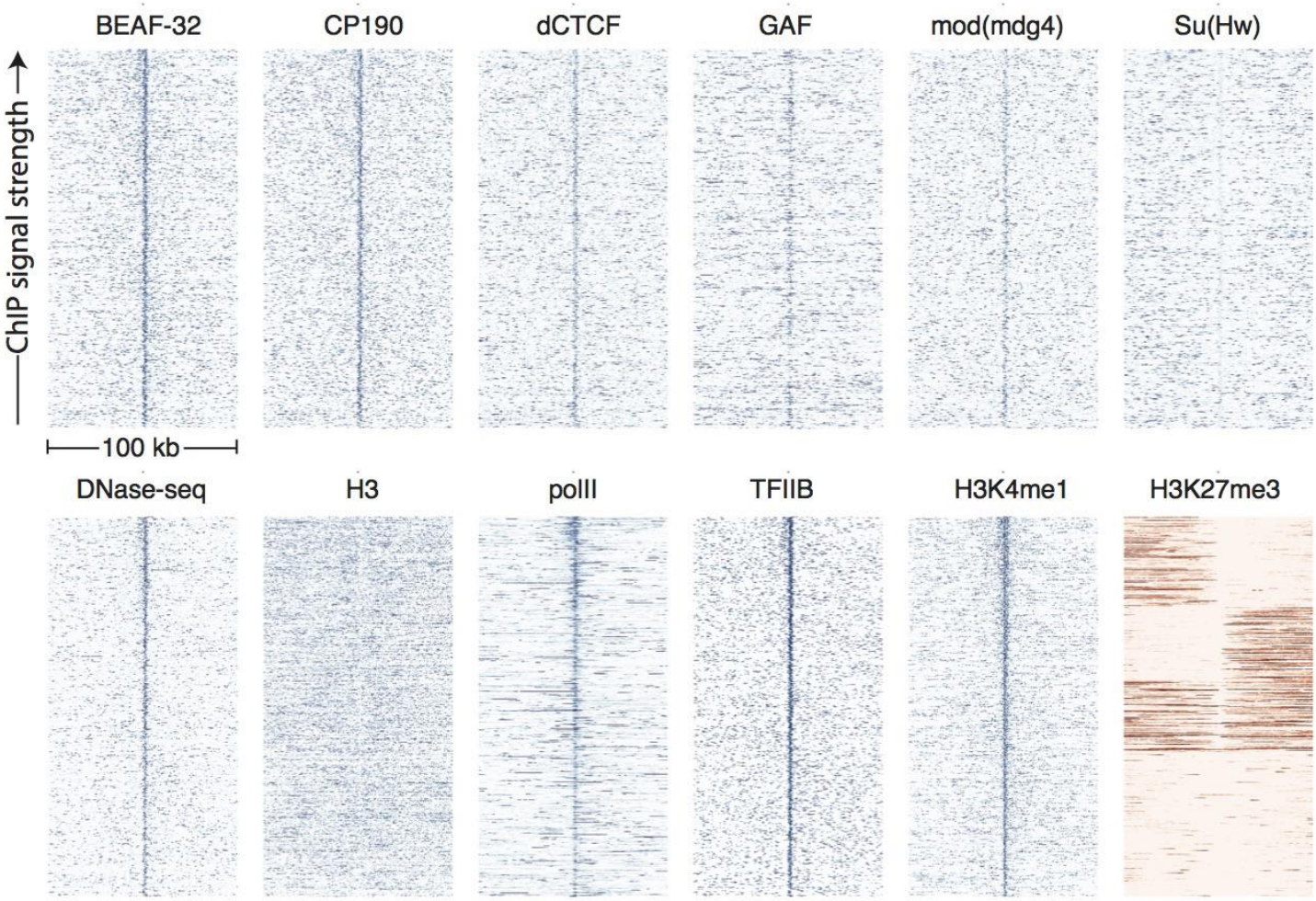


Figure S3: Genomic signals around H3K4me3 peaks. Data were prepared and displayed identically to Figure 3 for the top 937 peaks of H3K4me3 in stage 5 embryos, as identified by (X.-Y. Li et al., 2014). H3K4me3 is a proxy for active promoters.

Figure S4

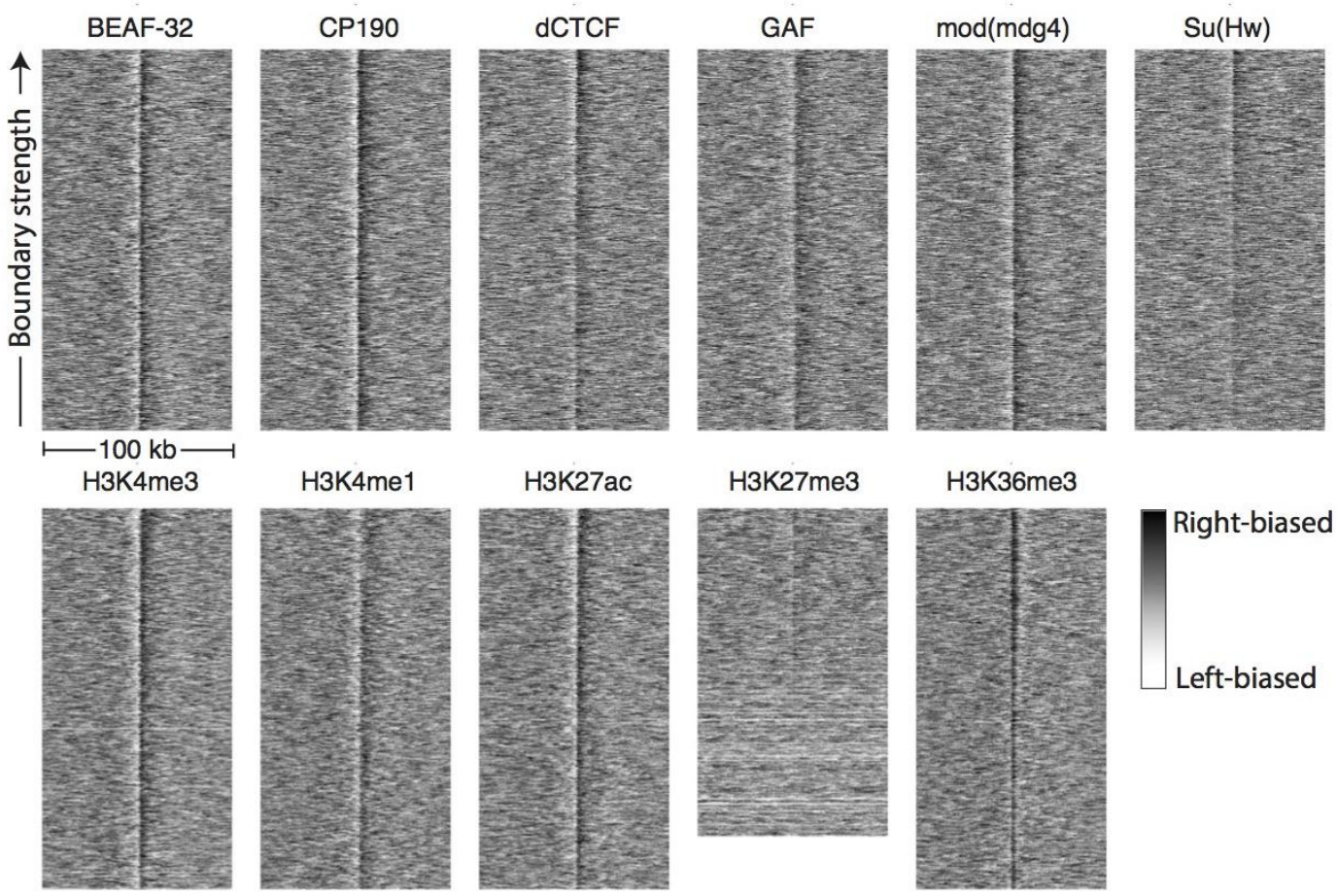
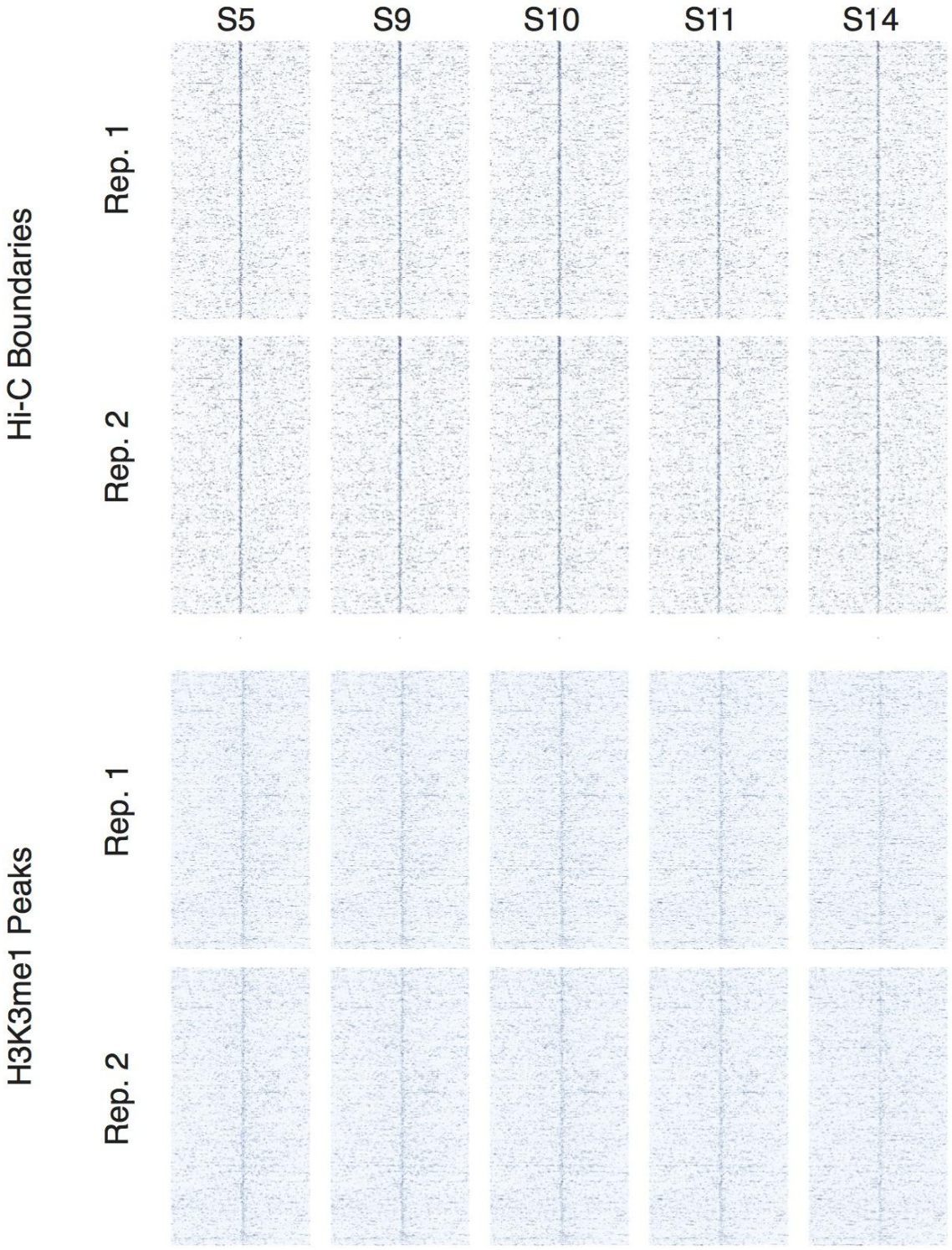


Figure S4: Directionality around peaks of genomic features. Hi-C directionality scores in 500 bp bins from aggregated nc14 data were plotted around the top 937 (where available) peaks of the genomic features shown in Figure 3. The signature of a boundary is a region of left-bias (white) transitioning sharply to a region of right-bias (black).

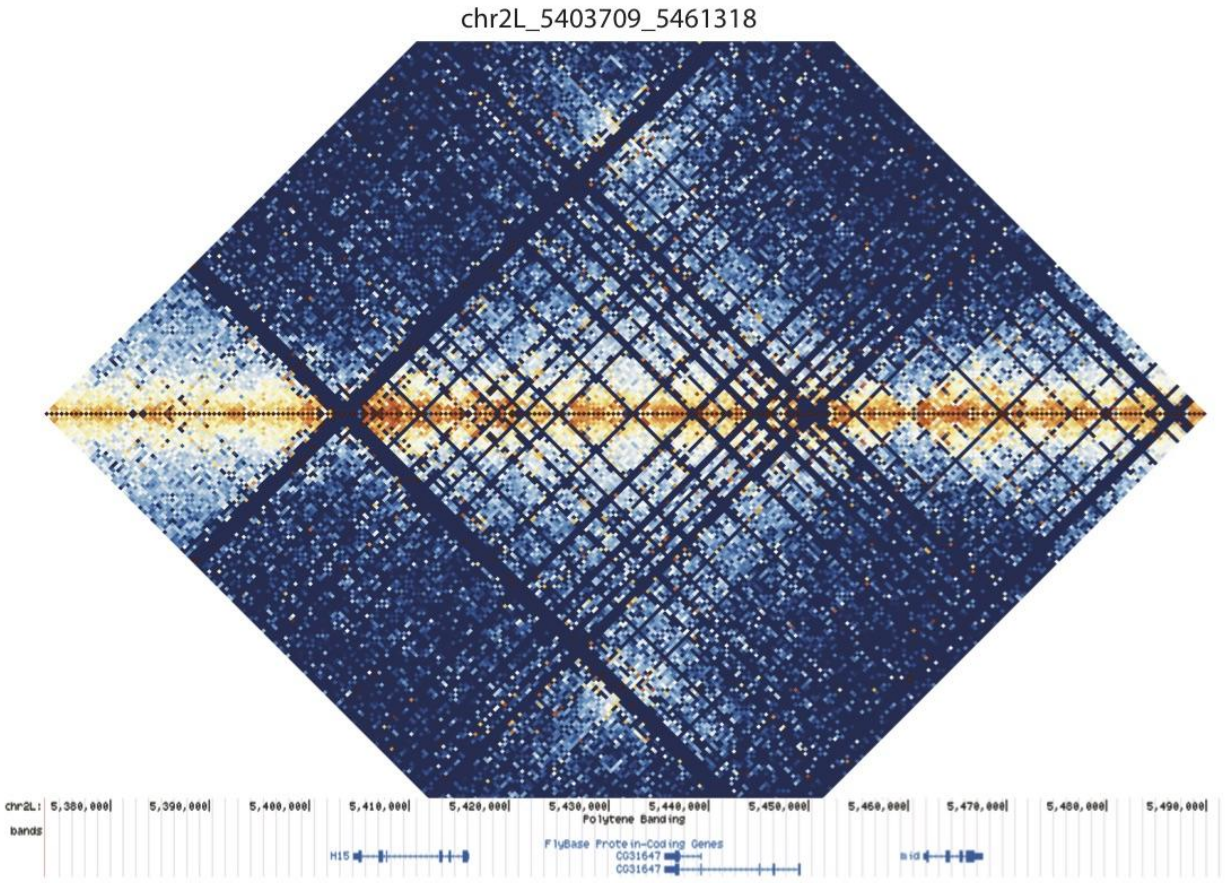
Figure S5

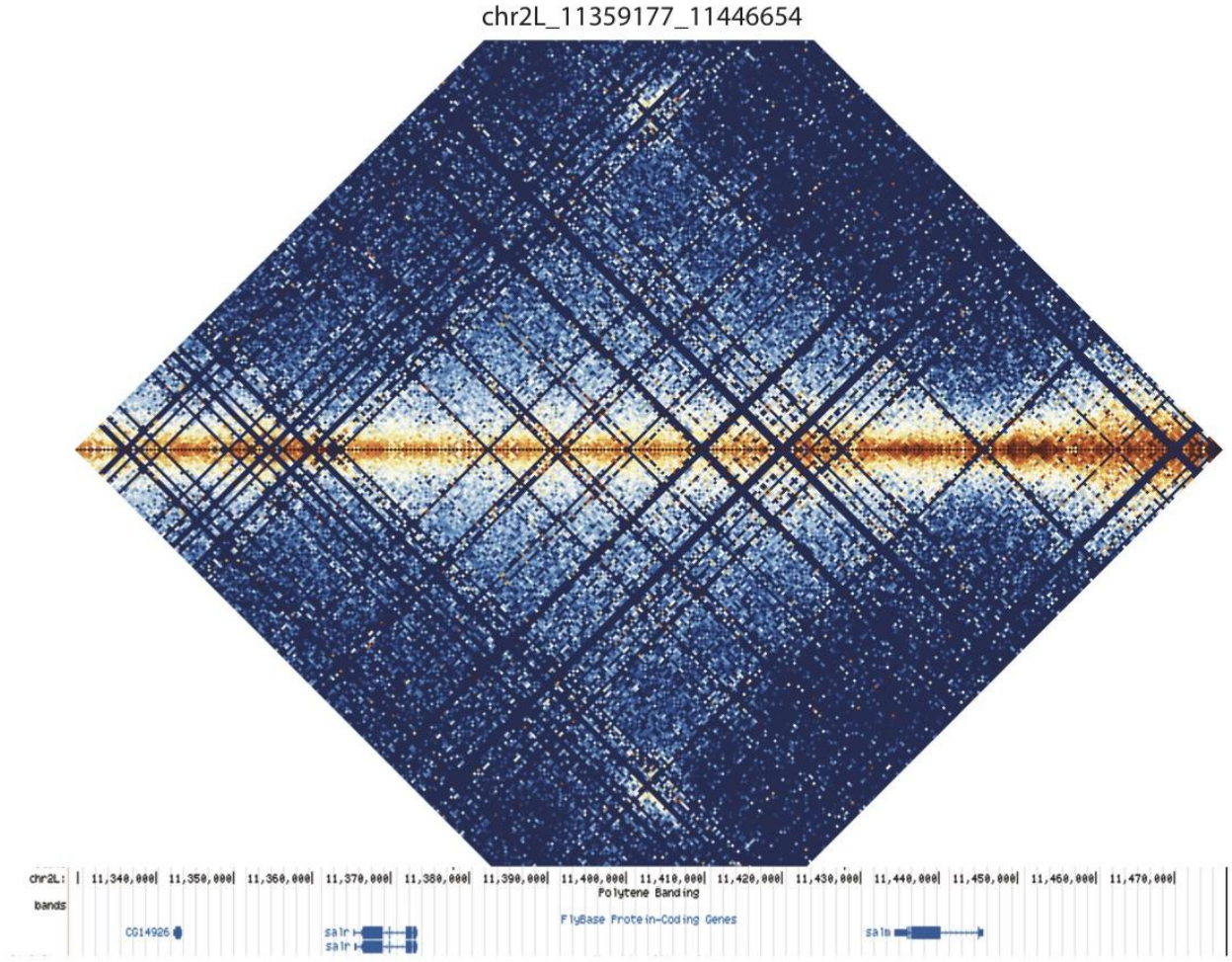


940

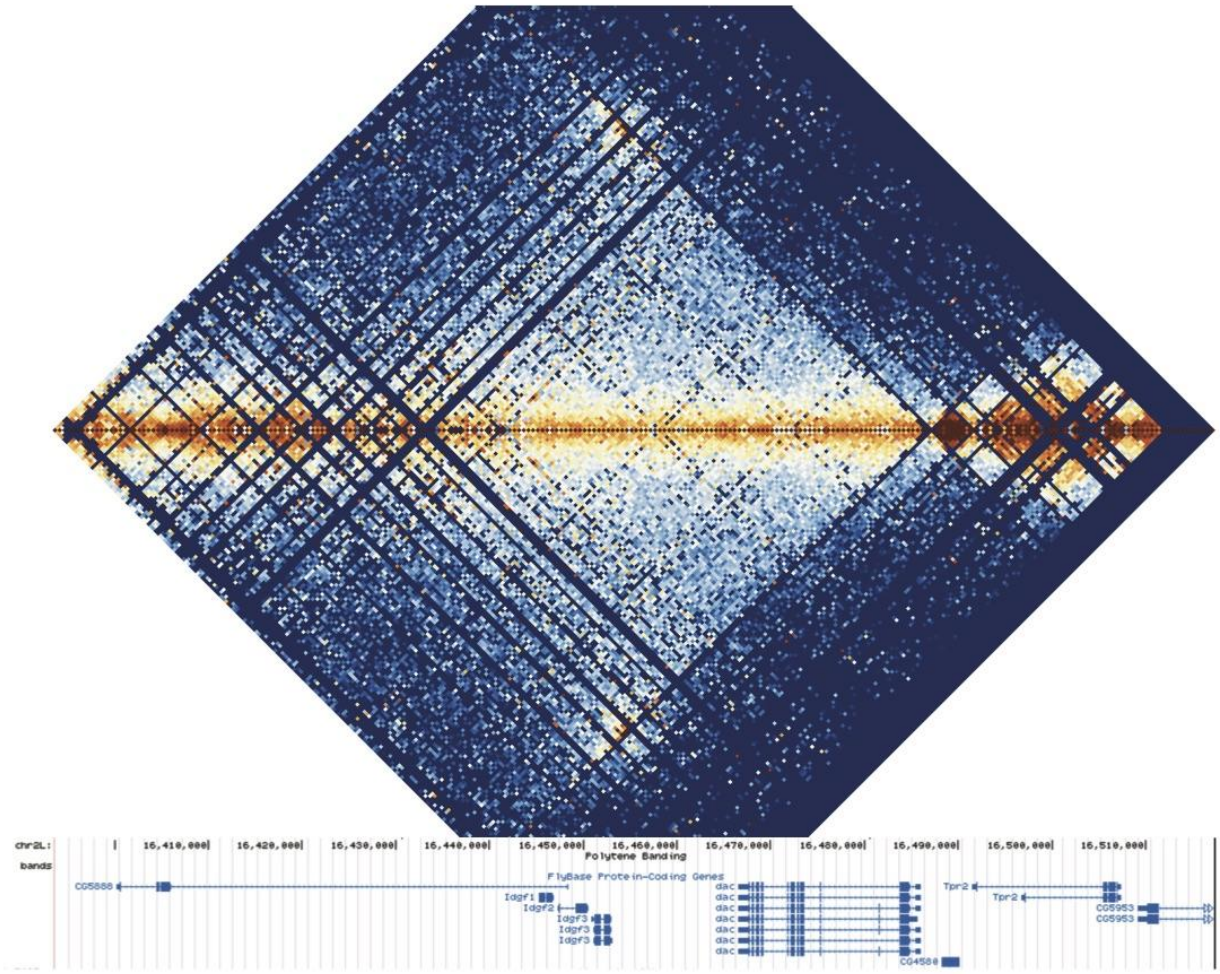
Figure S5: Developmental time series of DNase accessibility at TAD boundaries. DNase accessibility data from was plotted around 937 high-confidence boundaries and top 937 H3K4me1 peaks. Data covers stage 5 (~2 hours) through stage 14 (~10 hours) of embryonic development.

945 Figure S6

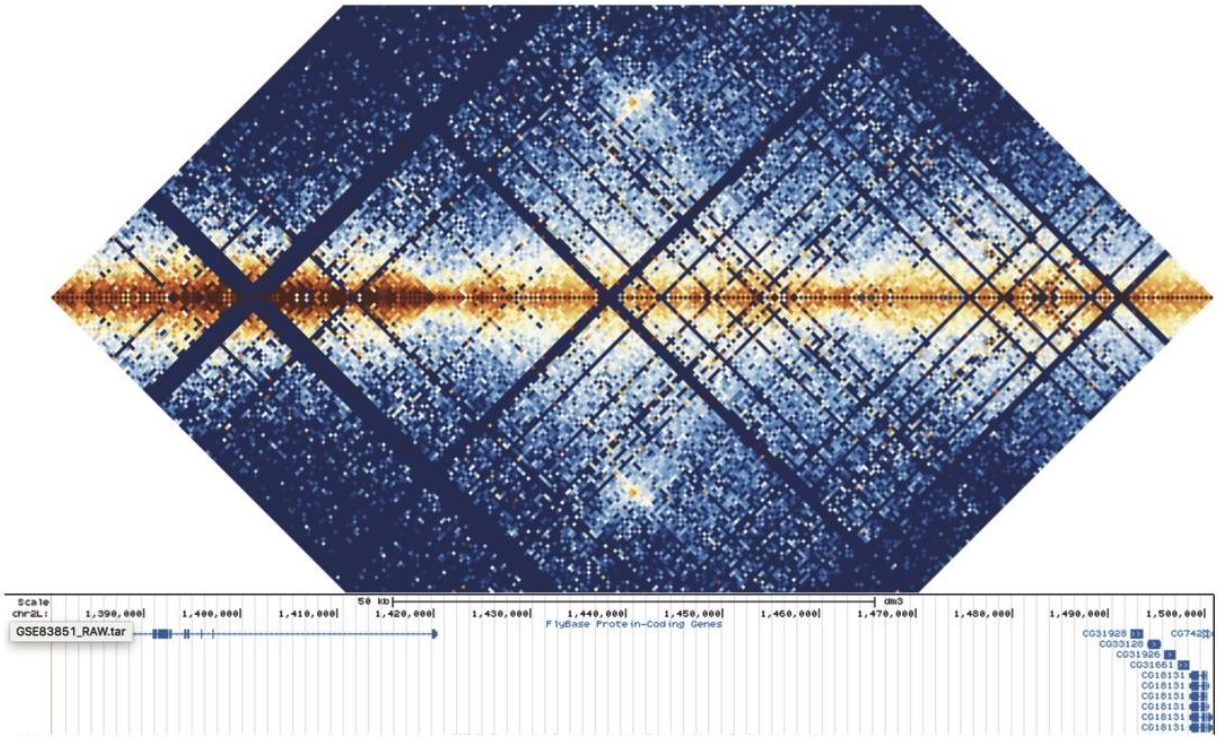




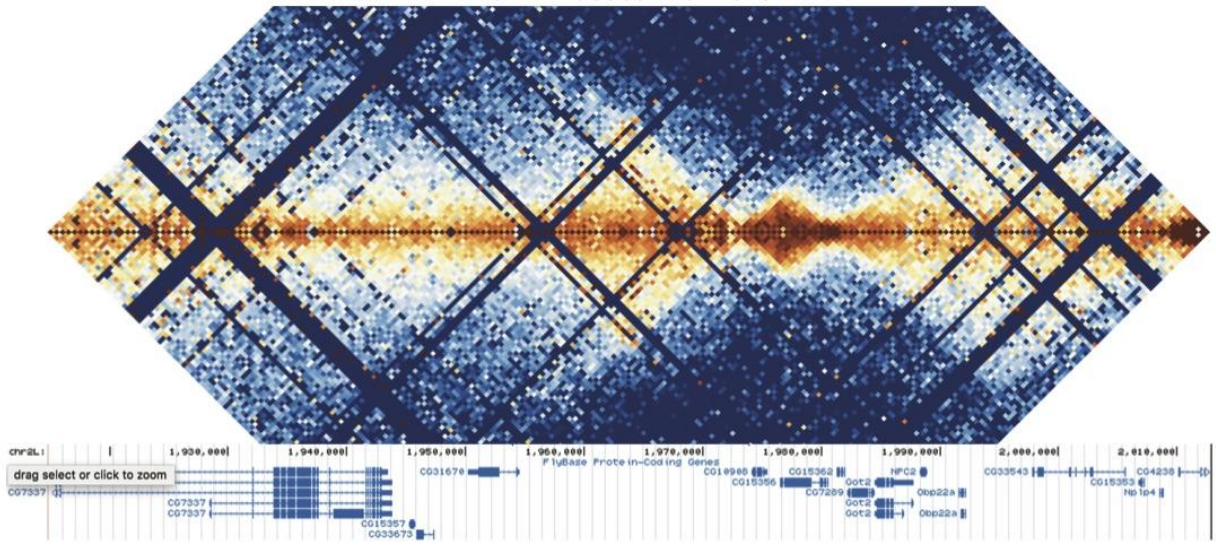
chr2L_16423824_16487341



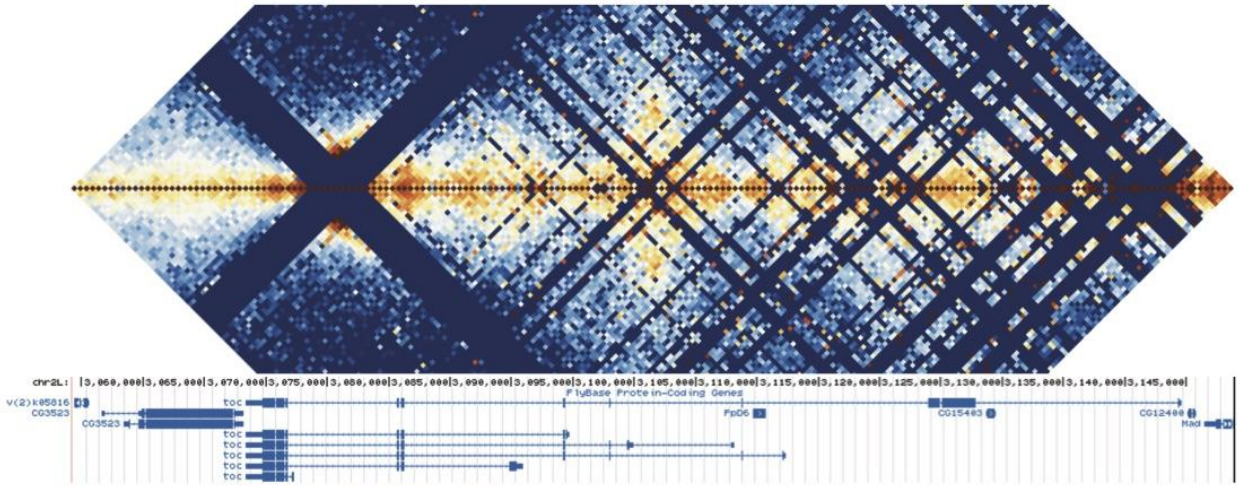
chr2L-1420658-1460903



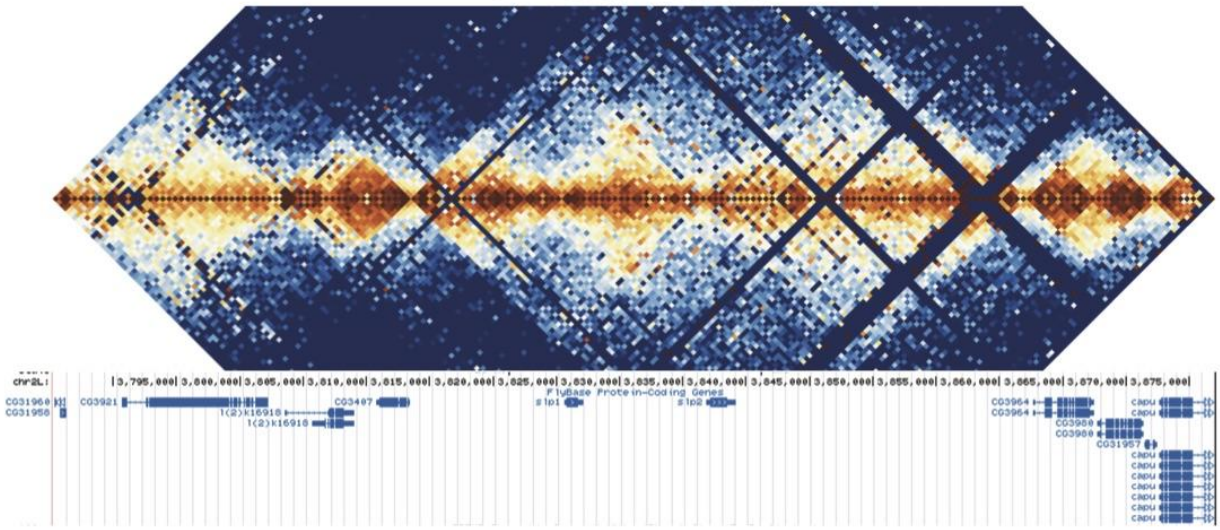
chr2L-1955032-1972848



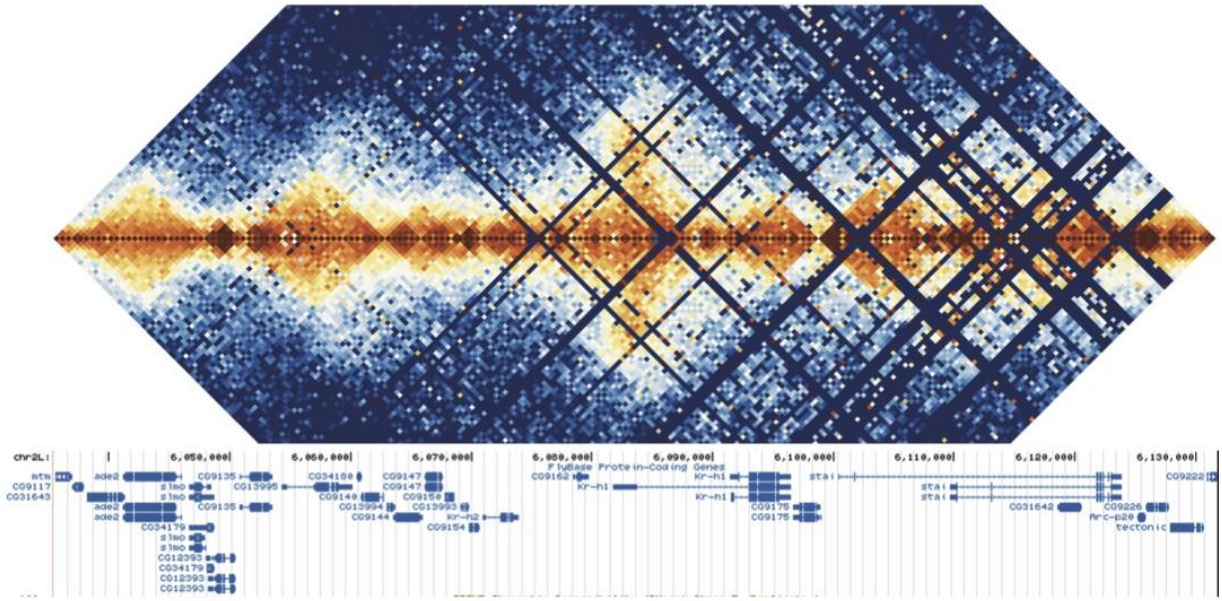
chr2L-3094402-3108884



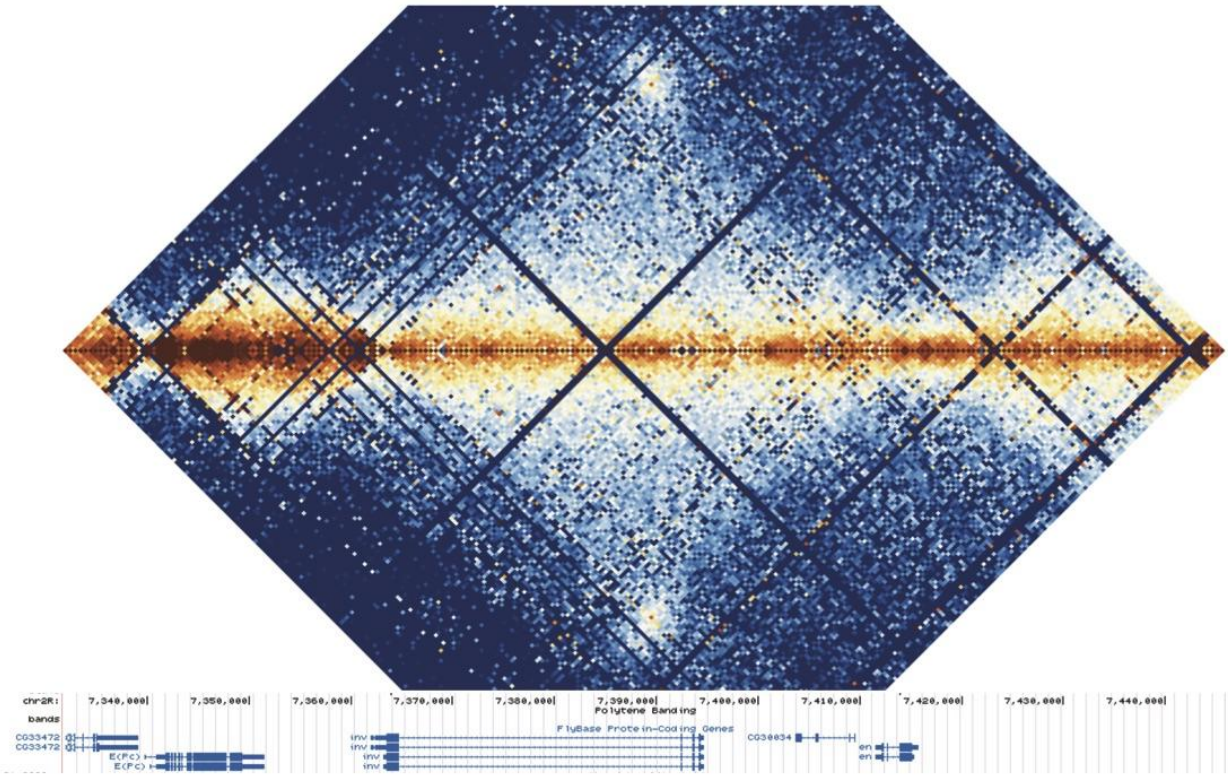
chr2L-3825377-3837035



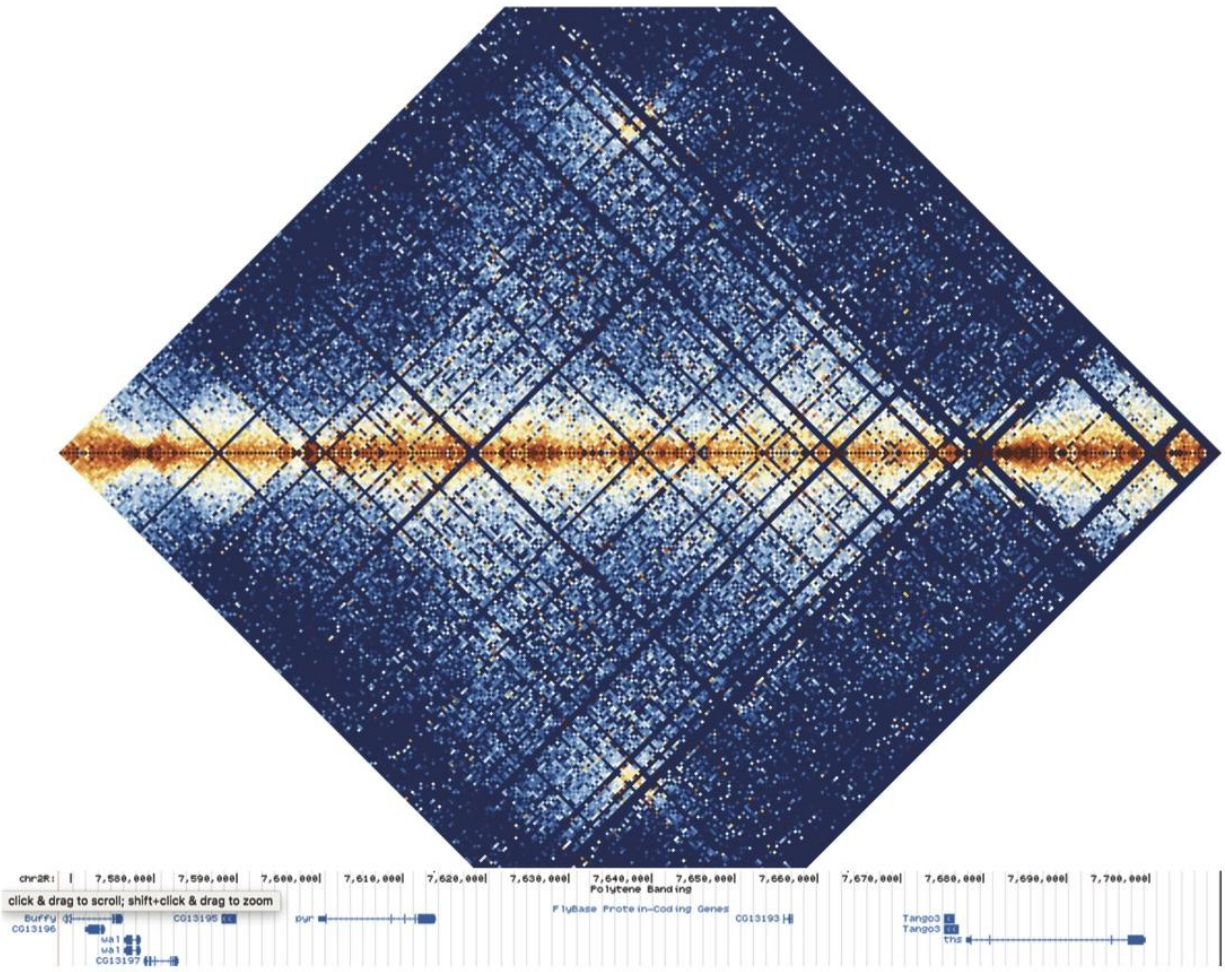
chr2L-6075562-6091803



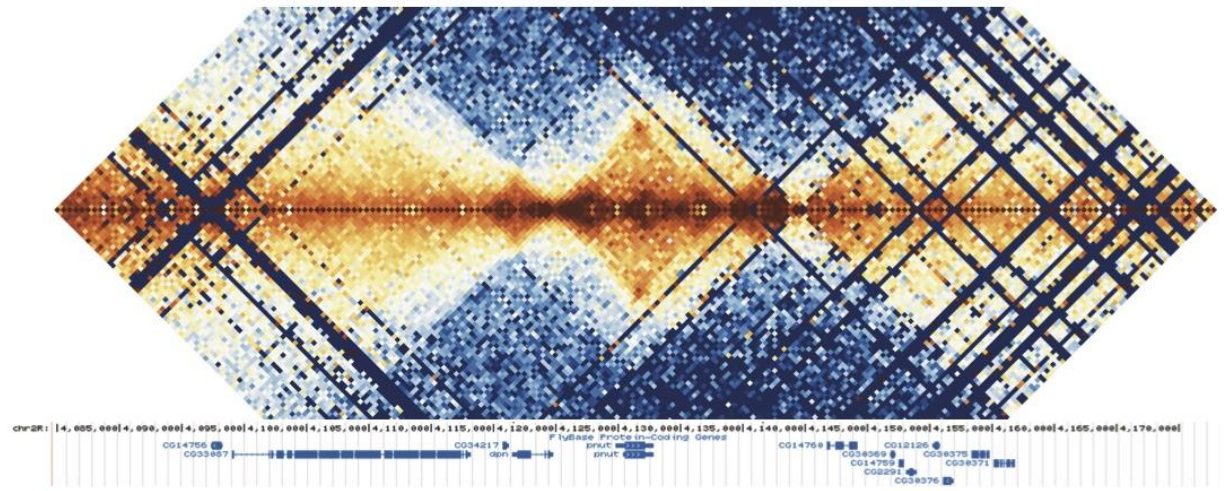
chr2R_7361890_7415913



chr2R_7598779_7678571

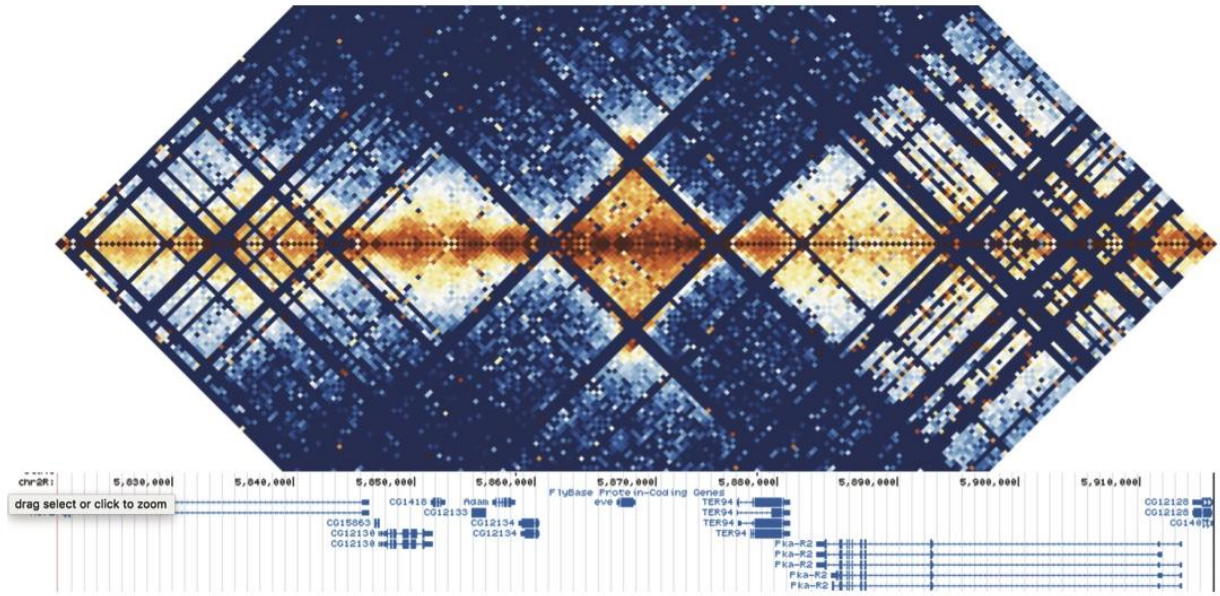


chr2R-4119816-4133364

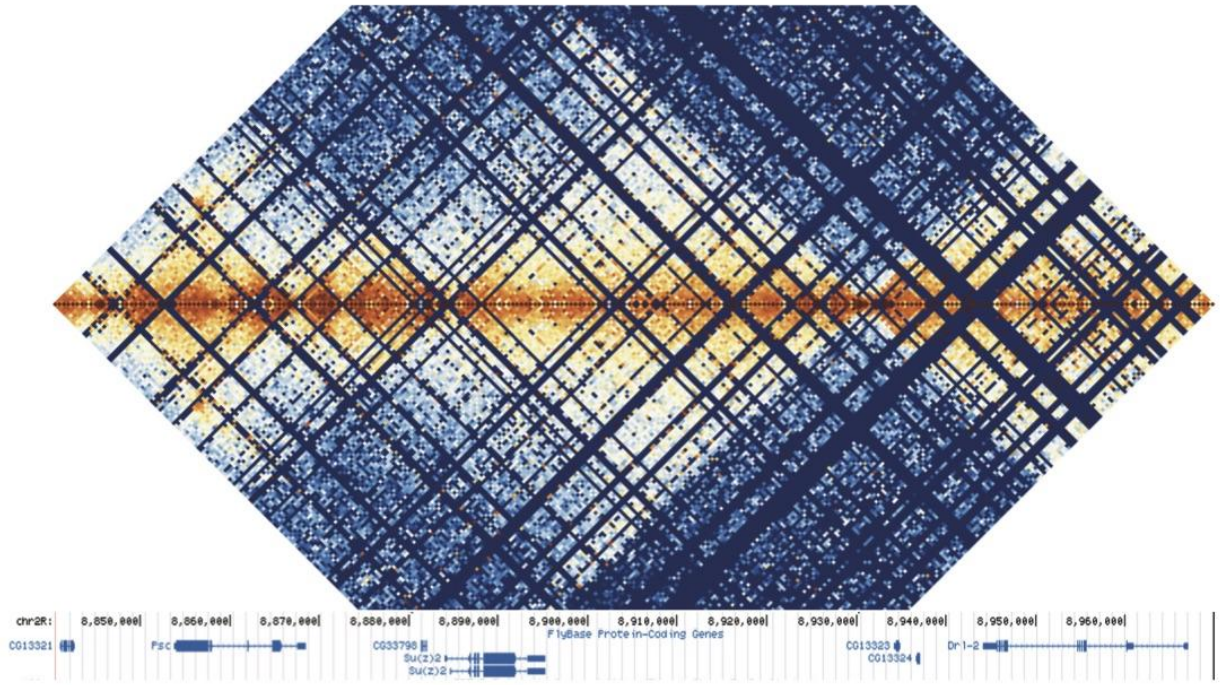


950

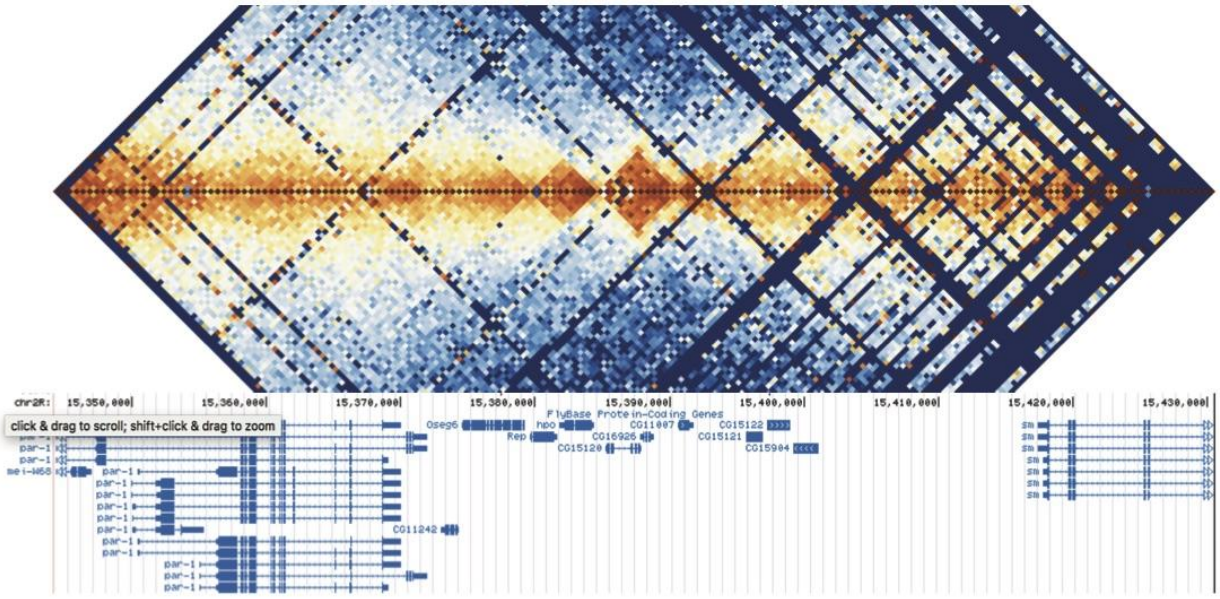
chr2R-5860693-5876110



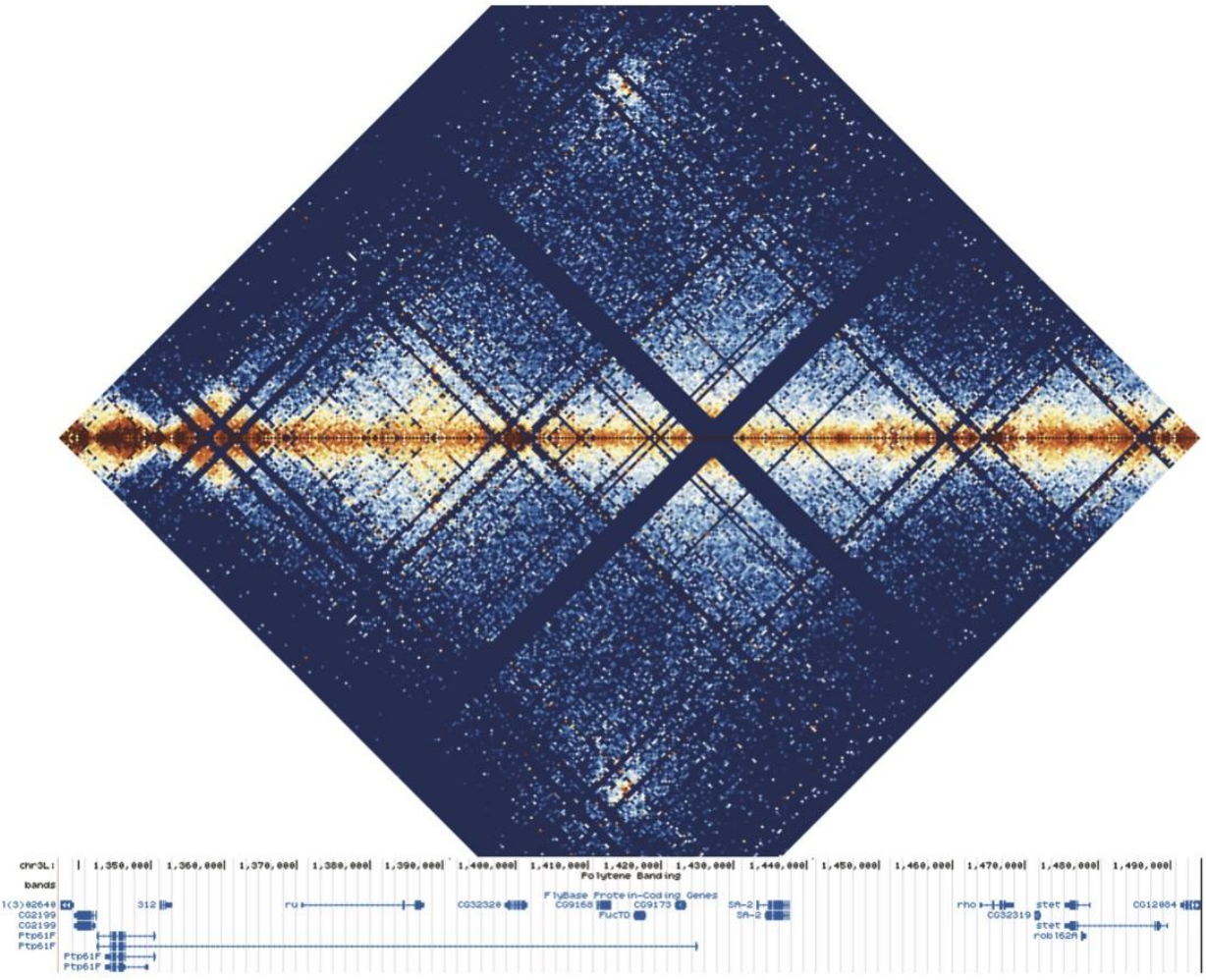
chr2R-8880833-8929825



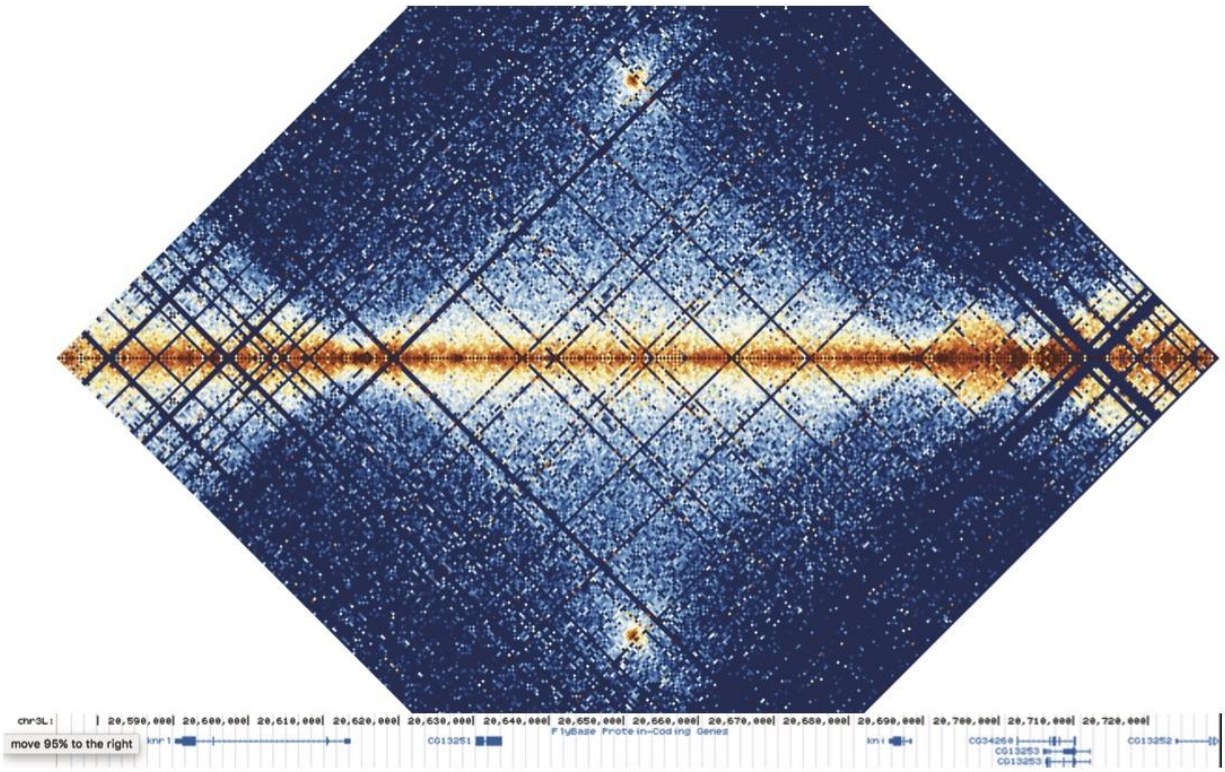
chr2R-15384369-15390479



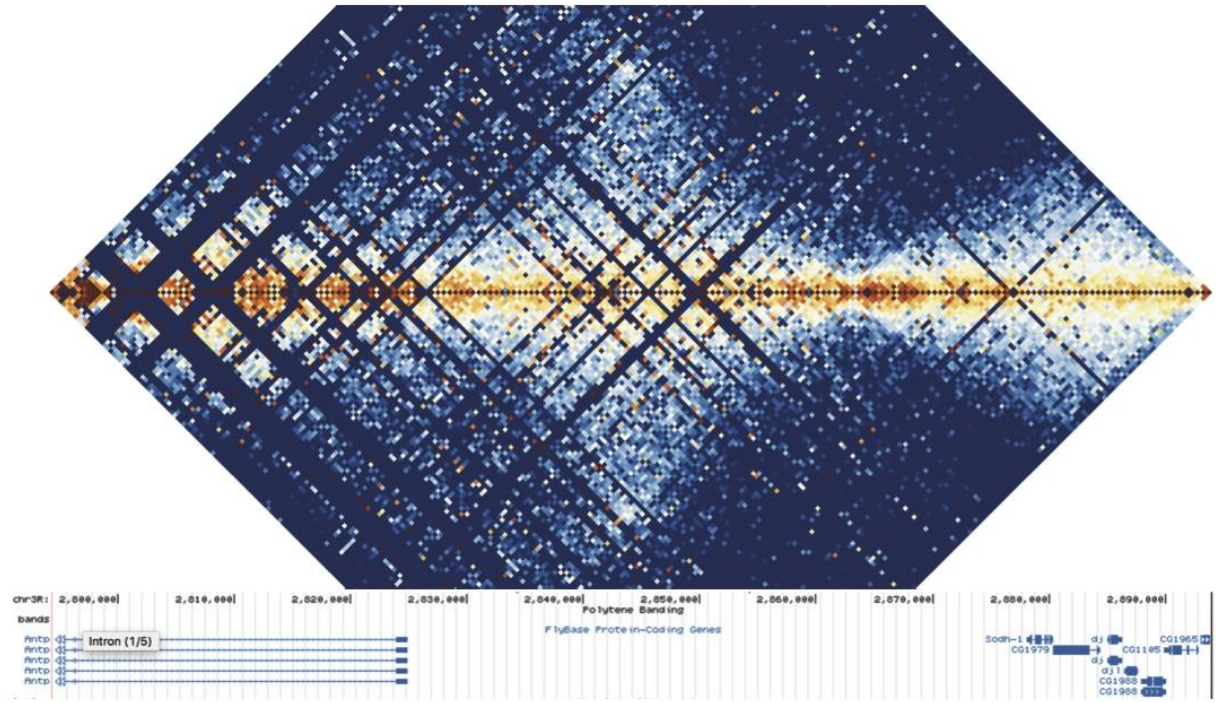
chr3L_1367546_1464138



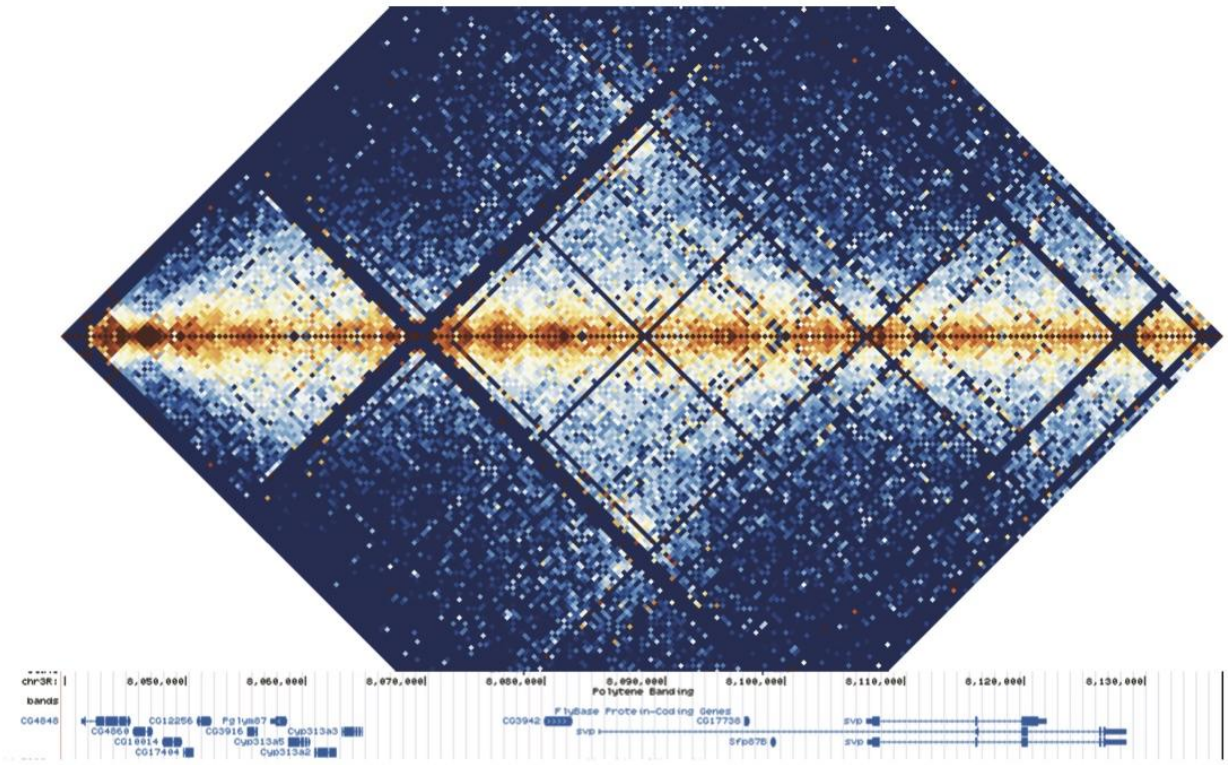
chr3L-20614828-20689547



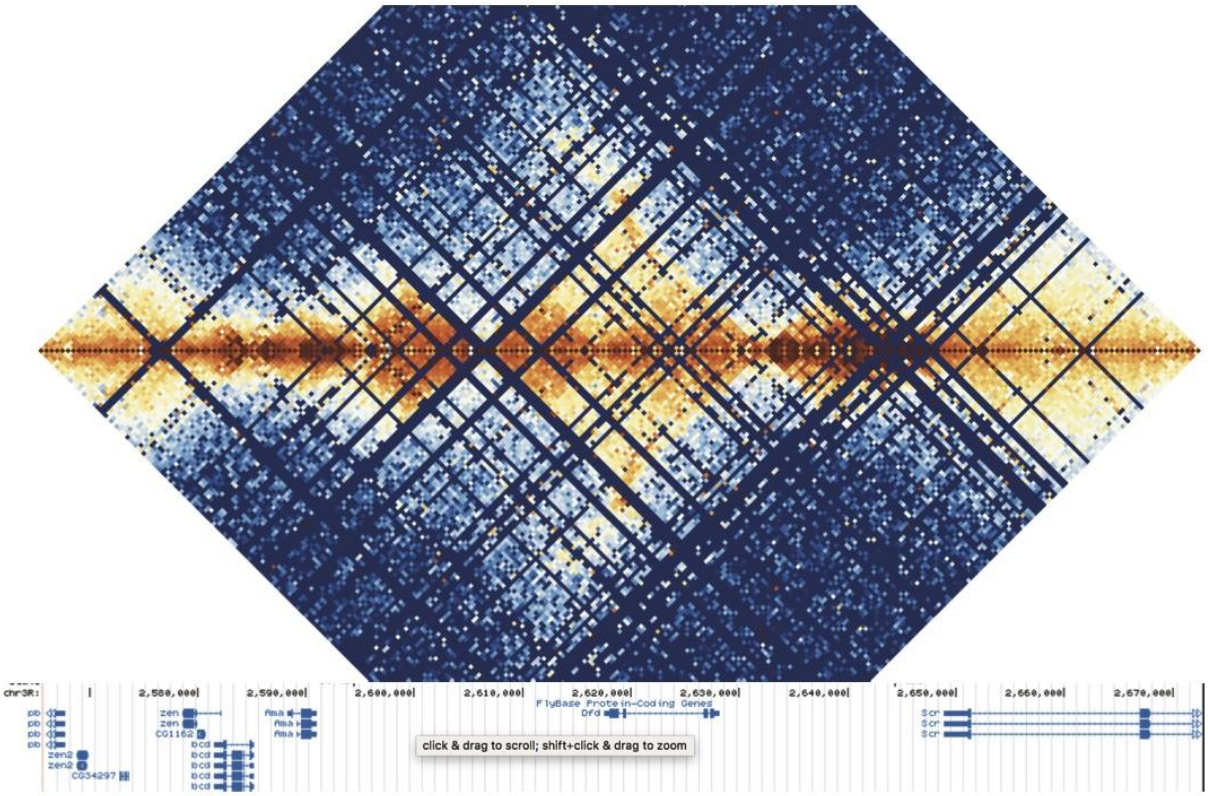
chr3R_2824593_2863924



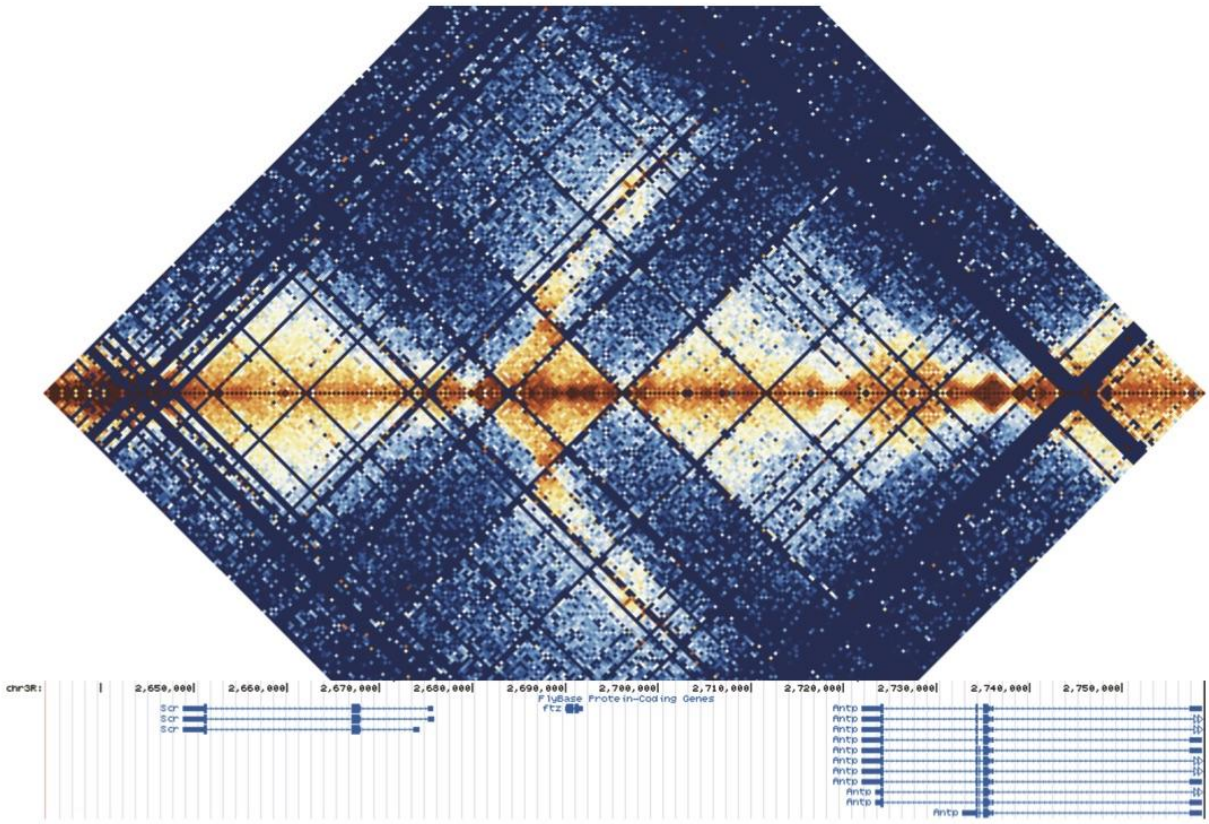
chr3R_8069847_8106465



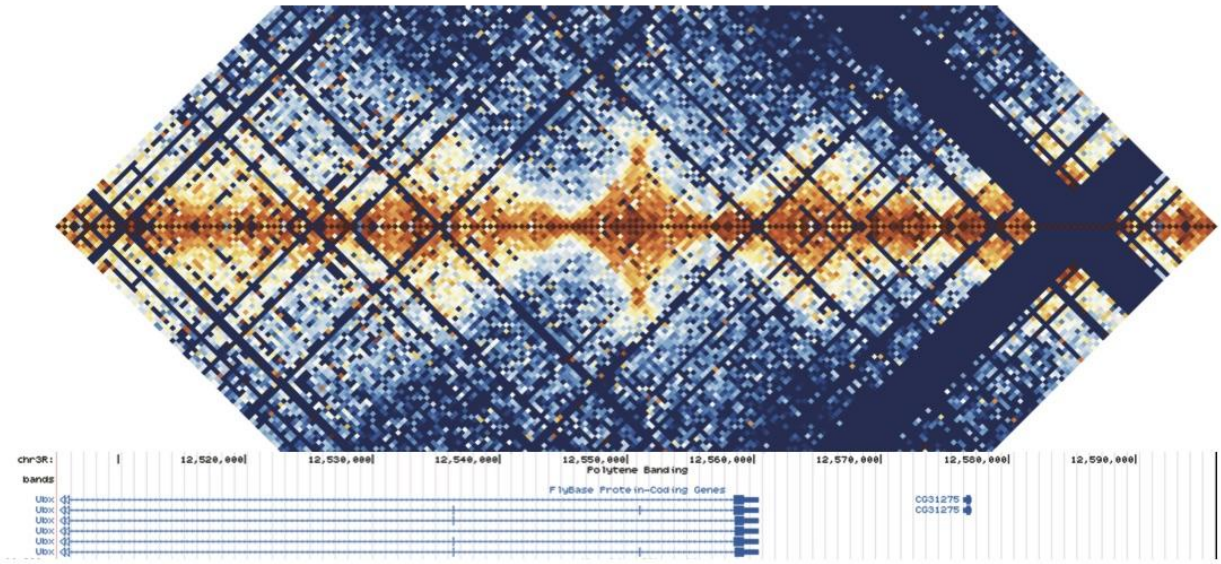
chr3R-2605932-2632790



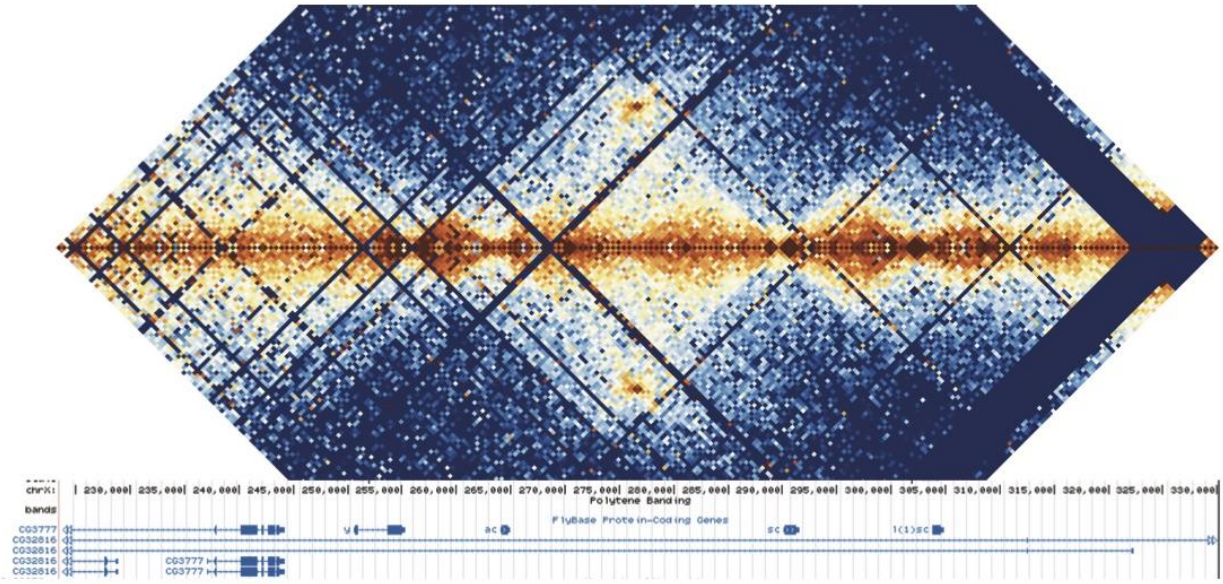
chr3R-2674246-2718838



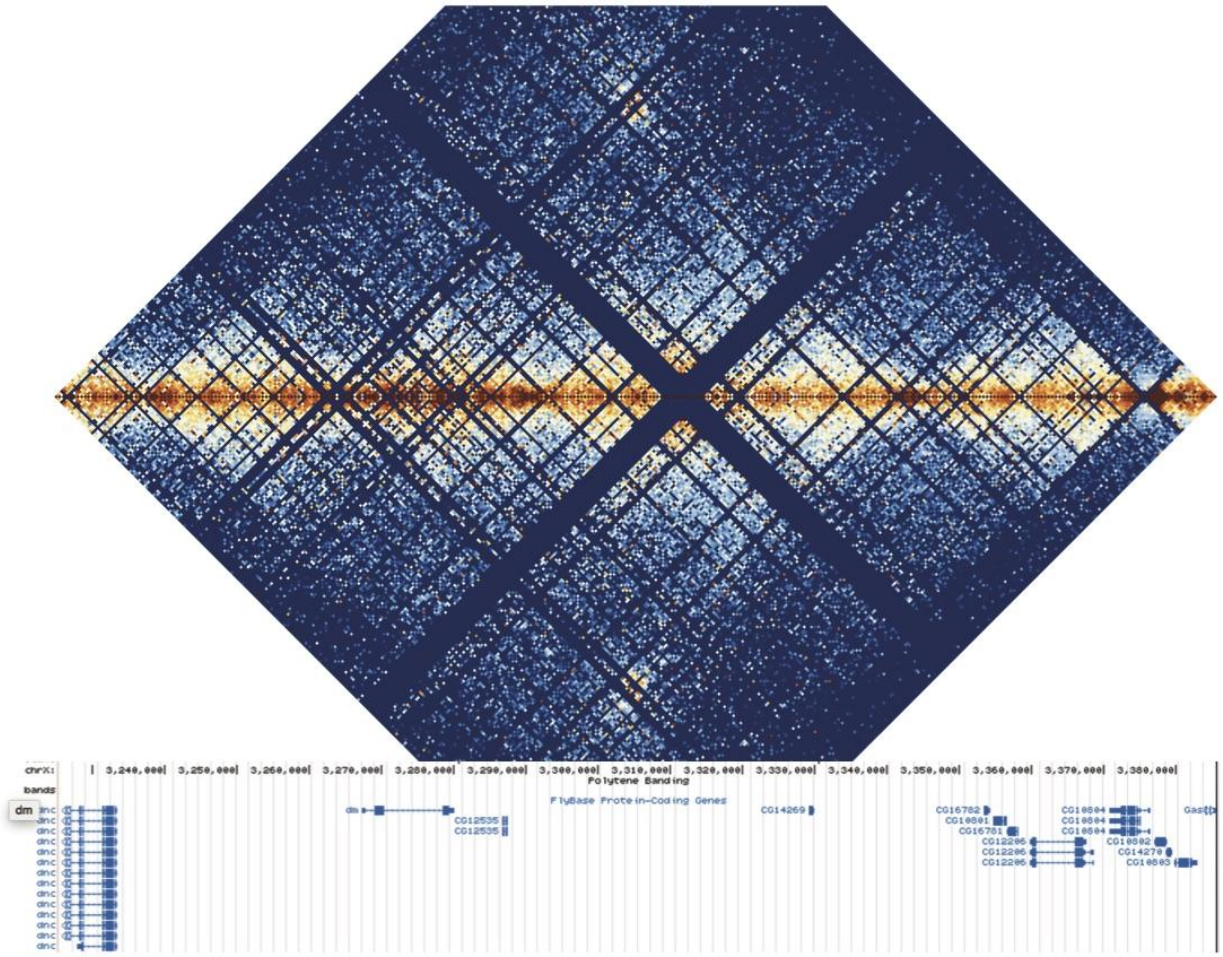
chr3R-12545340-12556303

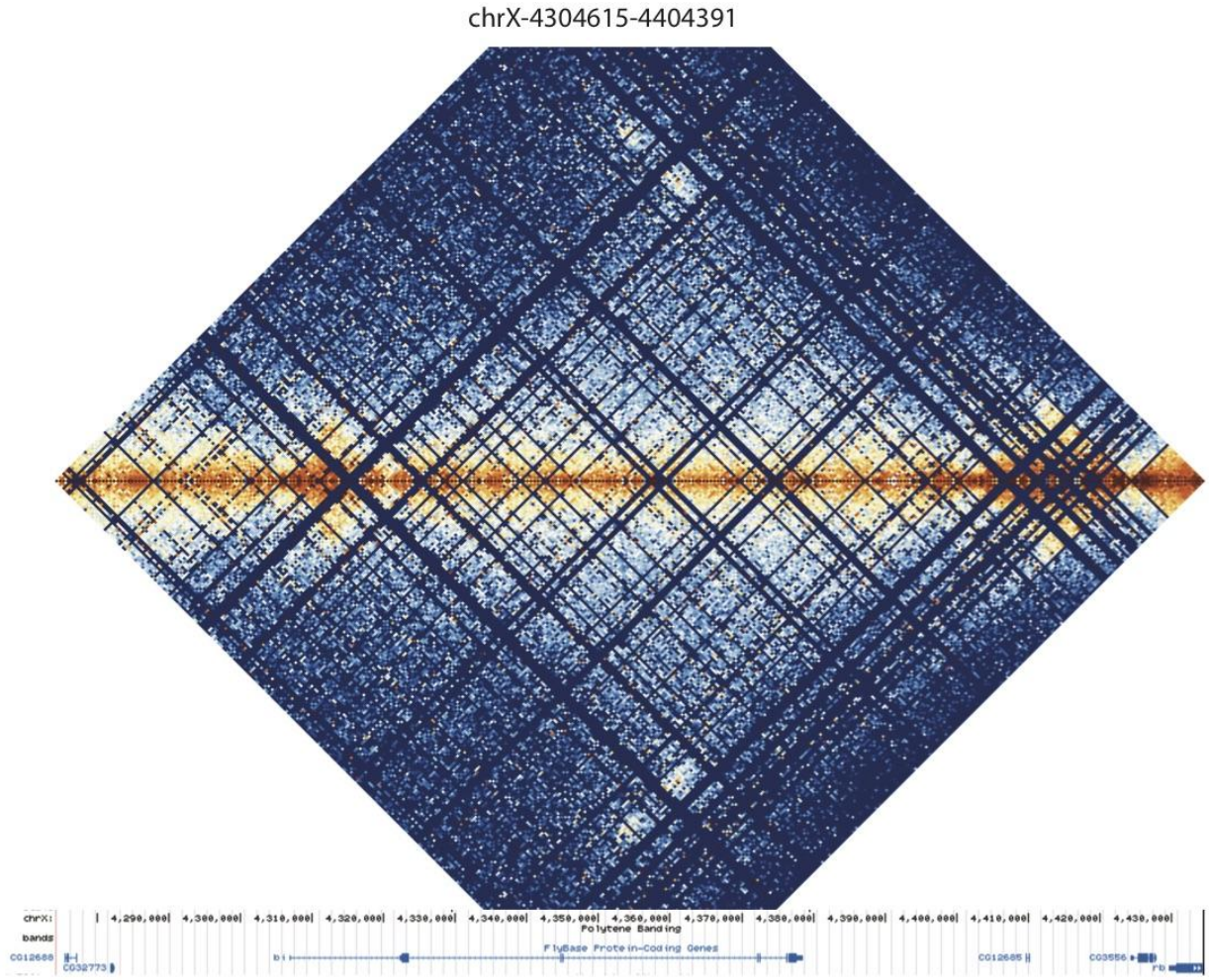


chrX-263672-290115

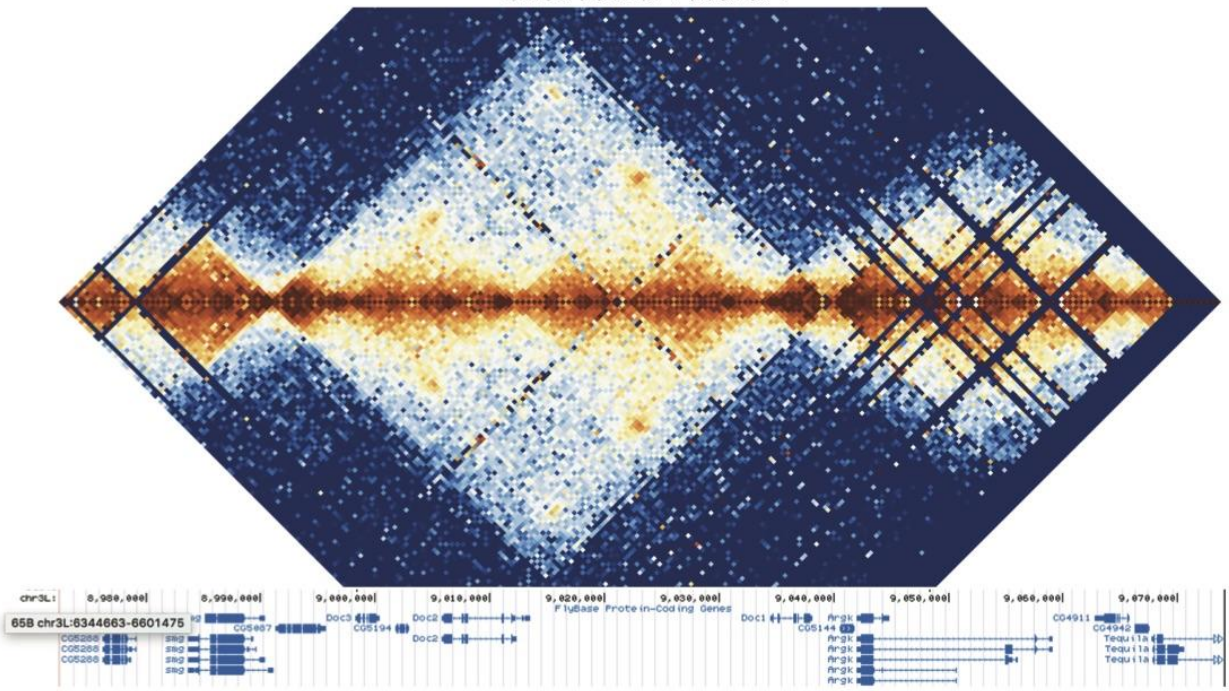


chrX-3265514-3345524

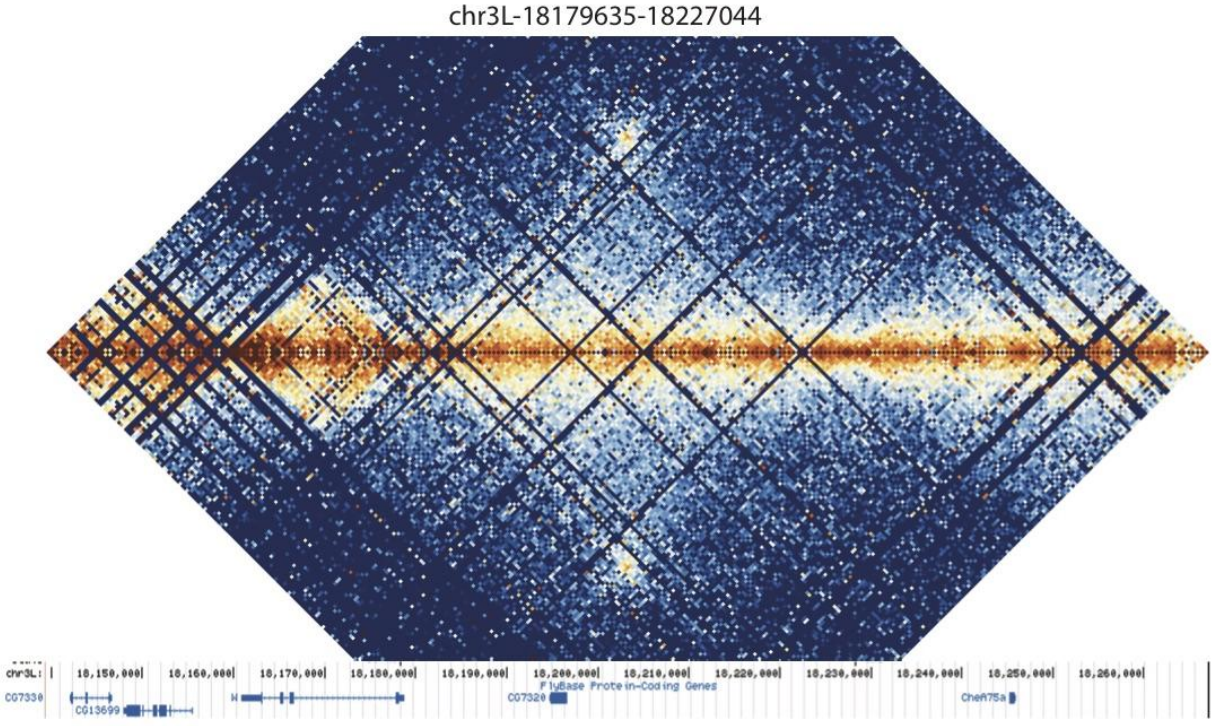


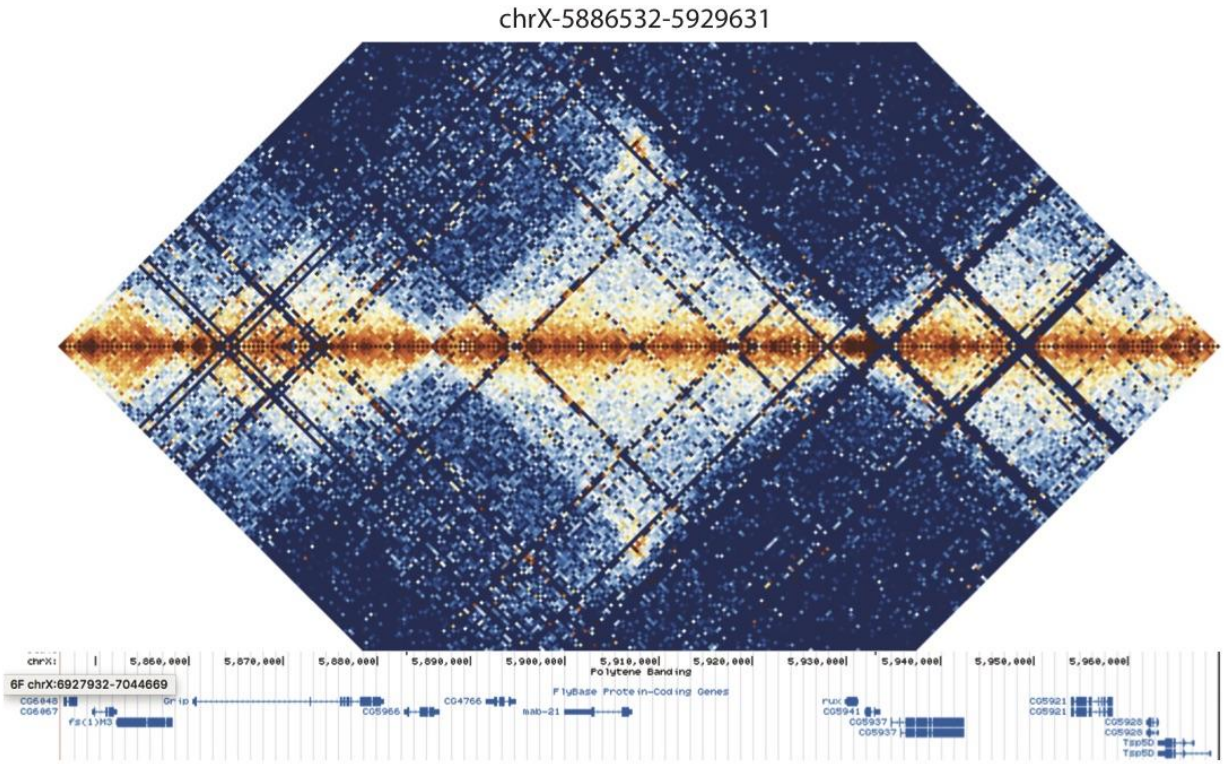


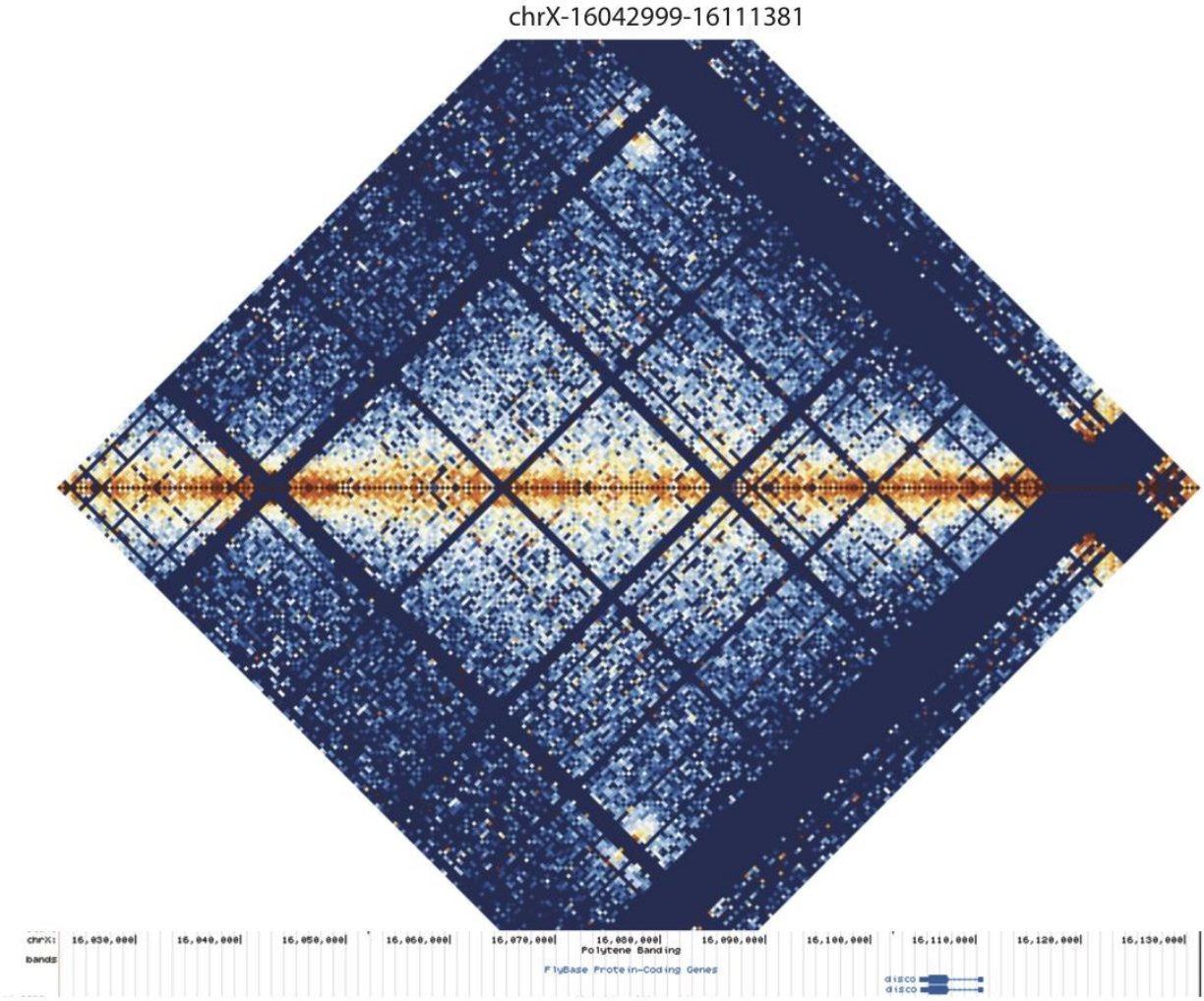
chr3L-9012641-9034071

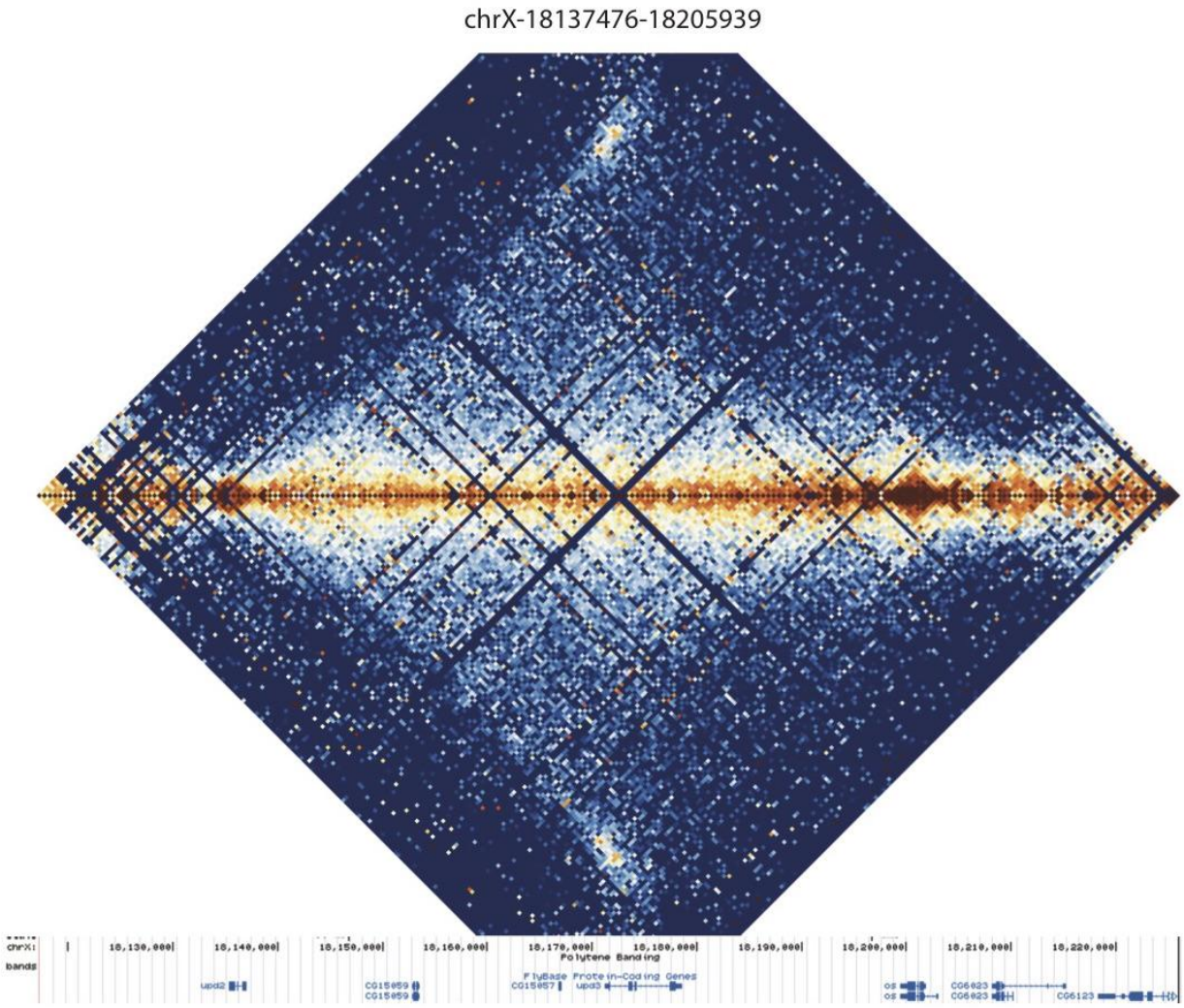


chr3L-18179635-18227044









960

Figure S6: Distal chromatin contacts in stage 5 embryos. Hi-C maps prepared as in Figure 7 for manually-curated distal interactions. Data shown at 500 bp resolution, all nc14 data aggregated. Gene tracks are from the UCSC genome browser dm3.

965

Figure S7

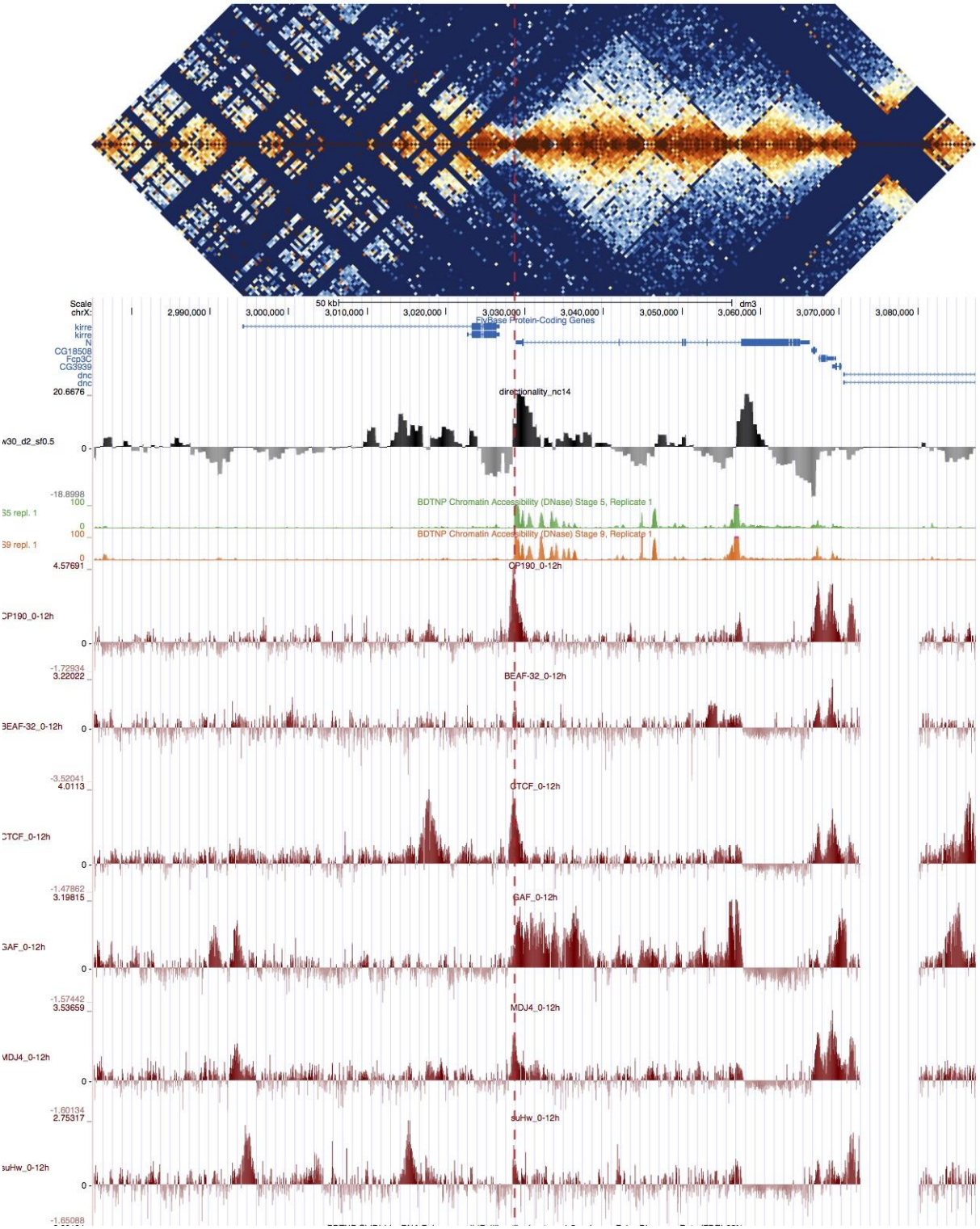


Figure S7: TAD structure corresponds to mapped polytene structure at the *Notch* locus. Hi-C map of the *Notch* locus, prepared as in Figure 4.

970

Figure S8

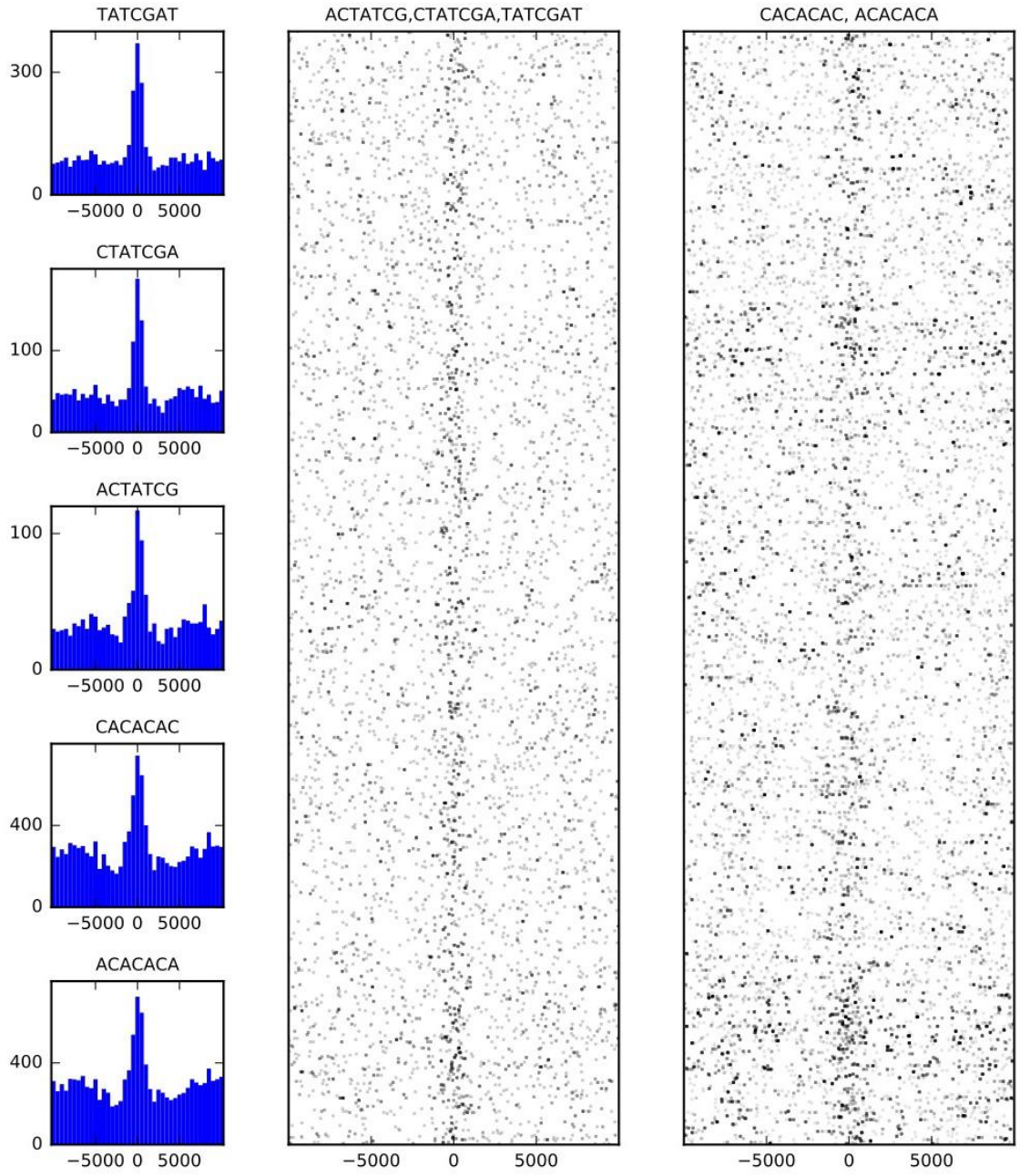


Figure S8: Sequence features of TAD boundary elements. (A) Histograms showing the frequency of enriched 7-mers in 5 kb windows around 937 high-confidence TAD boundaries. (**B**) Scatter plots of occurrences of words matching known BEAF-32 binding motifs (left) and CACA motif (right) in 10 kb windows around high-confidence TAD boundaries. Points are plotted with low opacity, such that darker points correspond to positions where multiple words occurs close together in sequence.