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

High-throughput assays of three-dimensional interactions of chromosomes have shed considerable light on the structure of animal chromatin. Despite this progress, the precise physical nature of observed structures and the forces that govern their establishment remain poorly understood. Here we present high resolution Hi-C data from early *Drosophila* embryos. We demonstrate that boundaries between topological domains of various sizes map to DNA elements that resemble classical insulator elements: short genomic regions sensitive to DNase digestion that are strongly bound by known insulator proteins and are frequently located between divergent promoters. Further, we show a striking correspondence between these elements and the locations of mapped polytene interband regions. We believe it is likely this relationship between insulators, topological boundaries, and polytene interbands extends across the genome, and we therefore propose a model in which decompaction of boundary-insulator-interband 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 and Bauer 1933; King and 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 (Benyajati and Worcel 1976; Laird and Chooi 1976; Laird et al. 1976; McKnight and Miller 1976, 1977; Vlassova et al. 1985; Belyaeva and Zhimulev 1994).

Detailed examination of polytene chromosomes in *Drosophila melanogaster* revealed a stereotyped organization, with compacted, DNA-rich "bands" alternating with extended, DNA-poor "interband" regions (Bridges 1934; Rabinowitz 1941; Lefevre 1976; Benyajati and Worcel 1976; Laird and Chooi 1976), 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 (Rebecca Kellum and Schedl 1991; Holdridge and Dorsett 1991; Geyer and Corces 1992; R. Kellum and Schedl 1992). Subsequent work showed that these elements could also block the spread of silenced chromatin states (Roseman, Pirrotta, and Geyer 1993; Sigrist and Pirrotta 1997; Mallin et al. 1998; Recillas-Targa et al. 2002; Kahn et al. 2006) 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 (Lindsley and Grell 1968; Lewis 1981; Parkhurst and Corces 1985, 1986; Spana, Harrison, and Corces 1988; Parkhurst et al. 1988; Zhao, Hart, and Laemmli 1995; Tatlana I. Gerasimova et al. 1995; Bell, West, and Felsenfeld 1999; Gaszner, Vazquez, and Schedl 1999; Scott, Taubman, and Geyer 1999; Büchner et al. 2000; Pai et al. 2004; Melnikova et al. 2004; Moon et al. 2005). Except for CTCF, which is found throughout bilateria, all of these proteins appear to be specific to arthropods (Heger, George, and Wiehe 2013).

Staining of polytene chromosomes with antibodies against such insulator proteins showed that many of them localize to polytene interbands (Belyaeva and Zhimulev 1994; Zhao, Hart, and Laemmli 1995; Byrd and Corces 2003; Eggert, Gortchakov, and Saumweber 2004; Pai et al. 2004; Gortchakov et al. 2005; Gilbert, Tan, and Hart 2006; Berkaeva et al. 2009; Vatolina, Demakov, et al. 2011), with some enriched at interband borders. Further, some, though not all, insulator protein mutants disrupt polytene chromosome structure (Roy, Gilbert, and 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 (Lieberman-Aiden et al. 2009; Fullwood et al. 2009; Rao et al. 2014; Beagrie et al. 2017). Principle among these are derivatives of the chromosome conformation capture (3C) assay (Dekker et al. 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 common 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 largely 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 (Sexton et al. 2012; Hou et al. 2012; Van Bortle et al. 2014; Ulianov et al. 2015; L. Li et al. 2015). 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, and 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 substantially remodeled at enhancers and promoters during early development (Harrison et al. 2011; X.-Y. Li et al. 2014) and were interested in the role three-dimensional chromatin structure plays in spatial patterning.

We therefore generated high-resolution Hi-C datasets derived from nuclear cycle 14 *Drosophila melanogaster* embryos , and from the anterior and posterior halves of hand-dissected embryos at the same developmental stage. 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 - 2000 bp) genomic regions that are strongly bound by (usually multiple) insulator proteins and are sensitive to DNase digestion. Additionally, we find that boundaries share the molecular features of polytene interband regions. Finally, we show that for a region in which the fine polytene banding pattern has been mapped to genomic positions, boundaries show precise colocalization with interband regions that separate compacted bands corresponding to TADs. We propose that this relationship between insulators, TADs and polytene interbands extends across the genome, and suggest a model in which the decompaction of these regions drives the organization of interphase fly 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 replicatesof hand-sorted cellular blastoderm (mitotic cycle 14; mid-stage 5) 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 data from the anterior and posterior halves separately, also in duplicate. In total, we produced ~452 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 and Henikoff 1998; G. Wilkie A Shermoen P. O’Farrell 1999; Duan et al. 2010): 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, and 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 high resolution. We visually inspected genome-wide maps of a number of genomic regions constructed using bins of 500 bp, and were able to see a conspicuous pattern of TADs across a wide range of sizes, some smaller than 5 kb (**Fig. 2, Fig. 2--figure supplements 1-5**). When we compared maps for several of these regions with available functional genomic data from embryos, we observed that the boundaries between these domains showed a remarkably consistent pattern: they were formed by short regions of DNA (500-2000 bp) that are nearly always associated with high chromatin accessibility, measured by DNase-seq (X.-Y. Li et al. 2011), strong occupancy by known insulator proteins as measured by chromatin immunoprecipitation (ChIP) (Nègre et al. 2010) (**Fig. 2, Fig. 2--figure supplements 1-5**) properties characteristic of classical *Drosophila* 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 3,122 boundaries in the genome for nc14 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 4,000-5,500 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 (Lieberman-Aiden et al. 2009; Sexton et al. 2012; Rao et al. 2014; Crane et al. 2015). 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.

Attempts to exhaustively and definitively identify features within genomic data are necessarily variable due to differences in the choice of algorithm, parameters, cutoffs, and unavoidable tradeoffs between sensitivity and accuracy. We therefore sought a representative set of TAD boundaries with which to analyze features of these elements. . Of our top 1000 computationally-identified domain boundaries, we found that 952 were matched by a manually-called boundary within 1 kb. This high level of agreement suggested that the computational approach robustly identified the domain features that are apparent by eye. By taking the union of our computational calls, applied with a stringent cutoff, and our manual calls, we developed a very conservative set of exceptionally high confidence boundaries. We emphasize that this set represents only a subset of the boundaries identified by manual and computational approaches, the complete lists of which are provided in Tables S4 and S5.

Comparing these 952 boundaries to other genomic datasets confirms our initial observations and reveals 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 (95.1%) of the examined boundaries appear to be associated with CP190 binding (**Fig. 3--figure supplement 1**). 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 (45.1%) of stage 5 boundaries (**Fig. 3, Fig. 3--figure supplement 1**).

It is striking that we observe that not only are sites of combinatorial insulator protein binding enriched at TAD boundaries, but they are highly predictive. Of our representative set of boundaries, 95.1% are are enriched >2-fold for CP190 binding within a 1.5 kb window. Conversely, of the strongest 1000 CP190 peaks, 75.2% are within 2 kb of a manual or computationally-called boundary (compared to 37.4% of the top 1000 RNAPII peaks). It is important to note that we do identify a small subset of boundaries that are not apparently associated with sites of insulator binding (~1-2% show no enrichment for CP190, BEAF-32, or dCTCF, depending on thresholds used), suggesting that there are multiple

phenomena that can create topological boundaries in flies (e.g., see **Fig. 6**). However, the overwhelming majority of topological boundaries identified in this study coincide with elements that match the properties of CP190-associated insulators.

An important confounding factor in sorting out the nature of topological boundaries is the strong enrichment of insulator protein binding near promoters, specifically between divergent promoters. 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; Schwartz et al. 2012). 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**), raising the possibility that transcriptional activity at promoters may drive topological boundary formation. However, a number of features of the data argue against this possibility. First and most critically, many of the topological boundaries (54.9%) we identify are not associated with RNAPII binding in nc14 embryos. Similarly, there are many active promoters that do not appear to form topological boundaries (e.g., **Fig. 8** and supplements). Hug et al. (Hug et al. 2017) pharmacologically inhibited transcription in early embryos and observe that TADs remain intact. Finally, topological boundaries are invariant between anterior and posterior sections of embryos despite substantial differences in the transcriptional profiles of these regions (see below). We further 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. 3--figure supplement 2**) (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. Together, these data argue that boundaries constitute a distinct class of genetic elements that are not formed by promoter transcription, but are instead frequently located near promoters, possibly as a result of selective pressure to insulate these proximal promoters from distal regulatory elements. While we cannot rule out any role for promoter-bound transcription machinery in the formation of topological boundaries (notably, TFIIB is enriched at 69.1% of boundaries), we think it is unlikely that transcriptional activity plays a major role in establishing the topological domains of interphase fly chromosomes.

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 and Corces 2012), both of which show strong association with the set of boundary sequences as a whole (**Fig. 4**).

Boundary elements correspond to polytene interbands

The examination 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, and 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.

The Zhimulev laboratory has extensively studied the nature and composition of polytene bands and interbands for decades. Using a combination of approaches, they have identified interbands as a set of ~5700 short decompacted regions that tend to be located near divergent promoters and are characterized by DNase hypersensitivity and the binding of characteristic proteins, including insulator proteins (Zhimulev et al. 2014). It was immediately apparent to us that these elements bore significant similarity to the topological boundary elements we identified. We thus sought to compare our Hi-C data to known polytene chromosome structures.

There is surprisingly little data mapping features of polytene structure to specific genomic coordinates at high resolution. Vatolina et al. (Vatolina, Boldyreva, 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 (Vatolina, Boldyreva, 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 5 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). There is a striking correspondence between the assignments of Vatolina et al. and our Hi-C data: faint polytene bands correspond to TADs, and interbands correspond to the boundary elements that separate the TADs.

This correspondence is not perfect. Specifically, the evidence in our Hi-C data for the separation between the major band 10A1-2 and the minor band 10A3 is weak, though that may be partly explained by the absence of MboI cut sites obscuring much of this region. This minor band is barely detectable in polytene spreads, and the combination of this and weak support in Hi-C data may suggest that this band is not real or is perhaps only present in a minority of nuclei. Similarly, the Vatolina et al. report that they only rarely observe the interband between bands 10A6 and 10A7, and we indeed observe substantial contact between these two putative bands in Hi-C maps (the light orange region near the peak of the "pyramid" formed by 10A6 and 10A7 in Figure 5), though each shows stronger intra- than inter-domain interactions. One possible explanation for this observation is that the interband separating these two domains is not constitutive but rather is formed in only a fraction of nuclei. The pattern exhibited by these two domains--adjacent domains that show a clear separation but also a substantial interaction signal--is one we observe frequently in our early embryonic Hi-C data, suggesting that variable boundaries may be common features of the fly genome.

Overall, the alignment between polytene band mapping and Hi-C data in this region supports a strong correspondence between these two types of data. For five interbands which were easily visible in polytene spreads (10A3/4-5, 10A4-5/10A6, 10A7/10A8-9, 10A8-9/10A10, 10A10-11/10B1) , we observe strong domain boundaries in Hi-C data. For two interbands supported by weaker evidence in polytene analysis, we observe in Hi-C maps a weak or non-existent boundary (10A1-2/10A3) and a boundary with significant interaction across it, possibly representing heterogeneity between nuclei matching heterogeneity in polytenes (10A6/10A7).

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. 5--figure supplement 1**), an arrangement consistent with the correspondence of boundaries and interbands.

The chromodomain-containing protein Chriz has been suggested as strongest diagnostic feature of polytene interbands (Zhimulev et al. 2014). Using publicly available ChIP datasets from Kc167 cells (derived from late embryonic tissue), we observed a strong enrichment of Chriz binding at our representative boundaries (87.9% >2-fold enriched within 1.5 kb, **Fig. 5--figure supplement 2A**). Further, Hi-C directionality around Chriz peaks shows the characteristic pattern of boundary formation, and Chriz profiles at representative loci shows substantial correspondence between

boundary regions and Chriz binding (**Fig. 5--figure supplement 2B and** C), offering further support for the association between 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, Hartl, and Kornberg 2015). Our Hi-C data allows the detection of fine structure within these inter-TAD regions, down to individual boundary elements. Owing to the dearth of finely mapped polytene regions, the association between topological boundaries and interband regions is necessarily based on a limited number of example loci. However, the combination of data from these loci with the close agreement of the molecular composition of these regions, specifically the strong localization of the interband marker Chriz to topological boundaries, leads us to propose that the precise relationship between topological boundaries, insulator elements, and decompacted interband regions we observe is a general one, and that it extends across the genome..

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 5** (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, protect transgenes from position effect variegation and block the spread of chromatin silencing states. This chromatin extension model for fly insulator function can potentially explain these defining characteristics via simple physical separation.

Hi-C data can elaborate fine polytene structures

We reasoned that if our Hi-C data is capable of resolving fine banding patterns such as that at the 10A1-B1 locus, we should be able to resolve the borders of major bands with precision. We focused on a region of chromosome 2L that had previously been shown by Eagen et al. to appear as a single ~500 kb TAD using Hi-C at 15 kb resolution, but contains a faint interband in Bridge's map. Our Hi-C data reveal an intricate structure at this locus (**Fig. 6A**). There are two large TADs on the left and right, divided by a series of smaller domains in the center. We suspect that this middle region accounts for the interband in Bridge's map, in a manner similar to the 10A1-1/10B1-2 region: a complex region consisting

of several minor bands bounded by decompacted boundary regions appears as a single interband region under optical microscopy.

This region provides examples of a number of interesting features that we observe in our Hi-C data. First, the large TADs are bounded on both sides by gene-rich regions consisting of a number of smaller topological domains (**Fig. 6B, D**). The boundaries of large and small domains in this region nearly all share the common features of boundary elements: DNase hypersensitivity and binding of diagnostic insulator (e.g. CP190) and interband (CHRIZ) proteins. This region also contains a prominent example of an exception to this pattern: a loop is formed that appears to generate boundaries not associated with these characteristic protein binding events (**Fig. 6C**, indicated by dotted yellow lines and loop). This example highlights a critical point: while the description we provide of the association between TAD boundaries, insulator elements, and decompacted interbands appears to describe the overwhelming majority of cases, there are counter-examples. Indeed given the extraordinary capacity of nature to innovate with respect to gene regulation and structures, we expect that animal genomes will continue to provide novel chromosome topological and structural features for biologists to uncover for years to come.

Topological boundaries are nearly 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. 7A**). 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. 7B**).

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 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 et al. 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. 7C**). Additionally, DNase-seq data from later embryonic stages that feature substantial tissue differentiation, transcription, and chromatin changes show highly consistent profiles at boundaries (**Fig. 7C, Fig. 7--figure supplement 1)**. It is also striking that we observe significant enrichment of insulator proteins and Chriz at boundaries, despite the fact that boundaries were identified from Hi-C data from carefully-staged nc14 embryos (2-3 hours), whereas these ChIP datasets are derived from 0-12 hour old embryos or late embryonic cultured cells (Chriz). Together, these results are consistent with a model in which insulator-mediated chromatin organization is a constitutive feature of interphase chromatin of *D. melanogaster* embryos.

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, and Jaynes 2009; Yang and Corces 2012; Kyrchanova and Georgiev 2014; Kravchenko et al. 2005), and a substantial literature exists demonstrating that many insulator proteins are able to interact with each other and to self-associate (Büchner et al. 2000; Gause, Morcillo, and Dorsett 2001; Ghosh, Gerasimova, and Corces 2001; Blanton, Gaszner, and Schedl 2003; Pai et al. 2004; Mohan et al. 2007; Golovnin et al. 2007; Vogelmann et al. 2014). In general, we do not observe looping interactions between domain boundaries in our Hi-C data. However, during manual calling of topological boundaries for the entire genome, we noted 46 prominent examples of interactions between non-adjacent domains (**Fig. 8 and Fig. 8--figure supplements 1-10, Table S3**), in addition to the previously noted clustering of PcG-regulated Hox gene clusters (Sexton et al. 2012). Because the interactions we observed were not of a uniform character,

we did not attempt to computationally search for all such phenomena in our data, nor do we claim that this list is necessarily complete. It is merely the union of two sets of "interesting" loci identified in two independent rounds of visual inspection Hi-C maps for the entire genome, and we feel it is informative with respect to the nature and significance of distal interaction in the fly embryo.

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. 8A**). 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, Stathopoulos, and Levine 2002; Calhoun and Levine 2003). 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. 8A**: 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. 8--figure supplement 1**). We also observe a similar arrangement near the *eve* locus (**Fig. 8--figure supplement 2**). 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 8B**, one of these loops brings together the promoters of *kni* and the related *knrl* genes. Other loops connect the *achaete* and *scute* genes (**Fig. 8--figure supplement 3**), *slp1* and *slp2* (**Fig. 8--figure supplement 4**), and the promoter of *Ubx* with an element in its first intron (**Fig. 8--figure supplement 5**).

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.

Discussion

Several Hi-C studies in flies have identified enrichments of insulator proteins at TAD boundaries (Sexton et al. 2012; Ulianov et al. 2015; Eagen, Hartl, and Kornberg 2015; Mourad and Cuvier 2016) These studies varied in their resolution (due to use of 4- vs. 6-cutter enzymes and sequencing depth), methods (solution vs. in situ Hi-C), and, critically, in the methods used to identify TAD boundaries. As a result, each study relied on distinct sets of boundaries for analyses of the molecular features of these structures. We explored several methods to identify topological domains and associated boundaries and found that no single approach was sufficient to exhaustively identify all of these features in the genome. Rather, by using a combination of visual inspection of Hi-C maps at a large number of loci, unbiased hand-calling, and computational searches, we consistently observed a very close, two-way association between sites of combinatorial insulator protein binding (insulators) and the boundaries between topological domains. This result supports prior studies which found enriched insulator protein binding at topological boundaries, and extends this finding by localizing boundaries to discrete insulator elements. Hi-C data are exceptionally complex and reveal many layers of genomic organization, and we suspect that many questions in this field will only be resolved by the combined work of multiple groups using distinct analysis strategies and techniques.

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.

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 (Zhimulev 1996; Vatolina, Boldyreva, et al. 2011), 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 (Pai et al. 2004; Tatiana I. Gerasimova et al. 2007), which is precisely what we observe for boundary elements. Third, interbands and boundary elements are highly sensitive to DNase digestion (Vatolina, Demakov, et al. 2011). Fourth, interbands have been shown to contain the promoters and 5' ends of genes (Jamrich,

Greenleaf, and Bautz 1977; H. Sass 1982; Heinz Sass and Bautz 1982; H. Sass and Bautz 1982; Rykowski et al. 1988), 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, Gilbert, and Hart 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 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 attempts to unify years of studies of polytene chromosomes with modern genomic methods (**Fig. 9**). 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, exploring how it fits into hierarchical levels of genome organization, and understanding its implications for genome function.

Materials and methods

Embryo collection, sorting, and sectioning

OregonR strain *D. melanogaster (RRID:FlyBase_FBst1000080)* 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-10, 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). 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 -80°C 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 (Lieberman-Aiden et al. 2009; Sexton et al. 2012; Rao et al. 2014; Crane et al. 2015) in addition to procedures we developed independently. All analysis was performed with custom Python, R, and Perl scripts (Stadler 2017) except where noted. Single-ends of demultiplexed reads were separately mapped using Bowtie (B. Langmead et al. 2009) (parameters: -m1 --best --strata) to the *D. melanogaster* genome dm6 (R6.17). 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 5' 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.

Computational 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 and 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.

Sample size determination

No explicit statistical method was used to compute sample size. All unique experiments were prepared in duplicate.

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.

Figure Legends

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 values of the normalized countswere histogram equalized, and maximum and minimum values were adjusted for optimal display.

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 3R:24924500-25174500 (dm3: 3R:20750000-20999999). 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.

Figure 2--figure supplement 1: High resolution Hi-C maps of additional example genomic regions from stage 5 *Drosophila melanogaster* embryos. Region 2L:10325000-10499999 (dm6 and dm3). Data were prepared identically to Figure 2, and all data is from the same sources.

Figure 2--figure supplement 2: High resolution Hi-C maps of additional example genomic regions from stage 5 *Drosophila melanogaster* embryos. Region 2R:23932500-24112500 (dm3: 2R:19820000-19999999). Data were prepared identically to Figure 2, and all data is from the same sources.

Figure 2--figure supplement 3: High resolution Hi-C maps of additional example genomic regions from stage 5 *Drosophila melanogaster* embryos. Region 3L:18806900-190569000 (dm3: 3L:18800000-19050000). Data were prepared identically to Figure 2, and all data is from the same sources.

Figure 2--figure supplement 4: High resolution Hi-C maps of additional example genomic regions from stage 5 *Drosophila melanogaster* embryos. Region 2R:9124500-9223500 (dm3: 2R:5012000-5111000). Data were prepared identically to Figure 2, and all data is from the same sources.

Figure 2--figure supplement 5: High resolution Hi-C maps of additional example genomic regions from stage 5 *Drosophila melanogaster* embryos. Region X:15606000-15856000 (dm3: X:15500000-15749999). Data were prepared identically to Figure 2, and all data is from the same sources.

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 952 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 Figure 3--figure supplement 2. Percentages are calculated as the percentage of boundaries with a >2-fold enrichment for the given signal within a 3 kb window centered on the boundary (±1.5 kb). 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).

Figure 3--figure supplement 1: 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).

Figure 3--figure supplement 2: Genomic signals around H3K4me3 peaks. Data were prepared and displayed identically to Figure 3 for the top 1000 peaks of H3K4me3 in stage 5 embryos, as identified by (X.-Y. Li et al. 2014). H3K4me3 is a proxy for active promoters.

Figure 3--figure supplement 3: Directionality around peaks of genomic features. Hi-C directionality scores in 500 bp bins from aggregated nc14 data were plotted around the top 1000 (where available) peaks of the genomic features shown

637 in Figure 3. The signature of a boundary is a region of left-bias (white) transitioning sharply to a region of right-bias
638 (black).
639
640 **Figure 3--figure supplement 4: Genomic signals around H3K4me1 peaks.** Data were prepared and displayed
641 identically to Figure 3 for the top 1000 peaks of H3K4me3 in stage 5 embryos, as identified by (X.-Y. Li et al. 2014).
642 H3K4me1 is a proxy for active enhancers.
643
644 **Figure 4: Sequence features of TAD boundary elements. (A)** Histograms showing the frequency of enriched 7-mers in
645 5 kb windows around 952 high-confidence TAD boundaries. (**B**) Scatter plots of occurrences of words matching known
646 BEAF-32 binding motifs (left) and CACA motif (right) in 10 kb windows around high-confidence TAD boundaries.
647 Points are plotted with low opacity, such that darker points correspond to positions where multiple words occurs close
648 together in sequence.
649
650 **Figure 5: Topological boundary elements correspond to polytene interbands.** Heat map of aggregate Hi-C data for all
651 nc14 datasets binned at 500 bp and UCSC browser data shown for the region X:11077500-11181000 (dm3: X:10971500-
652 11075000) for which Vatolina et al. provided fine-mapping of polytene banding structure. Hi-C and browser data were
653 prepared and sourced as indicated in Fig. 2. Dashed red lines are visual guides drawn from the interband assignments of
654 Vatolina et al. Top: accurately-scaled representations of the size of the mapped bands and interbands in base pairs
655 ("Genomic") and the corresponding physical distances in polytene chromosomes derived from electron microscopic
656 analysis of polytene chromosomes by Vatolina et al. Increased relative physical size of interband regions demonstrates
657 their lower compaction ratios.
658
659 **Figure 5--figure supplement 1: TAD structure corresponds to mapped polytene structure at the *Notch* locus.** Hi-C
660 map of the *Notch* locus, prepared as in Figure 5.
661 **Figure 5--figure supplement 2: Chriz protein binding in Kc167 cells is highly enriched at sites of embryonic nc14**
662 **topological boundaries.** (**A**) Heatmaps were prepared as in Fig. 3 using ChIP data for Chriz binding (modENCODE
663 dataset 277) in Kc167 cells around 952 representative nc14 boundaries, sorted either by boundary strength (left) or Chriz
664 signal. (**B**) Directionality of aggregate nc14 Hi-C data around top 1000 Chriz peaks. (**C**) Chriz binding compared to Hi-C

maps for two example regions: X:11077500-11181000 (dm3: X: 10971500-11075000) and 3R:24924500-25174500 (dm3: 3R:20750000-20999999).

Figure 6: Complex topological structure of a region of chromosome 2L. Hi-C maps using 500 bp bins of the region of chromosome 2L corresponding to polytene band 22A1-2. This regions was shown by Eagen et al. to comprise a single TAD in Hi-C data viewed at 15 kb resolution, and is occasionally observed to contain an interband in polytene spreads. (**A**) View of the entire region, revealing complex internal structure. (**B-D**) Zoomed-in views of three regions comprising the left border (B), complex middle section (C), and right border of the larger region corresponding to the band/TAD investigated by Eagen et al., with associated stage 5 DNase accessibility, CP190, and Chriz (kc167 cells) profiles. Coordinates for this region are identical in dm3 and dm6.

Figure 7: 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 952 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. (dm6: X:11077500-11181000, dm3: X: 10971500-11075000), the example region from Fig. 2 supplement 4 (dm6: 2R:9124500-9223500, dm3: 2R:5012000-5111000)), and the genomic regions surrounding the *eve* (dm6: 2R:9903060-10056959, dm3: 2R:5790565-5944464) and *ftz* (dm6: 3R:6769234-6961333, dm3: 3R:2594956-2787055) 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).

Figure 7--figure supplement 1: Developmental time series of DNase accessibility at TAD boundaries. DNase accessibility data from (X.-Y. Li et al. 2011) was plotted around 952 high-confidence boundaries and top 952 H3K4me1 peaks. Data covers stage 5 (~2 hours) through stage 14 (~10 hours) of embryonic development.

Figure 8: 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, Stathopoulos, and Levine 2002; Calhoun and Levine 2003). 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 et al. 1994).

Figure 8--figure supplement 1: Distal chromatin contacts in stage 5 embryos. Hi-C maps prepared as in Figure 8 for manually-curated distal interactions. Data shown at 500 bp resolution, all nc14 data aggregated. Gene tracks are from the UCSC genome browser and shown with dm6 coordinates. A complete list of observed distal interactions is available in Table S3, along with dm3 coordinates. .

Figure 9: 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 enhancer-promoter interactions, and limiting the spread of chromatin silencing states.

Supplementary File 1: Supplemental tables. S1: List of samples used to generate Hi-C libraries. S2: Sequencing and processing statistics for Hi-C libraries. S3: Location of loops and domain-skipping identified in nc14 embryos. S4: Manually called boundaries. S5: Computationally identified boundaries. S6: Representative boundary set, merge of manual and computational curations.

Supplementary File 2: Source data for ChIP and similar enrichment files, e.g. in Figure 3.

Supplementary File 3: Source data for Hi-C heatmaps.

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: 637–41, 662–66.

———. 1890. “Sur La Structure Intime Du Noyau Du Loxophyllum Meleagris.” *Zoologischer Anzeiger* 13: 110–15, 132–36.

Beagrie, Robert A., Antonio Scialdone, Markus Schueler, Dorothee C. A. Kraemer, Mita Chotalia, Sheila Q. Xie, Mariano Barbieri, et al. 2017. “Complex Multi-Enhancer Contacts Captured by Genome Architecture Mapping.” *Nature*. Nature Publishing Group. doi:10.1038/nature21411.

Bell, A. C., A. G. West, and G. Felsenfeld. 1999. “The Protein CTCF Is Required for the Enhancer Blocking Activity of Vertebrate Insulators.” *Cell* 98 (3): 387–96.

Belyaeva, E. S., and I. F. Zhimulev. 1994. “Molecular and Cytogenetical Characterization of the 1OA1-2 Band and Adjoining.” *Genetics* 136 (1): 1063–73.

Benyajati, Cheeptip, and Abraham Worcel. 1976. “Isolation, Characterization, and Structure of the Folded Interphase Genome of Drosophila Melanogaster.” *Cell* 9 (3): 393–407.

Berkaeva, Maria, Sergei Demakov, Yuri B. Schwartz, and Igor Zhimulev. 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): 745–54.

Blanton, Jason, Miklos Gaszner, and Paul Schedl. 2003. “Protein:protein Interactions and the Pairing of Boundary Elements in Vivo.” *Genes & Development* 17 (5): 664–75.

Blythe, Shelby A., and Eric F. Wieschaus. 2016. “Establishment and Maintenance of Heritable Chromatin Structure during Early Drosophila Embryogenesis.” *eLife* 5 (November). doi:10.7554/eLife.20148.

Bridges, Calvin Blackman. 1934. “Salivary Chromosome Maps with a Key to the Banding of the Chromosomes of Drosophila Melanogaster.” *The Journal of Heredity*, 60–64.

Büchner, K., P. Roth, G. Schotta, V. Krauss, H. Saumweber, G. Reuter, and R. Dorn. 2000. “Genetic and Molecular Complexity of the Position Effect Variegation Modifier mod(mdg4) in Drosophila.” *Genetics* 155 (1): 141–57.

Buenrostro, Jason D., Paul G. Giresi, Lisa C. Zaba, Howard Y. Chang, and William J. Greenleaf. 2013. “Transposition of Native

Chromatin for Fast and Sensitive Epigenomic Profiling of Open Chromatin, DNA-Binding Proteins and Nucleosome Position." *Nature Methods* 10 (12): 1213–18.

Byrd, Keith, and Victor G. Corces. 2003. "Visualization of Chromatin Domains Created by the Gypsy Insulator of Drosophila." *The Journal of Cell Biology* 162 (4): 565–74.

Calhoun, Vincent C., and Michael Levine. 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): 9878–83.

Calhoun, Vincent C., Angelike Stathopoulos, and Michael Levine. 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): 9243–47.

Contrino, Sergio, Richard N. Smith, Daniela Butano, Adrian Carr, Fengyuan Hu, Rachel Lyne, Kim Rutherford, et al. 2012. "modMine: Flexible Access to modENCODE Data." *Nucleic Acids Research* 40 (D1): 1082–88.

Crane, Emily, Qian Bian, Rachel Patton McCord, Bryan R. Lajoie, Bayly S. Wheeler, Edward J. Ralston, Satoru Uzawa, Job Dekker, and Barbara J. Meyer. 2015. "Condensin-Driven Remodelling of X Chromosome Topology during Dosage Compensation." *Nature* 523 (7559): 240–44.

Csink, A. K., and S. Henikoff. 1998. "Large-Scale Chromosomal Movements during Interphase Progression in Drosophila." *The Journal of Cell Biology* 143 (1): 13–22.

Dekker, Job, Karsten Rippe, Martijn Dekker, and Nancy Kleckner. 2002. "Capturing Chromosome Conformation." *Science* 295 (5558): 1306–11.

Duan, Zhijun, Mirela Andronescu, Kevin Schutz, Sean McIlwain, Yoo Jung Kim, Choli Lee, Jay Shendure, Stanley Fields, C. Anthony Blau, and William S. Noble. 2010. "A Three-Dimensional Model of the Yeast Genome." *Nature* 465 (7296). Nature Publishing Group: 363–67.

Eagen, Kyle P., Tom A. Hartl, and Roger D. Kornberg. 2015. "Stable Chromosome Condensation Revealed by Chromosome Conformation Capture." *Cell* 163 (4). Elsevier Inc.: 934–46.

Eggert, Harald, Andrej Gortchakov, and Harald Saumweber. 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): 4253–64.

Foe, Victoria E., and Bruce M. Alberts. 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.

Fujioka, Miki, Xian Wu, and James B. Jaynes. 2009. "A Chromatin Insulator Mediates Transgene Homing and Very Long-Range Enhancer-Promoter Communication." *Development* 136 (18): 3077–87.

Fullwood, Melissa J., Mei Hui Liu, You Fu Pan, Jun Liu, Han Xu, Yusoff Bin Mohamed, Yuriy L. Orlov, et al. 2009. "An Oestrogen-

Receptor-Alpha-Bound Human Chromatin Interactome." *Nature* 462 (7269). Nature Publishing Group: 58–64.

Gaszner, M., J. Vazquez, and P. Schedl. 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.

Gause, Maria, Patrick Morcillo, and Dale Dorsett. 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): 4807–17.

Gavrilov, Alexey a., Ekaterina S. Gushchanskaya, Olga Strelkova, Oksana Zhironkina, Igor I. Kireev, Olga V. Iarovaia, and Sergey V. Razin. 2013. "Disclosure of a Structural Milieu for the Proximity Ligation Reveals the Elusive Nature of an Active Chromatin Hub." *Nucleic Acids Research* 41 (6): 3563–75.

Gavrilov, Alexey, Sergey V. Razin, and Giacomo Cavalli. 2015. "In Vivo Formaldehyde Cross-Linking: It Is Time for Black Box Analysis." *Briefings in Functional Genomics* 14 (2): 163–65.

Gerasimova, Tatiana I., Elissa P. Lei, Ashley M. Bushey, and Victor G. Corces. 2007. "Coordinated Control of dCTCF and Gypsy Chromatin Insulators in Drosophila." *Molecular Cell* 28 (5): 761–72.

Gerasimova, Tatlana I., David A. Gdula, Denis V. Gerasimov, Olga Simonova, and Victor G. Corces. 1995. "A Drosophila Protein That Imparts Directionality on a Chromatin Insulator Is an Enhancer of Position-Effect Variegation." *Cell* 82 (4): 587–97.

Geyer, P. K., and V. G. Corces. 1992. "DNA Position-Specific Repression of Transcription by a Drosophila Zinc Finger Protein." *Genes and Development* 6 (10): 1865–73.

Ghosh, D., T. I. Gerasimova, and V. G. Corces. 2001. "Interactions between the Su(Hw) and Mod(mdg4) Proteins Required for Gypsy Insulator Function." *The EMBO Journal* 20 (10): 2518–27.

Gilbert, Matthew K., Yian Yee Tan, and Craig M. Hart. 2006. "The Drosophila Boundary Element-Associated Factors BEAF-32A and BEAF-32B Affect Chromatin Structure." *Genetics* 173 (3): 1365–75.

Golovnin, Anton, Alexander Mazur, Marina Kopantseva, Maria Kurshakova, Pavel V. Gulak, Brian Gilmore, William G. F. Whitfield, Pamela Geyer, Vincenzo Pirrotta, and Pavel Georgiev. 2007. "Integrity of the Mod(mdg4)-67.2 BTB Domain Is Critical to Insulator Function in Drosophila Melanogaster." *Molecular and Cellular Biology* 27 (3): 963–74.

González-Gaitán, M., M. Rothe, E. A. Wimmer, H. Taubert, and H. Jäckle. 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): 8567–71.

Gortchakov, A. A., H. Eggert, M. Gan, J. Mattow, I. F. Zhimulev, and H. Saumweber. 2005. "Chriz, a Chromodomain Protein Specific for the Interbands of Drosophila Melanogaster Polytene Chromosomes." *Chromosoma* 114 (1): 54–66.

Guo, Ya, Quan Xu, Daniele Canzio, Jia Shou, Jinhuan Li, David U. Gorkin, Inkyung Jung, et al. 2015. "CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function." *Cell* 162 (4). Elsevier: 900–910.

G. Wilkie A Shermoen P. O'Farrell, I. Davis. 1999. "Transcribed Genes Are Localized according to Chromosomal Position within Polarized Drosophila Embryonic Nuclei." *Current Biology: CB* 9 (21): 1263–66.

Harrison, Melissa M., Xiao-Yong Li, Tommy Kaplan, Michael R. Botchan, and Michael B. Eisen. 2011. "Zelda Binding in the Early Drosophila Melanogaster Embryo Marks Regions Subsequently Activated at the Maternal-to-Zygotic Transition." *PLoS Genetics* 7 (10): e1002266.

Heger, Peter, Rebecca George, and Thomas Wiehe. 2013. "Successive Gain of Insulator Proteins in Arthropod Evolution." *Evolution; International Journal of Organic Evolution* 67 (10): 2945–56.

Heitz, E., and H. Bauer. 1933. "Beweise Fur Die Chromosomennatur Der Kernschleifen in Den Knauelkernen von Bibio Hortulanus." *Zeitschrift Fur Zellforschung* 17: 67–82.

Holdridge, C., and D. Dorsett. 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.

Hou, Chunhui, Li Li, Zhaohui S. Qin, and Victor G. Corces. 2012. "Gene Density, Transcription, and Insulators Contribute to the Partition of the Drosophila Genome into Physical Domains." *Molecular Cell* 48 (3): 471–84.

Hug, Clemens B., Alexis G. Grimaldi, Kai Kruse, and Juan M. Vaquerizas. 2017. "Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription." *Cell* 169 (2). Elsevier: 216–28.e19.

Jamrich, M., A. L. Greenleaf, and E. K. Bautz. 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): 2079–83.

Kahn, Tatyana G., Yuri B. Schwartz, Gaetano I. Dellino, and Vincenzo Pirrotta. 2006. "Polycomb Complexes and the Propagation of the Methylation Mark at the Drosophila Ubx Gene." *The Journal of Biological Chemistry* 281 (39): 29064–75.

Kellum, Rebecca, and Paul Schedl. 1991. "A Position-Effect Assay for Boundaries of Higher Order Chromosomal Domains." *Cell* 64 (5): 941–50.

Kellum, R., and P. Schedl. 1992. "A Group of Scs Elements Function as Domain Boundaries in an Enhancer-Blocking Assay." *Molecular and Cellular Biology* 12 (5): 2424–31.

Kent, W. James, Charles W. Sugnet, Terrence S. Furey, Krishna M. Roskin, Tom H. Pringle, Alan M. Zahler, and David Haussler. 2002. "The Human Genome Browser at UCSC." *Genome Research* 12 (6): 996–1006.

King, R. L., and H. W. Beams. 1934. "Somatic Synapsis in Chironomus, with Special Reference to the Individuality of Chromosomes." *Journal of Morphology* 5 (3): 577–91.

Kravchenko, Elena, Ekaterina Savitskaya, Oksana Kravchuk, Alexander Parshikov, Pavel Georgiev, and Mikhail Savitsky. 2005. "Pairing between Gypsy Insulators Facilitates the Enhancer Action in Trans throughout the Drosophila Genome." *Molecular and Cellular Biology* 25 (21): 9283–91.

Kyrchanova, Olga, and Pavel Georgiev. 2014. "Chromatin Insulators and Long-Distance Interactions in Drosophila." *FEBS Letters*

588 (1). Federation of European Biochemical Societies: 8–14.

Laird, Charles D., and W. Yean Chooi. 1976. “Morphology of Transcription Units in Drosophila Melanogaster.” *Chromosoma* 58 (2): 193–218.

Laird, Charles D., Linda E. Wilkinson, Victoria E. Foe, and W. Yean Chooi. 1976. “Analysis of Chromatin-Associated Fiber Arrays.” *Chromosoma* 58 (2): 169–90.

Langmead, Ben, and Steven L. Salzberg. 2012. “Fast Gapped-Read Alignment with Bowtie 2.” *Nature Methods* 9 (4): 357–59.

Langmead, B., C. Trapnell, M. Pop, and S. Salzberg. 2009. “Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome.” *Genome Biology* 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, Ed. 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, Erez, Nynke L. van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, et al. 2009. “Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome.” *Science* 326 (5950): 289–93.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. “The Sequence Alignment/Map Format and SAMtools.” *Bioinformatics* 25 (16): 2078–79.

Li, Li, Xiaowen Lyu, Chunhui Hou, Naomi Takenaka, Huy Q. Nguyen, Chin Tong Ong, Caelin Cubeñas-Potts, et al. 2015. “Widespread Rearrangement of 3D Chromatin Organization Underlies Polycomb-Mediated Stress-Induced Silencing.” *Molecular Cell* 58 (2): 216–31.

Lindsley, D. L., and E. H. Grell. 1968. “Genetic Variations of Drosophila Melanogaster.” *Carnegie Inst. Wash. Publication*. http://garfield.library.upenn.edu/classics1983/A1983QS51100002.pdf.

Li, Xiao-Yong, Melissa M. Harrison, Jacqueline E. Villalta, Tommy Kaplan, and Michael B. Eisen. 2014. “Establishment of Regions of Genomic Activity during the Drosophila Maternal to Zygotic Transition.” *eLife* 3: e03737.

Li, Xiao Yong, Stewart MacArthur, Richard Bourgon, David Nix, Daniel A. Pollard, Venky N. Iyer, Aaron Hechmer, et al. 2008. “Transcription Factors Bind Thousands of Active and Inactive Regions in the Drosophila Blastoderm.” *PLoS Biology* 6 (2): 0365–88.

Li, Xiao-Yong, Sean Thomas, Peter J. Sabo, Michael B. Eisen, John a. Stamatoyannopoulos, and Mark D. Biggin. 2011. “The Role of Chromatin Accessibility in Directing the Widespread, Overlapping Patterns of Drosophila Transcription Factor Binding.” *Genome Biology* 12 (4). BioMed Central Ltd: R34.

Lowenstein, Michael G., Thomas D. Goddard, and John W. Sedat. 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): 5678–92.

Lupiáñez, Darío G., Katerina Kraft, Verena Heinrich, Peter Krawitz, Francesco Brancati, Eva Klopocki, Denise Horn, et al. 2015. “Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions.” *Cell* 161 (5): 1012–25.

Mallin, Daniel R., Jane S. Myung, J. Scott Patton, and Pamela K. Geyer. 1998. “Polycomb Group Repression Is Blocked by the Drosophila Suppressor of Hairy-Wing [SU(HW)] Insulator.” *Genetics* 148 (1): 331–39.

McKnight, Steven L., and Oscar L. Miller. 1976. “Ultrastructural Patterns of RNA Synthesis during Early Embryogenesis of Drosophila Melanogaster.” *Cell* 8 (2): 305–19.

———. 1977. “Electron Microscopic Analysis of Chromatin Replication in the Cellular Blastoderm Drosophila Melanogaster Embryo.” *Cell* 12 (3): 795–804.

Melnikova, Larisa, François Juge, Natalia Gruzdeva, Aleksandr Mazur, Giacomo Cavalli, and Pavel Georgiev. 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): 14806–11.

Mohan, Man, Marek Bartkuhn, Martin Herold, Angela Philippen, Nina Heinl, Imke Bardenhagen, Joerg Leers, et al. 2007. “The Drosophila Insulator Proteins CTCF and CP190 Link Enhancer Blocking to Body Patterning.” *The EMBO Journal* 26 (19): 4203–14.

Moon, Hanlim, Galina Filippova, Dmitry Loukinov, Elena Pugacheva, Qi Chen, Sheryl T. Smith, Adam Munhall, et al. 2005. “CTCF Is Conserved from Drosophila to Humans and Confers Enhancer Blocking of the Fab-8 Insulator.” *EMBO Reports* 6 (2): 165–70.

Mourad, Raphaël, and Olivier Cuvier. 2016. “Computational Identification of Genomic Features That Influence 3D Chromatin Domain Formation.” *PLoS Computational Biology* 12 (5). Public Library of Science: e1004908.

Nègre, Nicolas, Christopher D. Brown, Parantu K. Shah, Pouya Kheradpour, Carolyn A. Morrison, Jorja G. Henikoff, Xin Feng, et al. 2010. “A Comprehensive Map of Insulator Elements for the Drosophila Genome.” *PLoS Genetics* 6 (1): e1000814.

Newport, J., and M. Kirschner. 1982. “A Major Developmental Transition in Early Xenopus Embryos: I. Characterization and Timing of Cellular Changes at the Midblastula Stage.” *Cell* 30 (3): 675–86.

Pai, Chi Y., Elissa P. Lei, Dolanchanpa Ghosh, and Victor G. Corces. 2004. “The Centrosomal Protein CP190 Is a Component of the Gypsy Chromatin Insulator.” *Molecular Cell* 16 (5): 737–48.

Painter, Theophilus 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., and V. G. Corces. 1985. “Forked, Gypsys, and Suppressors in Drosophila.” *Cell* 41 (2): 429–37.

———. 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.

Parkhurst, S. M., D. A. Harrison, M. P. Remington, C. Spana, R. L. Kelley, R. S. Coyne, and V. G. Corces. 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): 1205–15.

Rabinowitz, M. 1941. “Studies on the Cytology and Early Embryology of the Egg of Drosophila Melanogaster.” *Journal of Morphology* 69. doi:10.1002/jmor.1050690102.

Ramirez, Fidel, Vivek Bhardwaj, Jose Villaveces, Laura Arrigoni, Bjoern A. Gruening, Kin Chung Lam, Bianca Habermann, Asifa Akhtar, and Thomas Manke. 2017. “High-Resolution TADs Reveal DNA Sequences Underlying Genome Organization in Flies.” *bioRxiv* 49 (0): 115063.

Rao, Suhas S. P., Miriam H. Huntley, Neva C. Durand, Elena K. Stamenova, Ivan D. Bochkov, James T. Robinson, Adrian L. Sanborn, et al. 2014. “A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping.” *Cell*, December. Elsevier Inc., 1–16.

Recillas-Targa, Félix, Michael J. Pikaart, Bonnie Burgess-Beusse, Adam C. Bell, Michael D. Litt, Adam G. West, Miklos Gaszner, and Gary Felsenfeld. 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): 6883–88.

Roseman, R. R., V. Pirrotta, and P. K. Geyer. 1993. “The su(Hw) Protein Insulates Expression of the Drosophila Melanogaster White Gene from Chromosomal Position-Effects.” *The EMBO Journal* 12 (2): 435–42.

Roy, Swarnava, Matthew K. Gilbert, and Craig M. Hart. 2007. “Characterization of BEAF Mutations Isolated by Homologous Recombination in Drosophila.” *Genetics* 176 (2): 801–13.

Rykowski, M. C., S. J. Parmelee, D. A. Agard, and J. W. Sedat. 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): 461–72.

Sass, H. 1982. “RNA Polymerase B in Polytene Chromosomes: Immunofluorescent and Autoradiographic Analysis during Stimulated and Repressed RNA Synthesis.” *Cell* 28 (2): 269–78.

Sass, H., and E. K. Bautz. 1982. “Interbands of Polytene Chromosomes: Binding Sites and Start Points for RNA Polymerase B (II).” *Chromosoma* 86 (1): 77–93.

Sass, Heinz, and Ekkehard K. F. Bautz. 1982. “Immunoelectron Microscopic Localization of RNA Polymerase B on Isolated Polytene Chromosomes of Chironomus Tentans.” *Chromosoma* 85 (5). Springer-Verlag: 633–42.

Schwartz, Yuri B., Daniela Linder-Basso, Peter V. Kharchenko, Michael Y. Tolstorukov, Maria Kim, Hua-Bing Li, Andrey a. Gorchakov, et al. 2012. “Nature and Function of Insulator Protein Binding Sites in the Drosophila Genome.” *Genome Research* 22 (11): 2188–98.

Scott, K. C., A. D. Taubman, and P. K. Geyer. 1999. “Enhancer Blocking by the Drosophila Gypsy Insulator Depends upon Insulator Anatomy and Enhancer Strength.” *Genetics* 153 (2): 787–98.

Sexton, Tom, Eitan Yaffe, Ephraim Kenigsberg, Frédéric Bantignies, Benjamin Leblanc, Michael Hoichman, Hugues Parrinello, Amos Tanay, and Giacomo Cavalli. 2012. “Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome.” *Cell* 148 (3): 458–72.

Sigrist, Christian J. A., and Vincenzo Pirrotta. 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): 209–21.

Spana, C., D. A. Harrison, and V. G. Corces. 1988. “The Drosophila Melanogaster Suppressor of Hairy-Wing Protein Binds to Specific Sequences of the Gypsy Retrotransposon.” *Genes & Development* 2 (11): 1414–23.

Stadler, Michael. 2017. “Hi-C.” *Bitbucket*. https://bitbucket.org/Stadlerlab/hi-c/. 95e304b.

Tadros, Wael, and Howard D. Lipshitz. 2009. “The Maternal-to-Zygotic Transition: A Play in Two Acts.” *Development* 136 (18). The Company of Biologists Ltd: 3033–42.

Ulianov, Sergey V., Ekaterina E. Khrameeva, Alexey a. Gavrilov, Ilya M. Flyamer, Pavel Kos, Elena a. Mikhaleva, Aleksey a. Penin, et al. 2015. “Active Chromatin and Transcription Play a Key Role in Chromosome Partitioning into Topologically Associating Domains.” *Genome Research* 26 (1): 70–84.

Van Bortle, Kevin, Michael H. Nichols, Li Li, Chin-Tong Ong, Naomi Takenaka, Zhaohui S. Qin, and Victor G. Corces. 2014. “Insulator Function and Topological Domain Border Strength Scale with Architectural Protein Occupancy.” *Genome Biology* 15 (5): R82.

Vatolina, Tatyana Yu, Lidiya V. Boldyreva, Olga V. Demakova, Sergey A. Demakov, Elena B. Kokoza, Valeriy F. Semeshin, Vladimir N. Babenko, Fedor P. Goncharov, Elena S. Belyaeva, and Igor F. Zhimulev. 2011. “Identical Functional Organization of Nonpolytene and Polytene Chromosomes in Drosophila Melanogaster.” *PloS One* 6 (10). doi:10.1371/journal.pone.0025960.

Vatolina, Tatyana Yu, S. A. Demakov, V. F. Semeshin, I. V. Makunin, V. N. Babenko, E. S. Belyaeva, and I. F. Zhimulev. 2011. “Identification and Molecular Genetic Characterization of the Polytene Chromosome Interbands in Drosophila Melanogaster.” *Russian Journal of Genetics* 47 (5): 521–32.

Vlassova, Inna E., Galima H. Umbetova, Vladimir H. Zimmermann, Carlos Alonso, Elena S. Belyaeva, and Igor F. Zhimulev. 1985. “Immunofluorescence Localization of DNA:RNA Hybrids in Drosophila Melanogaster Polytene Chromosomes.” *Chromosoma* 91 (3-4): 251–58.

Vogelmann, Jutta, Antoine Le Gall, Stephanie Dejardin, Frederic Allemand, Adrien Gamot, Gilles Labesse, Olivier Cuvier, et al. 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.

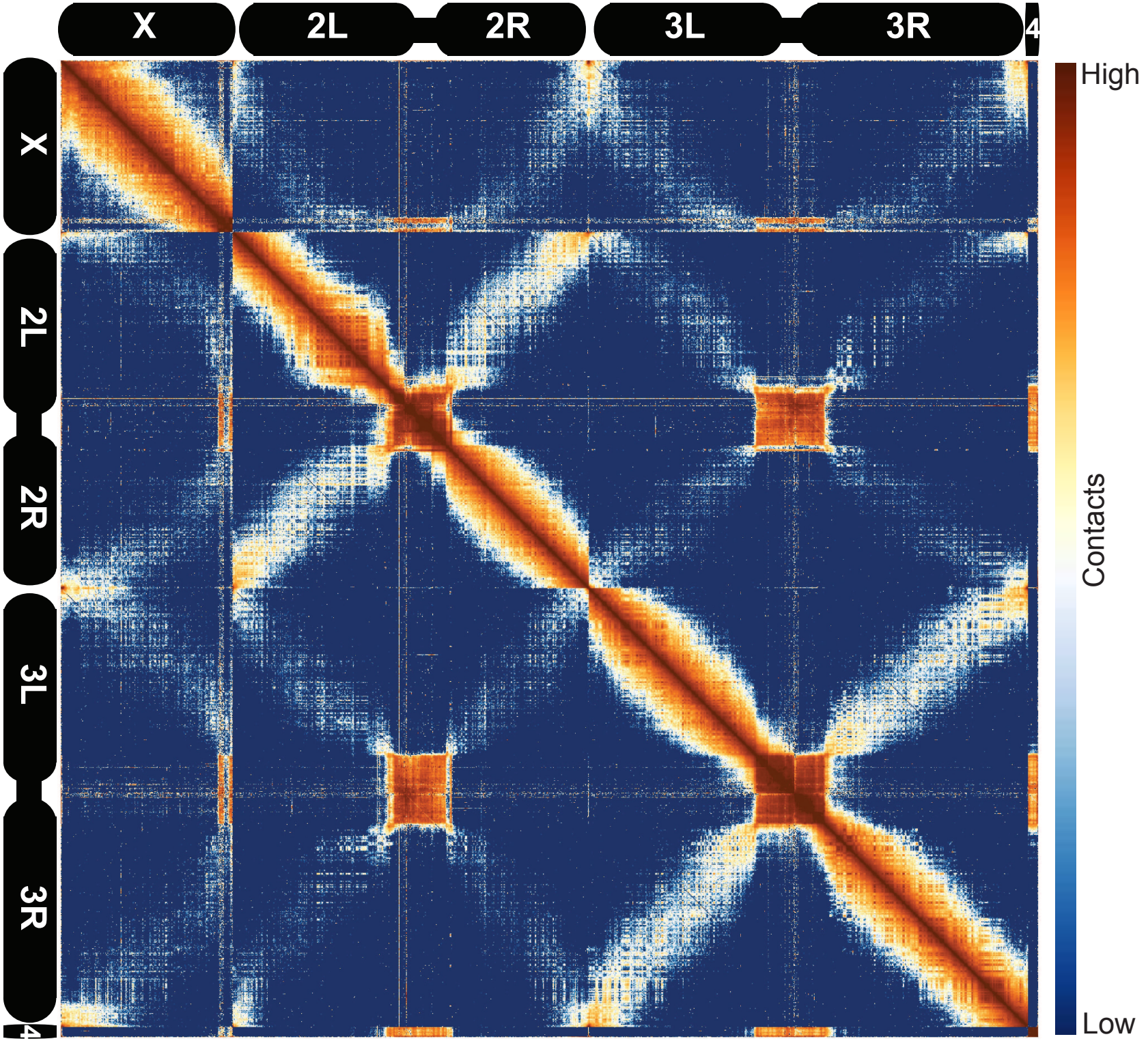
Yang, Jingping, and Victor G. Corces. 2012. “Insulators, Long-Range Interactions, and Genome Function.” *Current Opinion in Genetics & Development* 22 (2): 86–92.

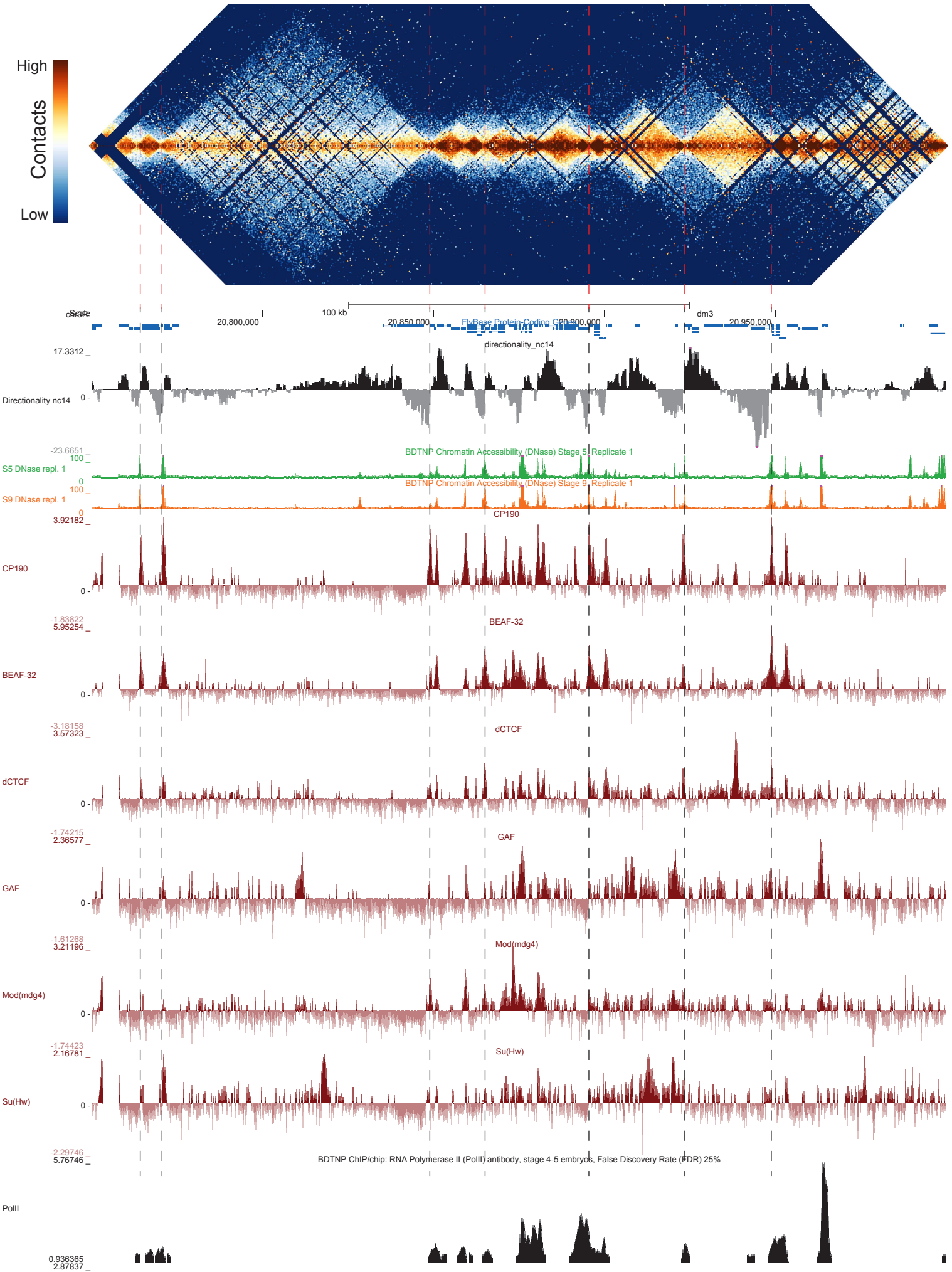
Zhao, Keji, Craig M. Hart, and Ulrich K. Laemmli. 1995. “Visualization of Chromosomal Domains with Boundary Element-Associated Factor BEAF-32.” *Cell* 81 (6): 879–89.

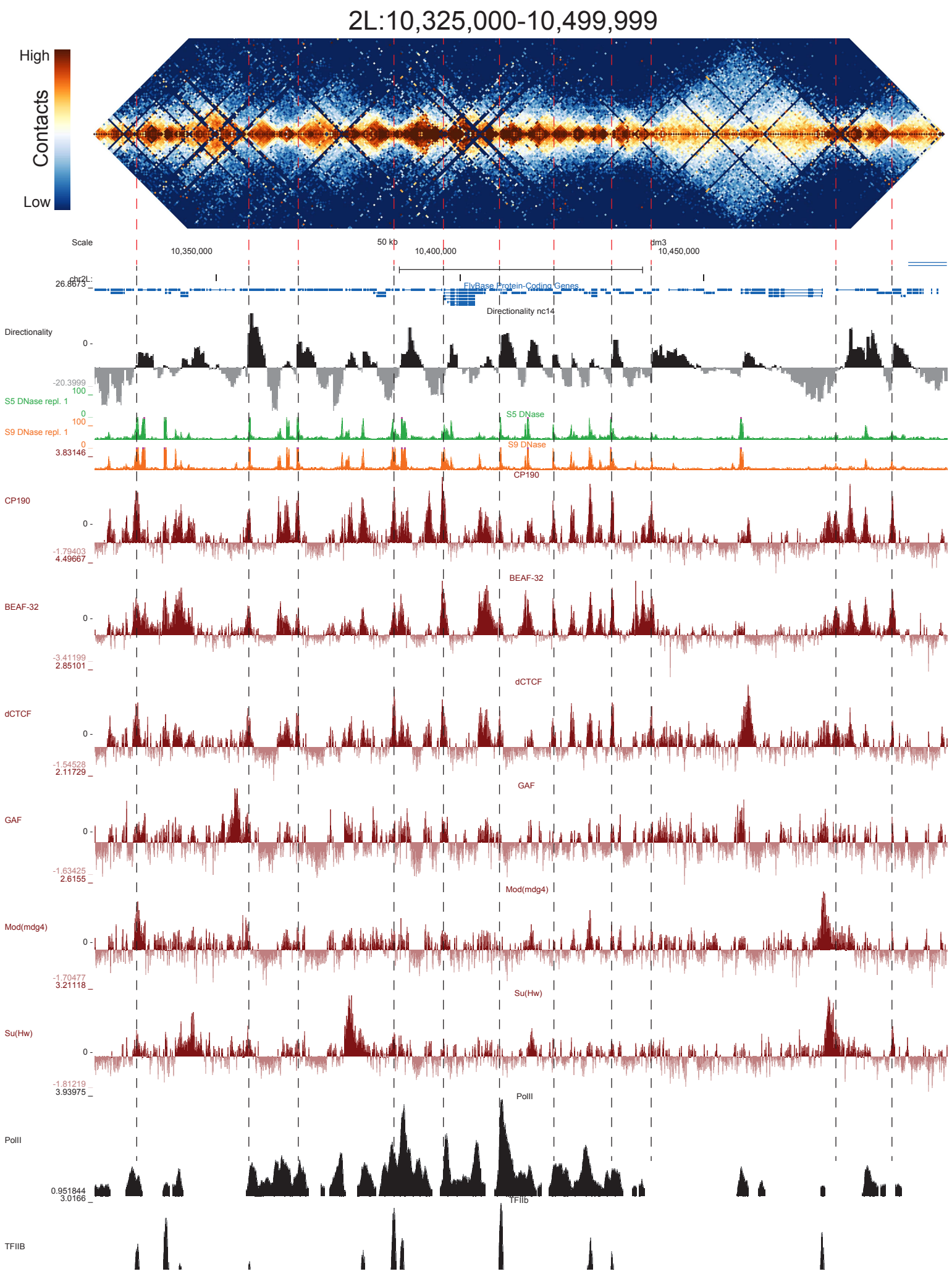
Zhimulev, Igor F. 1996. “Morphology and Structure of Polytene Chromosomes.” *Advances in Genetics* 34 (January): 1–490.

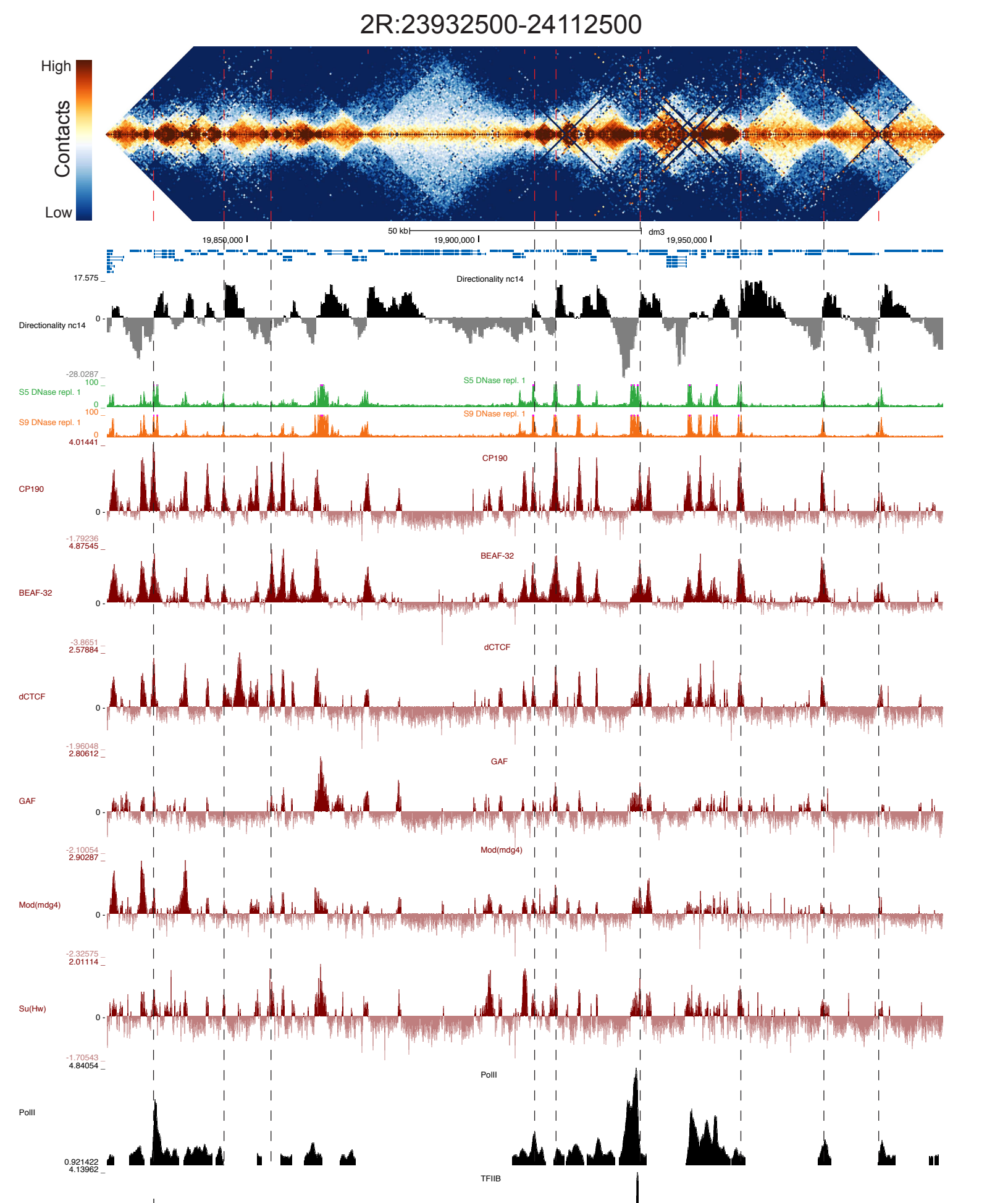
Zhimulev, Igor F., Tatyana Yu Zykova, Fyodor P. Goncharov, Varvara A. Khoroshko, Olga V. Demakova, Valeriy F. Semeshin, Galina V. Pokholkova, et al. 2014. “Genetic Organization of Interphase Chromosome Bands and Interbands in Drosophila Melanogaster.” *PloS One* 9 (7). doi:10.1371/journal.pone.0101631.

Figure 1

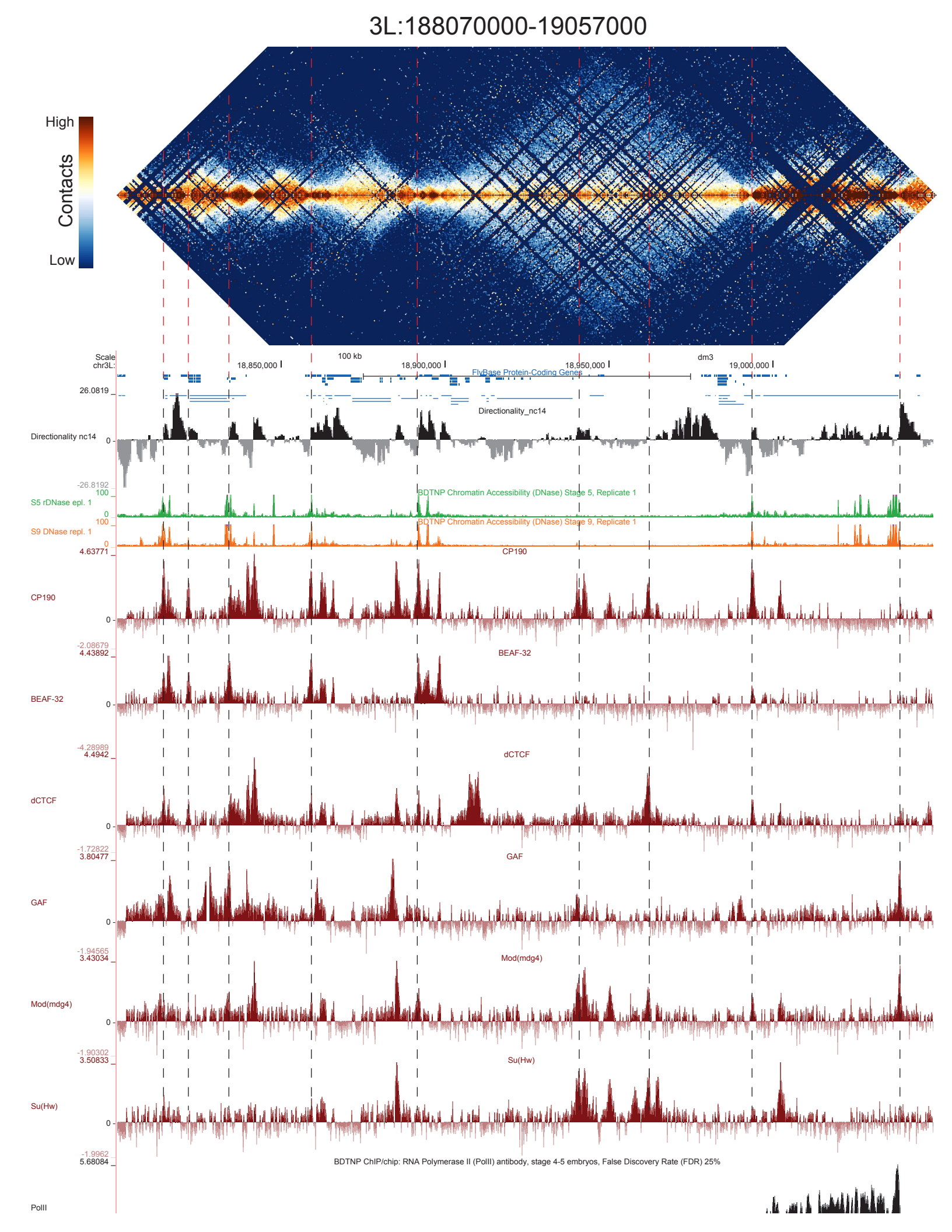


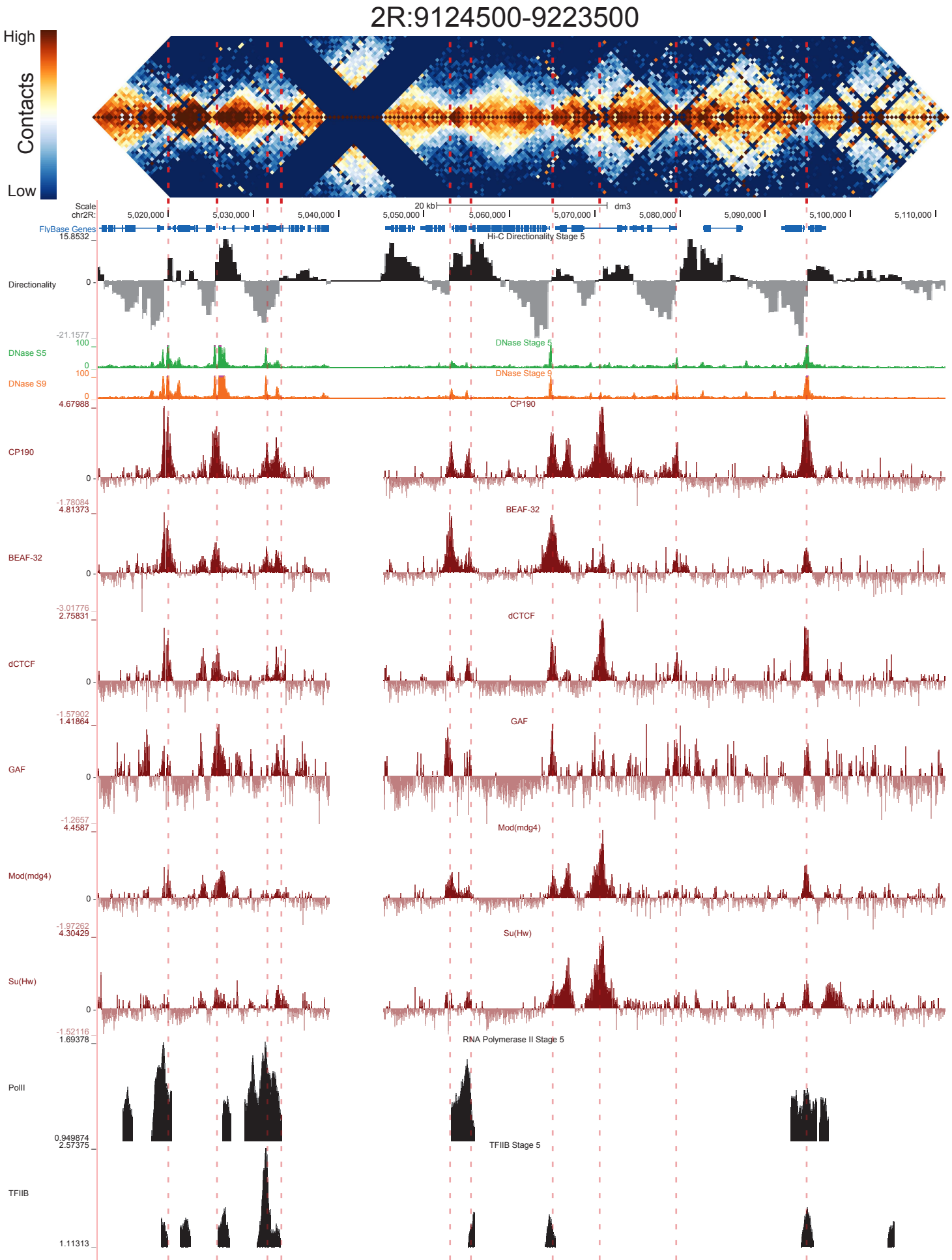

chr3R:24924500-25174500
High
Contacts
Low
chr3R
Scale
100 kb
20,800,000
20,850,000
20,900,000
20,950,000
dm3
FlyBase Protein-Coding Genes
directionality_nc14
17.3312
0
Directionality nc14
-23.6651
100
S5 DNase repl. 1
BDTNP Chromatin Accessibility (DNase) Stage 5, Replicate 1
0
100
S9 DNase repl. 1
BDTNP Chromatin Accessibility (DNase) Stage 9, Replicate 1
0
3.92182
CP190
0
CP190
-1.83822
5.95254
BEAF-32
0
BEAF-32
-3.18158
3.57323
dCTCF
0
dCTCF
-1.74215
2.36577
GAF
0
GAF
-1.61268
3.21196
Mod(mdg4)
0
Mod(mdg4)
-1.74423
2.16781
Su(Hw)
0
Su(Hw)
-2.29746
5.76746
BDTNP ChIP/chip: RNA Polymerase II (PolII) antibody, stage 4-5 embryos, False Discovery Rate (FDR) 25%
PolII
0.936365
2.87837


2L:10,325,000-10,499,999
High
Contacts
Low
Scale
50 kb
10,350,000
10,400,000
10,450,000
dm3
chr2L:
26.8673
FlyBase Protein-Coding Genes
Directionality nc14
Directionality
0
-20.3999
100
S5 DNase repl. 1
S5 DNase
0
100
S9 DNase repl. 1
S9 DNase
0
3.83146
CP190
CP190
0
-1.79403
4.49667
BEAF-32
BEAF-32
0
-3.41199
2.85101
dCTCF
dCTCF
0
-1.54528
2.11729
GAF
GAF
0
-1.63425
2.6155
Mod(mdg4)
Mod(mdg4)
0
-1.70477
3.21118
Su(Hw)
Su(Hw)
0
-1.81219
3.93975
PolII
PolII
0.951844
3.0166
TFIIb
TFIIB


2R:23932500-24112500
High
Contacts
Low
50 kb
dm3
19,850,000
19,900,000
19,950,000
17.575
Directionality nc14
0
Directionality nc14
-28.0287
100
S5 DNase repl. 1
0
S5 DNase repl. 1
100
S9 DNase repl. 1
0
S9 DNase repl. 1
4.01441
CP190
0
CP190
-1.79236
4.87545
BEAF-32
0
BEAF-32
-3.8651
2.57884
dCTCF
0
dCTCF
-1.96048
2.80612
GAF
0
GAF
-2.10054
2.90287
Mod(mdg4)
0
Mod(mdg4)
-2.32575
2.01114
Su(Hw)
0
Su(Hw)
-1.70543
4.84054
PolII
0.921422
4.13962
PolII
TFIIB

3L:18807000-19057000




2R:9124500-9223500
High
Contacts
Low
Scale
chr2R:
5,020,000
5,030,000
5,040,000
20 kb
5,050,000
5,060,000
5,070,000
dm3
5,080,000
5,090,000
5,100,000
5,110,000
FlyBase Genes
15.8532
Hi-C Directionality Stage 5
Directionality
0
-21.1577
DNase Stage 5
DNase S5
100
0
DNase Stage 9
DNase S9
100
0
4.67988
CP190
CP190
0
-1.78084
4.81373
BEAF-32
BEAF-32
0
-3.01776
2.75831
dCTCF
dCTCF
0
-1.57902
1.41864
GAF
GAF
0
-1.2657
4.4587
Mod(mdg4)
Mod(mdg4)
0
-1.97262
4.30429
Su(Hw)
Su(Hw)
0
-1.52116
1.69378
RNA Polymerase II Stage 5
PolII
0.949874
2.57375
TFIIB Stage 5
TFIIB
1.11313

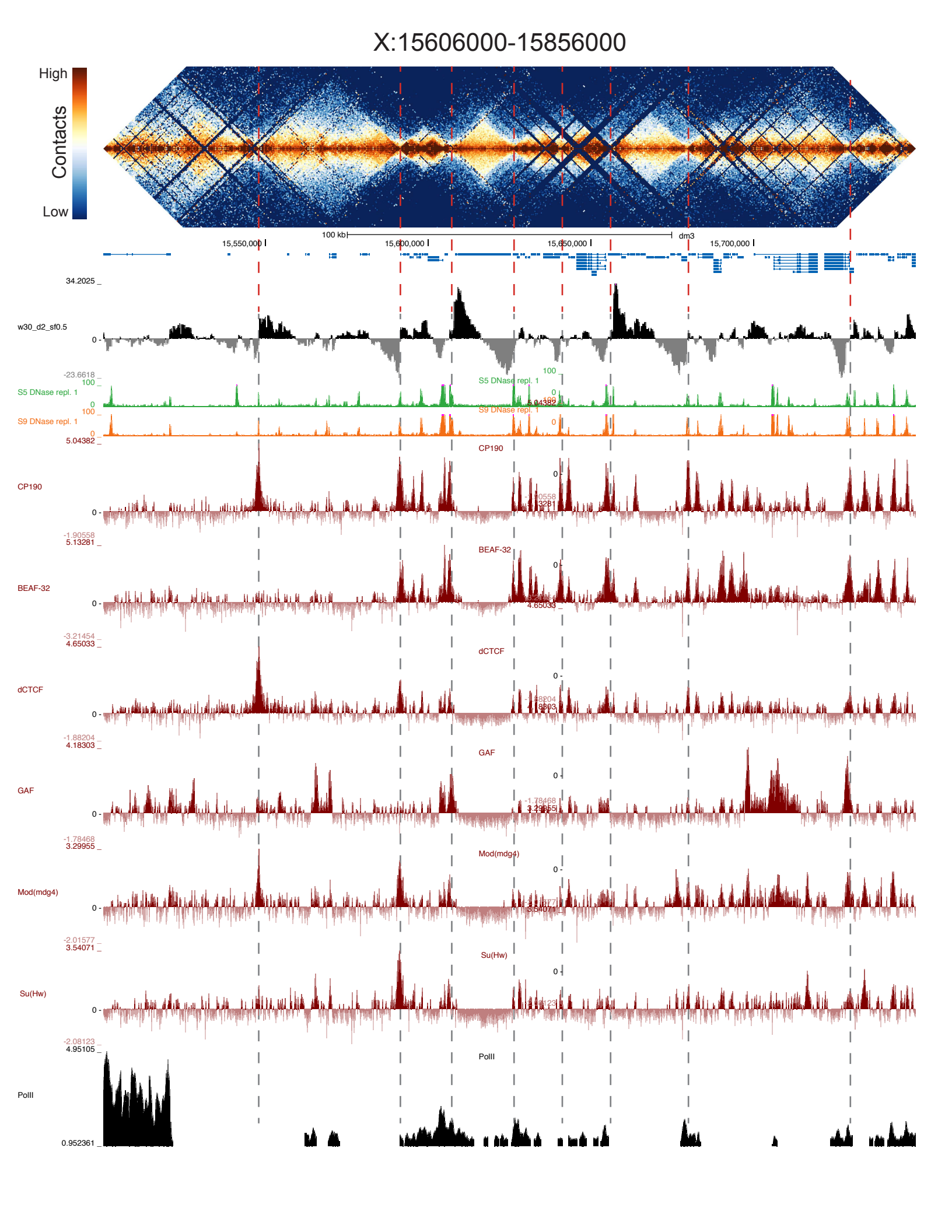

X:15606000-15856000
High
Contacts
Low
100 kb
dm3
15,550,000
15,600,000
15,650,000
15,700,000
34.2025
w30_d2_sf0.5
0
-23.6618
S5 DNase repl. 1
100
0
S9 DNase repl. 1
100
0
5.04382
CP190
0
-1.90558
5.13281
BEAF-32
0
-3.21454
4.65033
dCTCF
0
-1.88204
4.18303
GAF
0
-1.78468
3.29955
Mod(mdg4)
0
-2.01577
3.54071
Su(Hw)
0
-2.08123
4.95105
PolII
0.952361

Figure 3

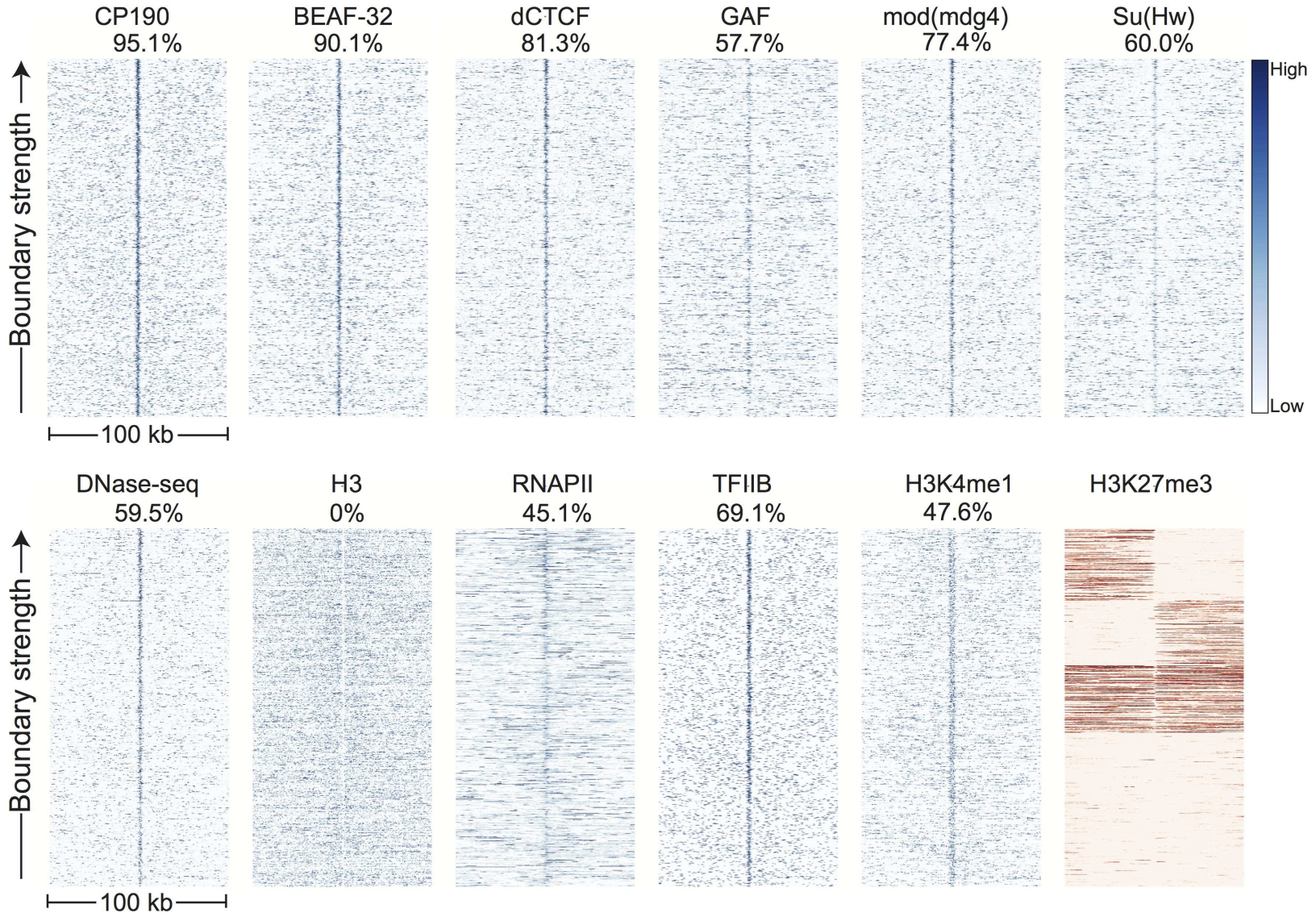
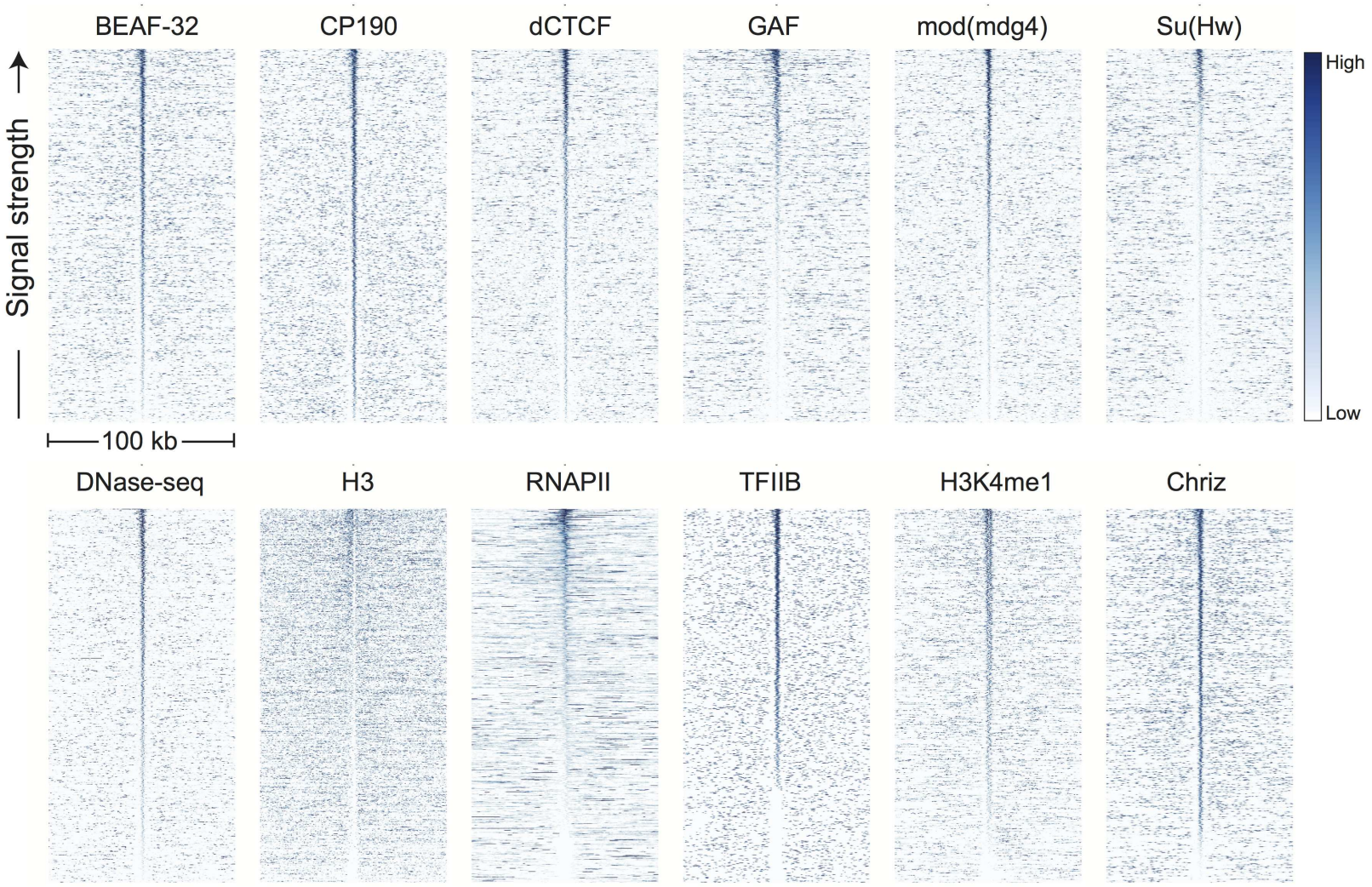
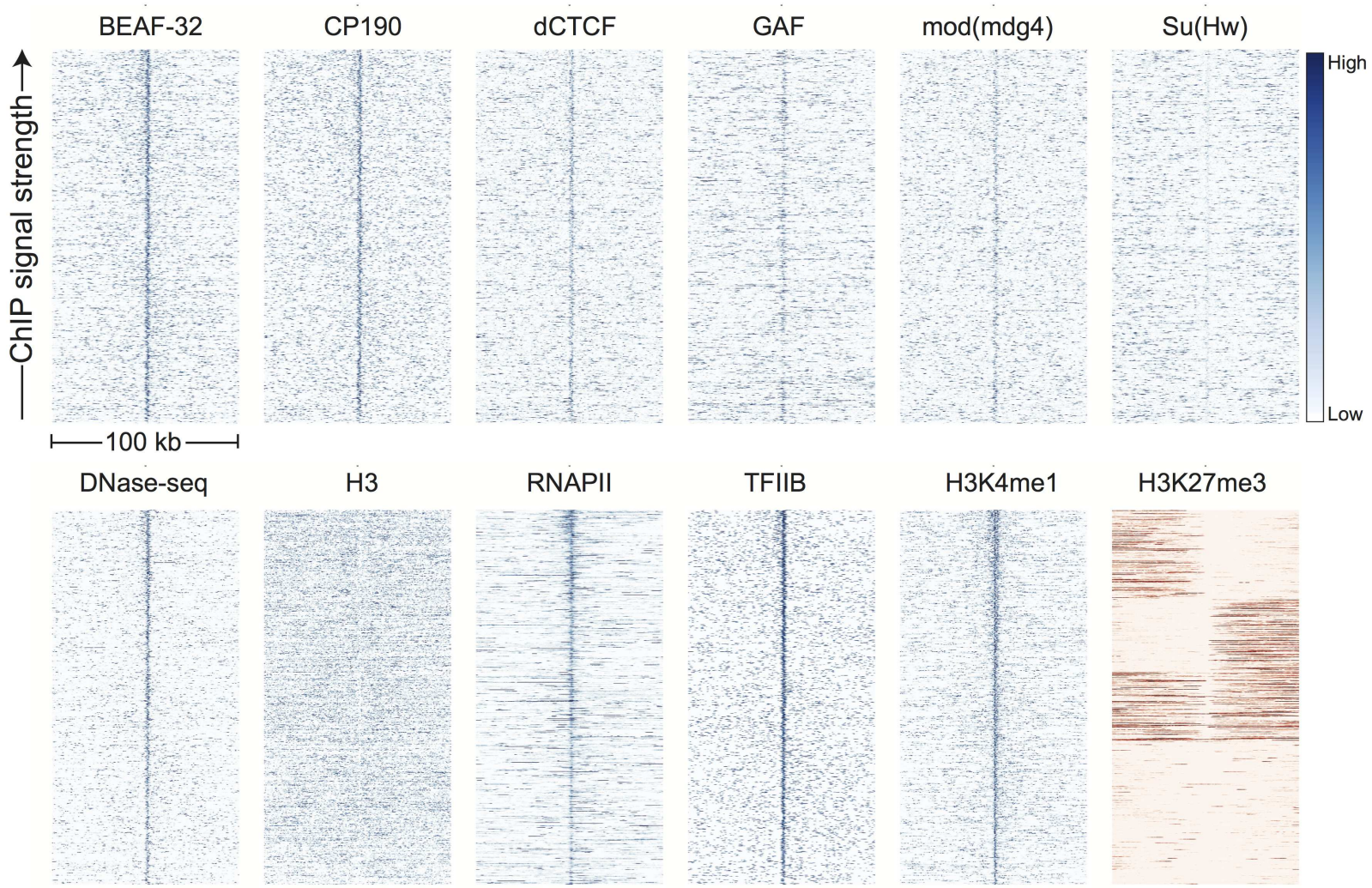
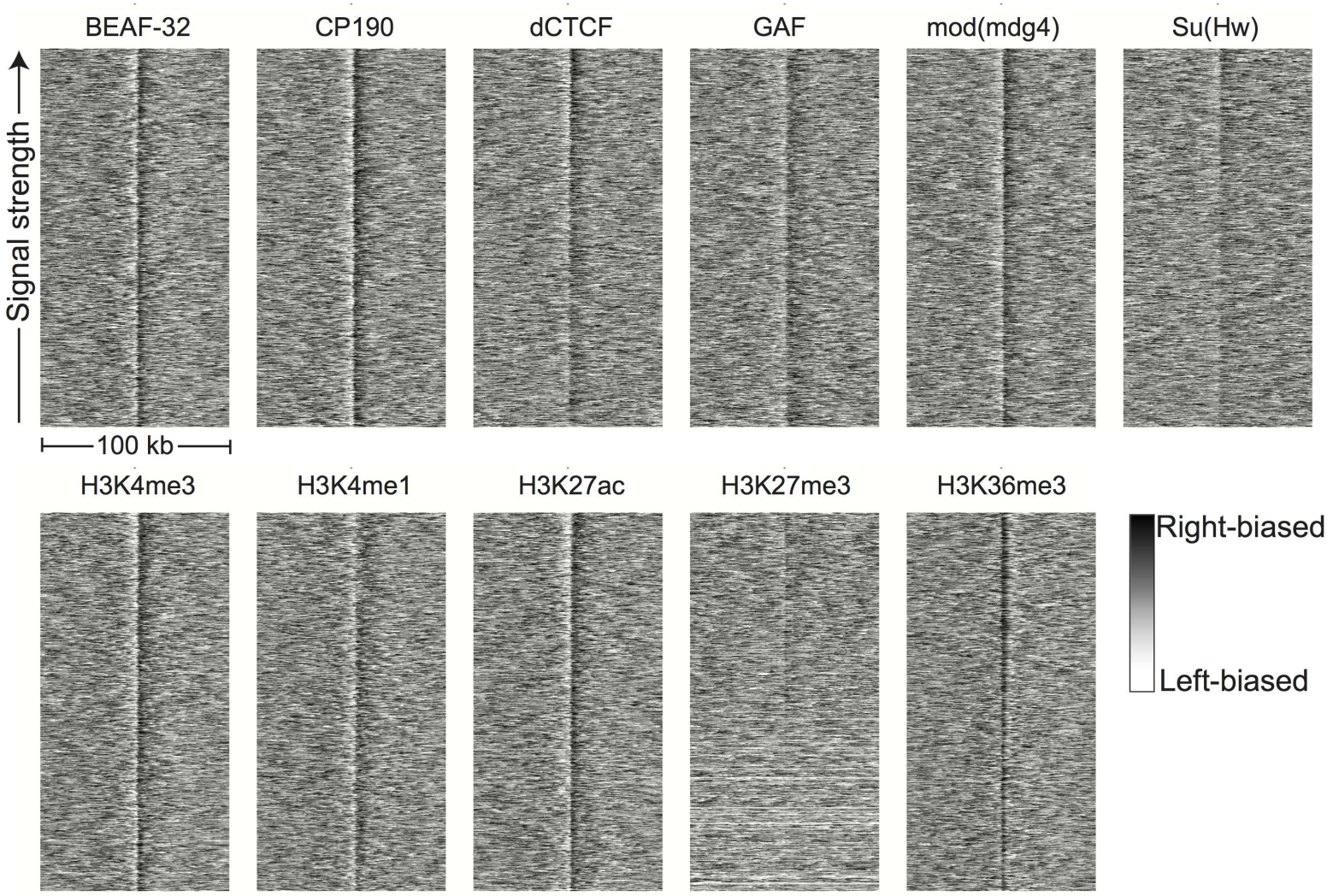
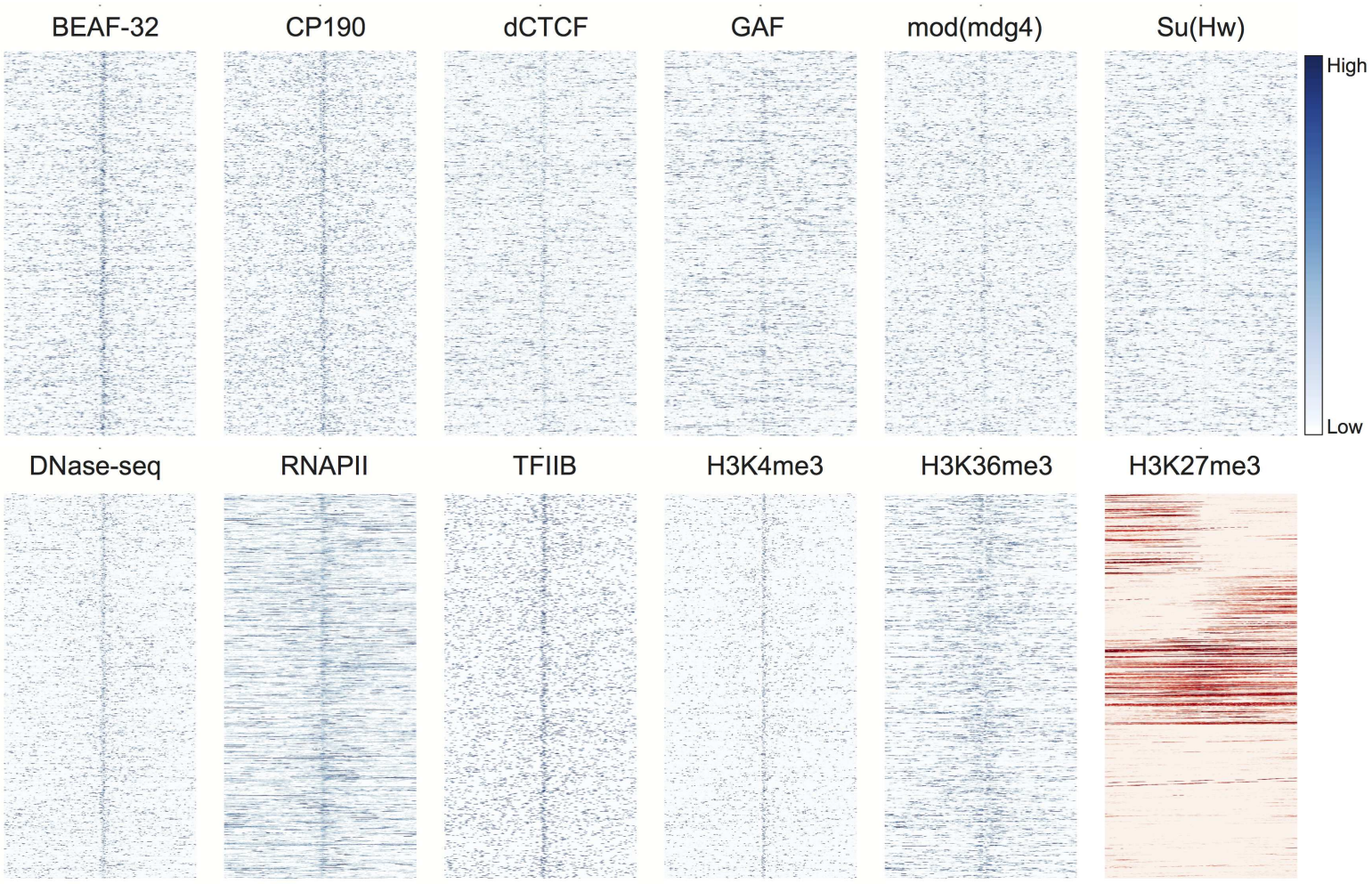

CP190
95.1%
BEAF-32
90.1%
dCTCF
81.3%
GAF
57.7%
mod(mdg4)
77.4%
Su(Hw)
60.0%
High
Low
Boundary strength
100 kb
DNase-seq
59.5%
H3
0%
RNAPII
45.1%
TFIIB
69.1%
H3K4me1
47.6%
H3K27me3
Boundary strength
100 kb

Figure 3-supplement 1




BEAF-32
CP190
dCTCF
GAF
mod(mdg4)
Su(Hw)
ChIP signal strength
High
Low
100 kb
DNase-seq
H3
RNAPII
TFIIB
H3K4me1
H3K27me3


BEAF-32
CP190
dCTCF
GAF
mod(mdg4)
Su(Hw)
Signal strength
100 kb
H3K4me3
H3K4me1
H3K27ac
H3K27me3
H3K36me3
Right-biased
Left-biased


BEAF-32
CP190
dCTCF
GAF
mod(mdg4)
Su(Hw)
High
Low
DNase-seq
RNAPII
TFIIB
H3K4me3
H3K36me3
H3K27me3

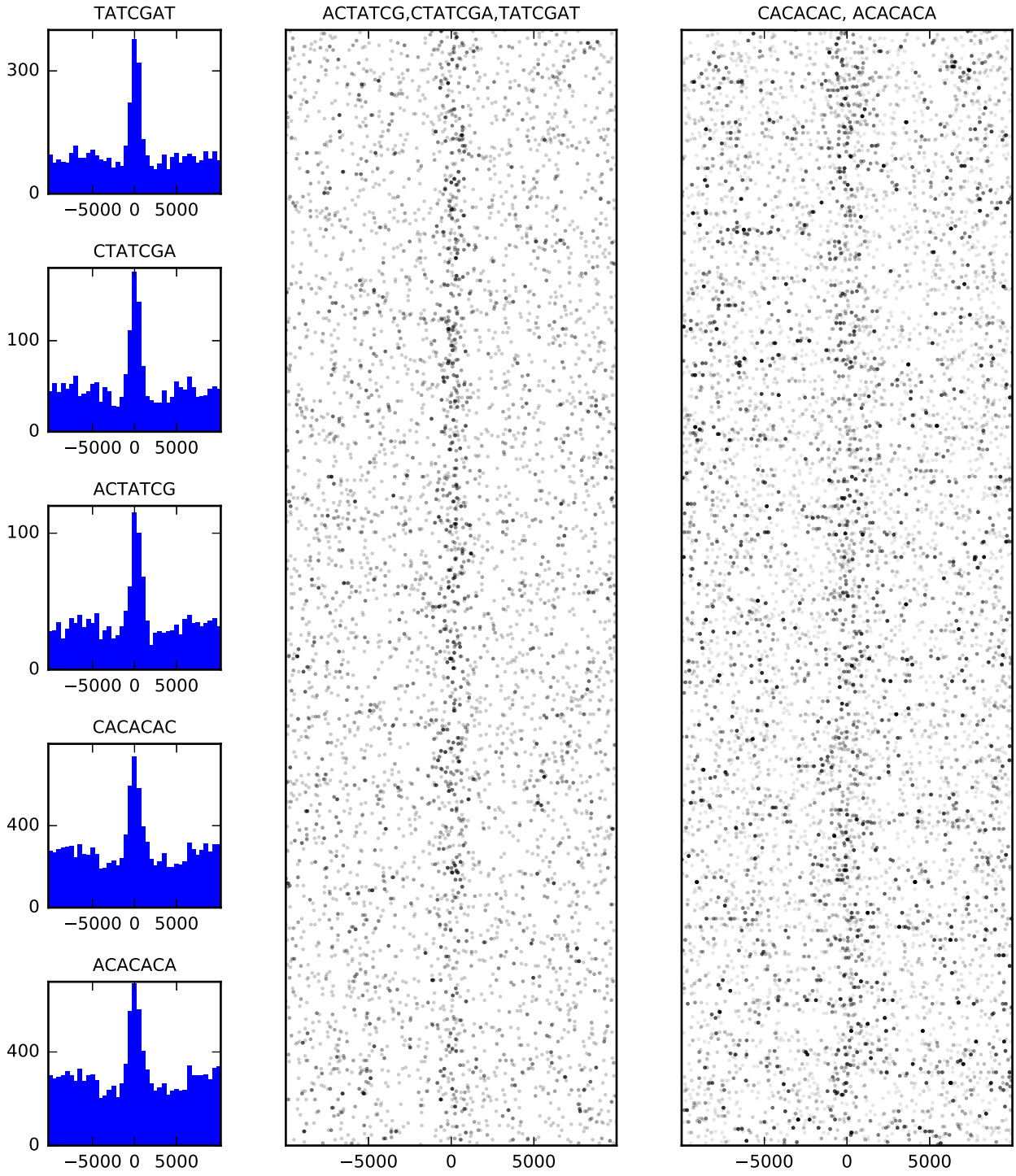
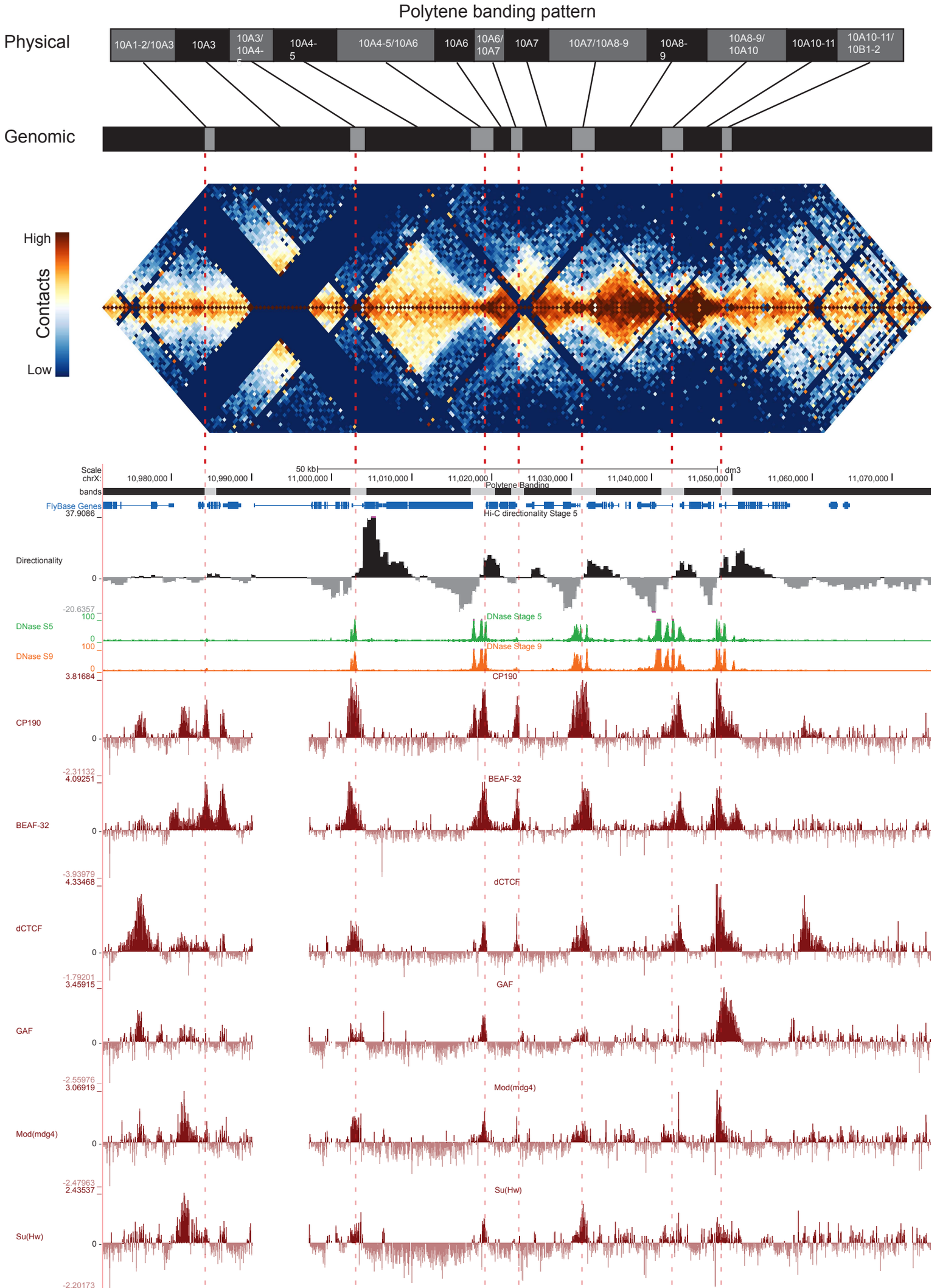
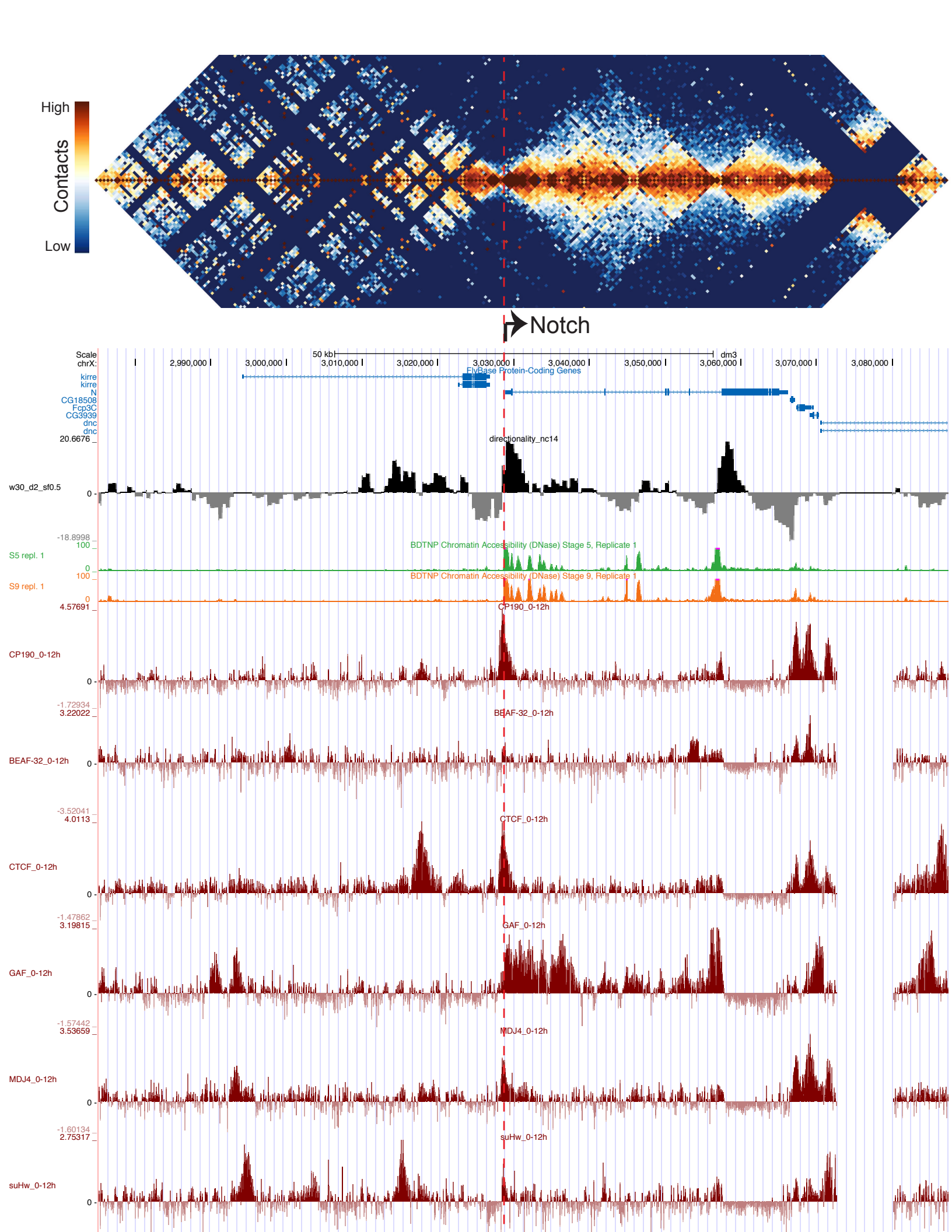
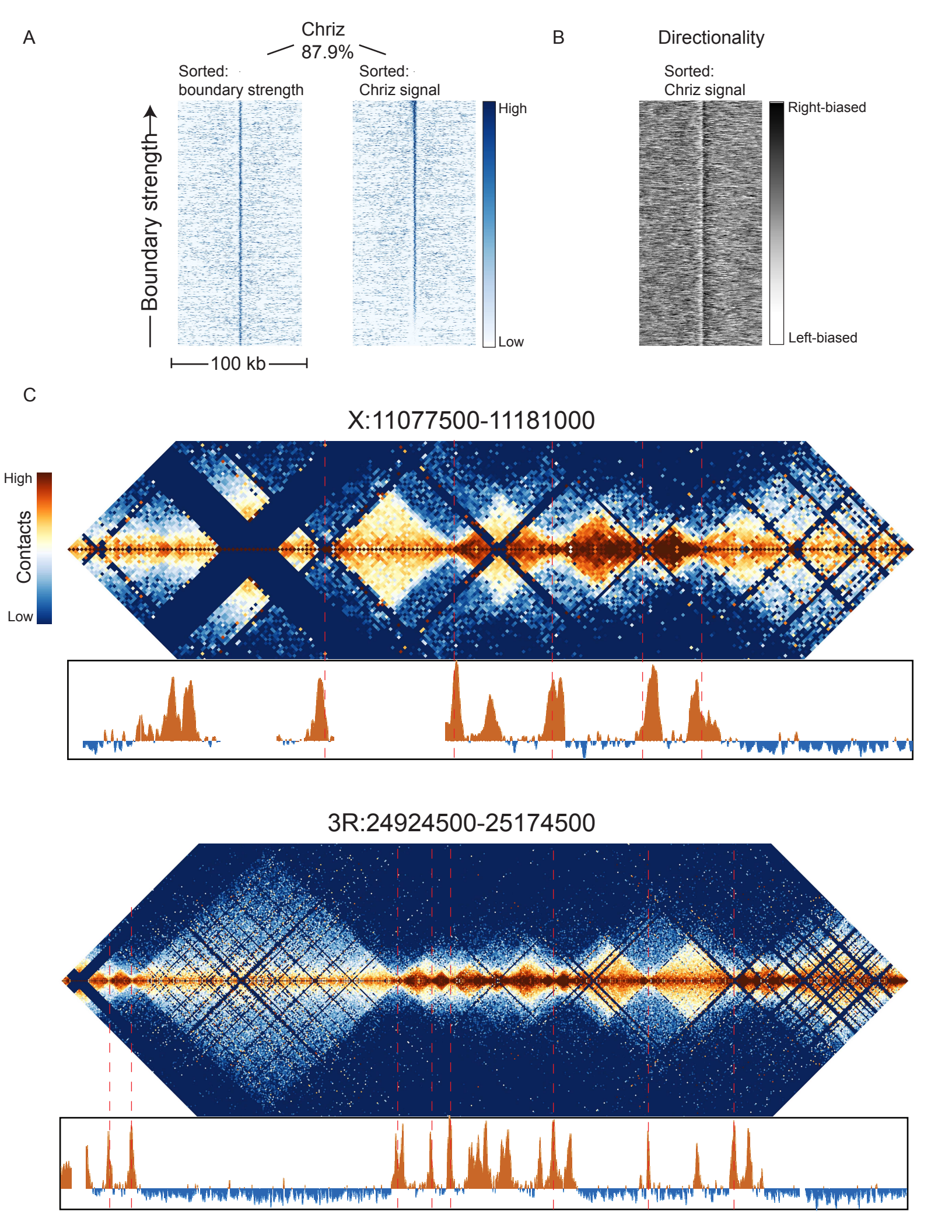

TATCGAT
300
0
−5000 0 5000
CTATCGA
100
0
−5000 0 5000
ACTATCG
100
0
−5000 0 5000
CACACAC
400
0
−5000 0 5000
ACACACA
400
0
−5000 0 5000
ACTATCG,CTATCGA,TATCGAT
−5000 0 5000
CACACAC, ACACACA
−5000 0 5000

Figure 5


Polytene banding pattern
Physical
10A1-2/10A3
10A3
10A3/ 10A4-5
10A4-5
10A4-5/10A6
10A6
10A6/ 10A7
10A7
10A7/10A8-9
10A8-9
10A8-9/ 10A10
10A10-11
10A10-11/ 10B1-2
Genomic
High
Contacts
Low
Scale
chrX:
10,980,000
10,990,000
50 kb
11,000,000
11,010,000
11,020,000
11,030,000
11,040,000
11,050,000
dm3
11,060,000
11,070,000
bands
Polytene Banding
FlyBase Genes
37.9086
Hi-C directionality Stage 5
Directionality
0
-20.6357
DNase S5
100
0
DNase Stage 5
DNase S9
100
0
DNase Stage 9
3.81684
CP190
CP190
0
-2.31132
4.09251
BEAF-32
BEAF-32
0
-3.93979
4.33468
dCTCF
dCTCF
0
-1.79201
3.45915
GAF
GAF
0
-2.55976
3.06919
Mod(mdg4)
Mod(mdg4)
0
-2.47963
2.43537
Su(Hw)
Su(Hw)
0
-2.20173


High
Contacts
Low
Notch
Scale
chrX:
50 kb
dm3
2,990,000
3,000,000
3,010,000
3,020,000
3,030,000
3,040,000
3,050,000
3,060,000
3,070,000
3,080,000
FlyBase Protein-Coding Genes
kirre
kirre
N
CG18508
Fcp3C
CG3939
dnc
dnc
20.6676
directionality_nc14
w30_d2_sf0.5
0
-18.8998
100
S5 repl. 1
BDTNP Chromatin Accessibility (DNase) Stage 5, Replicate 1
0
100
S9 repl. 1
BDTNP Chromatin Accessibility (DNase) Stage 9, Replicate 1
0
4.57691
CP190_0-12h
CP190_0-12h
0
-1.72934
3.22022
BEAF-32_0-12h
BEAF-32_0-12h
0
-3.52041
4.0113
CTCF_0-12h
CTCF_0-12h
0
-1.47862
3.19815
GAF_0-12h
GAF_0-12h
0
-1.57442
3.53659
MDJ4_0-12h
MDJ4_0-12h
0
-1.60134
2.75317
suHw_0-12h
suHw_0-12h
0


A
Chriz
87.9%
Sorted:
boundary strength
Sorted:
Chriz signal
Boundary strength
High
Low
100 kb
B
Directionality
Sorted:
Chriz signal
Right-biased
Left-biased
C
X:11077500-11181000
High
Contacts
Low
3R:24924500-25174500

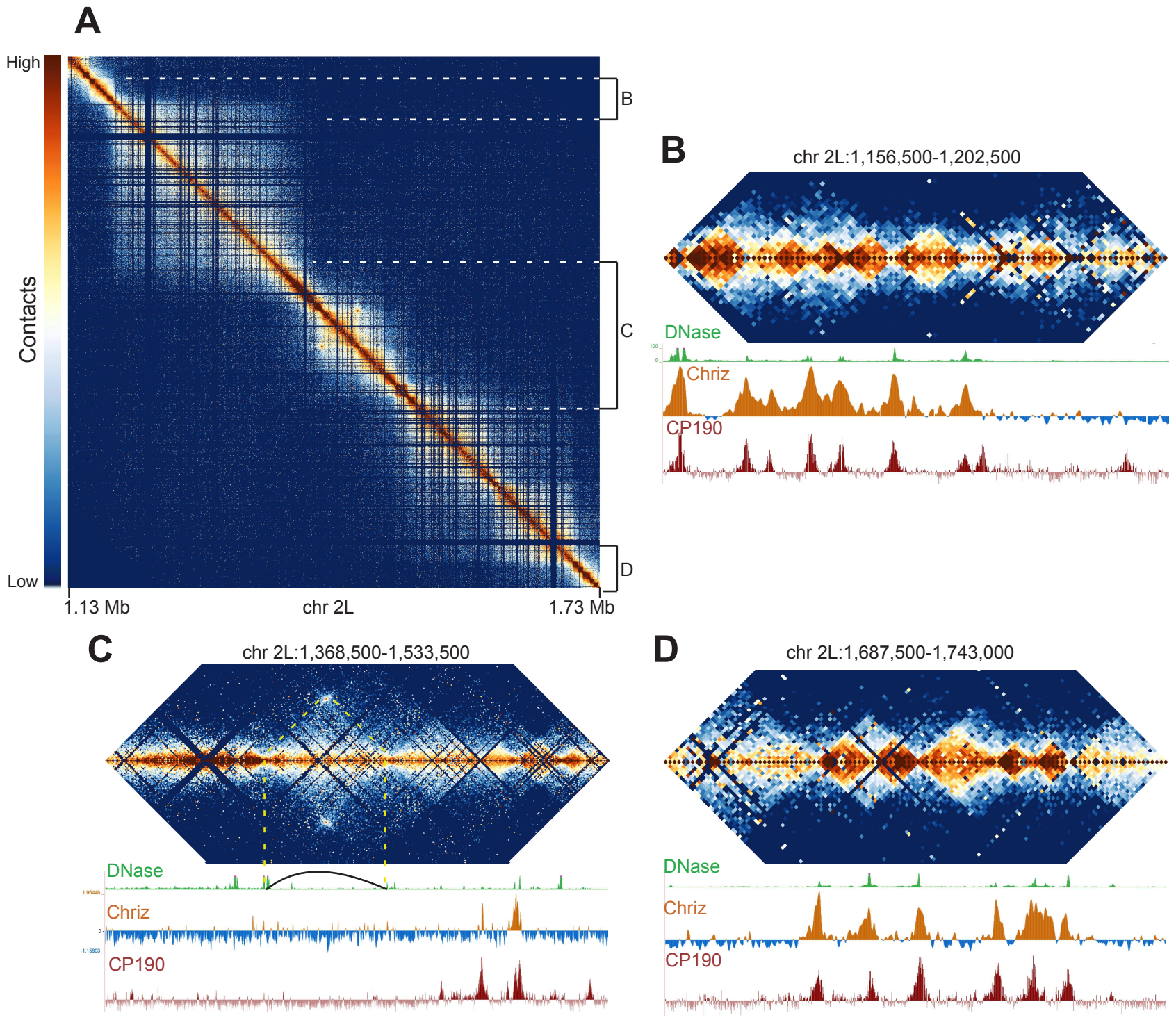
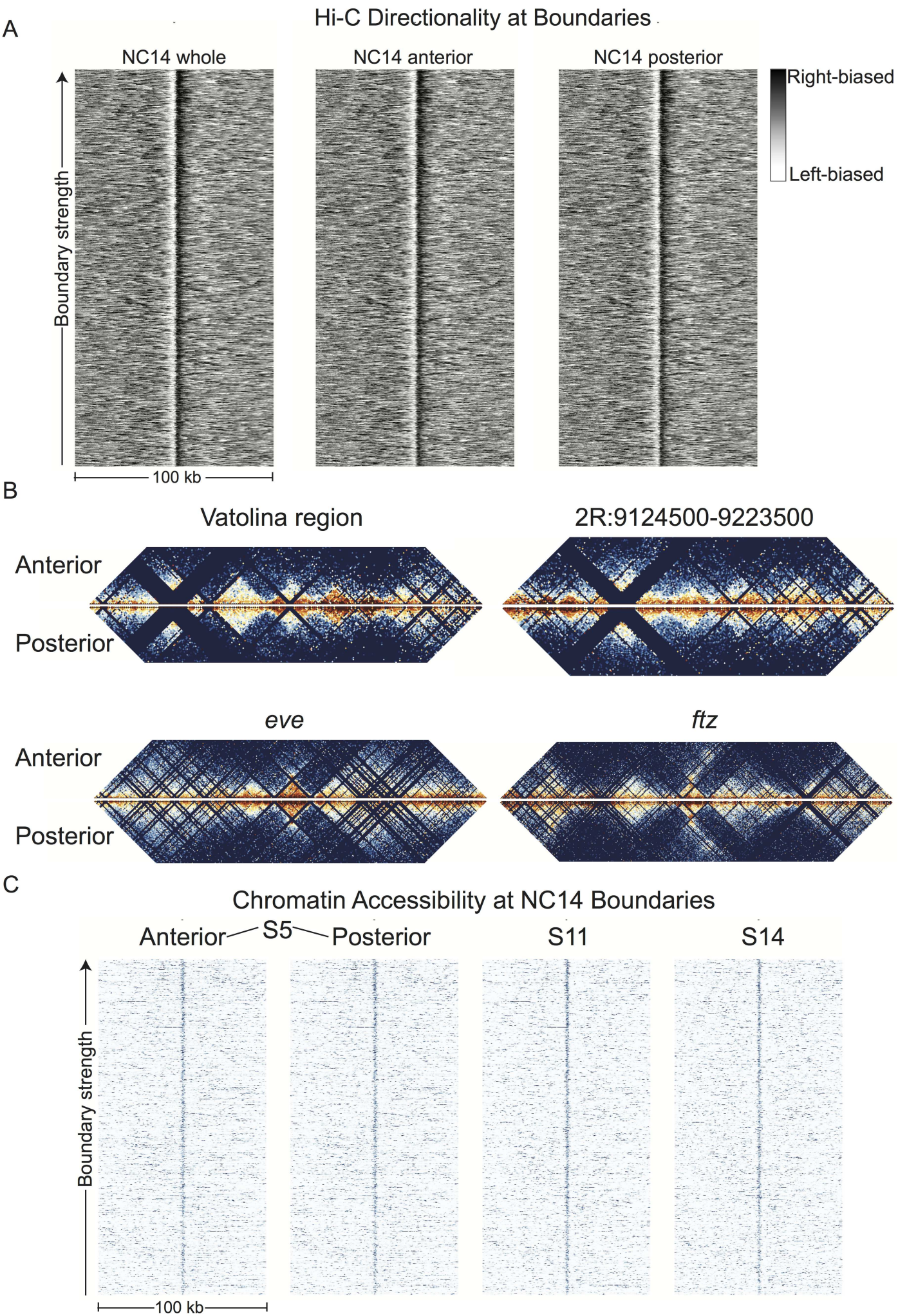

A
High
Low
Contacts
B
C
D
1.13 Mb
chr 2L
1.73 Mb
B
chr 2L:1,156,500-1,202,500
DNase
Chriz
CP190
C
chr 2L:1,368,500-1,533,500
DNase
Chriz
CP190
D
chr 2L:1,687,500-1,743,000
DNase
Chriz
CP190

Figure 6


A
Hi-C Directionality at Boundaries
NC14 whole
NC14 anterior
NC14 posterior
Right-biased
Left-biased
Boundary strength
100 kb
B
Vatolina region
2R:9124500-9223500
Anterior
Posterior
eve
ftz
Anterior
Posterior
C
Chromatin Accessibility at NC14 Boundaries
Anterior
S5
Posterior
S11
S14
Boundary strength
100 kb

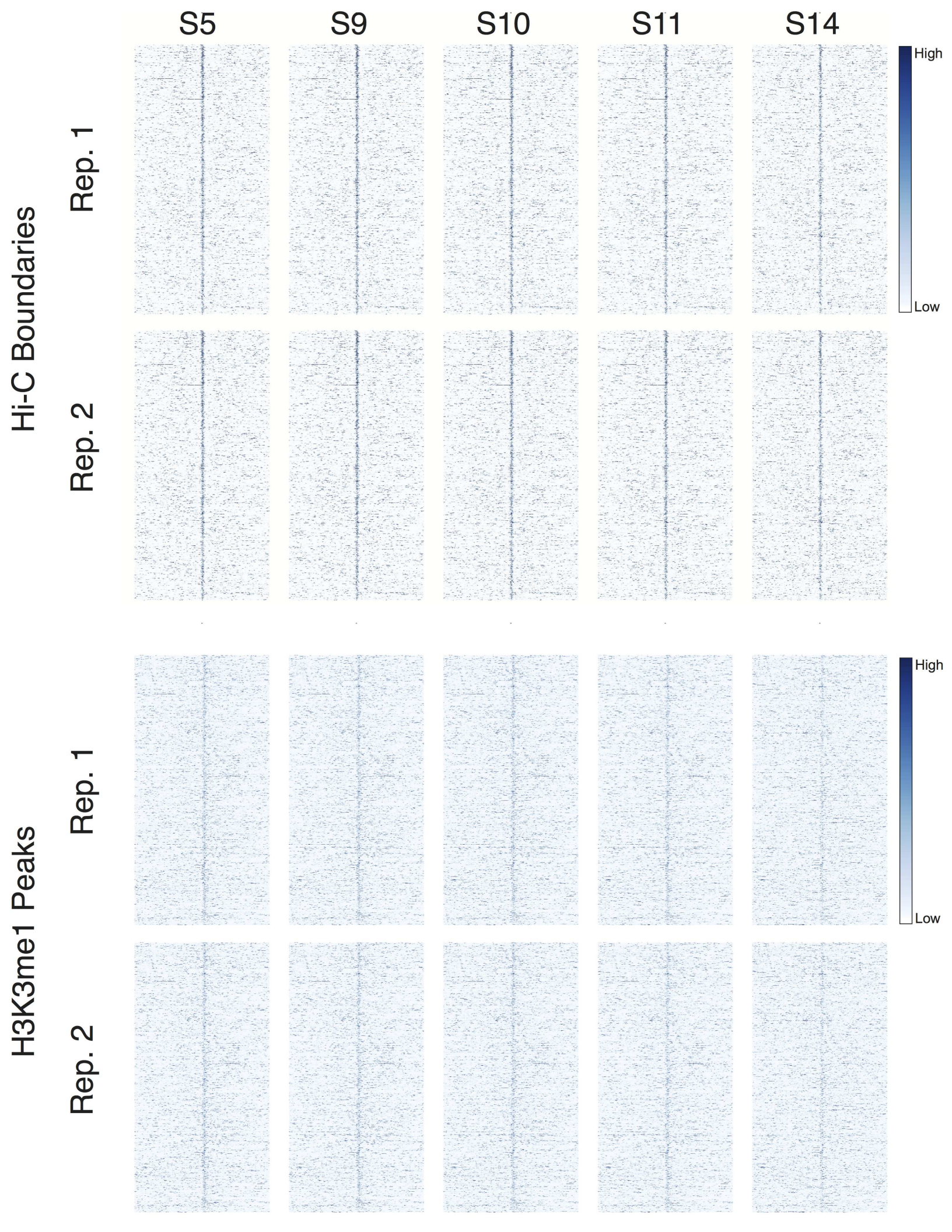
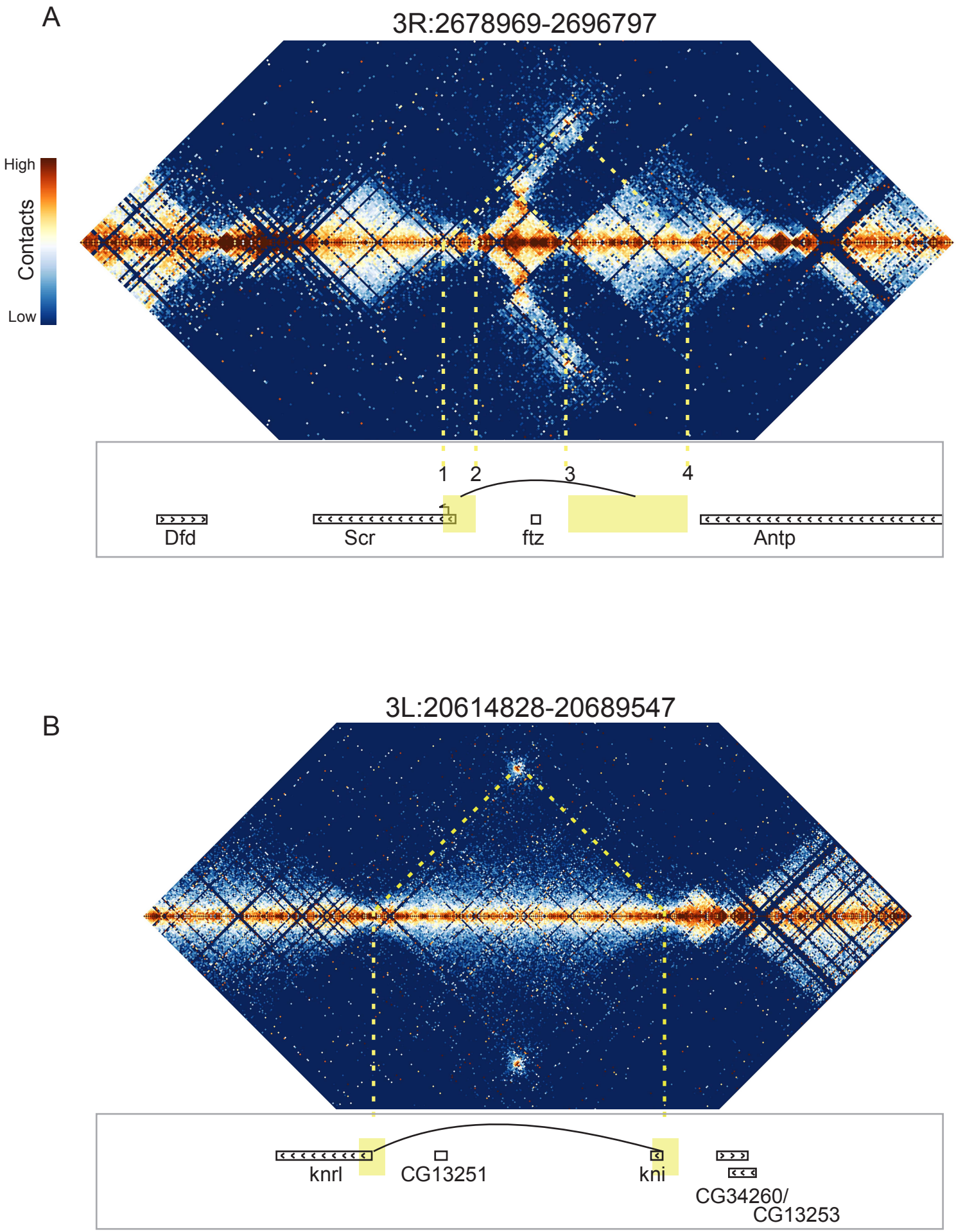
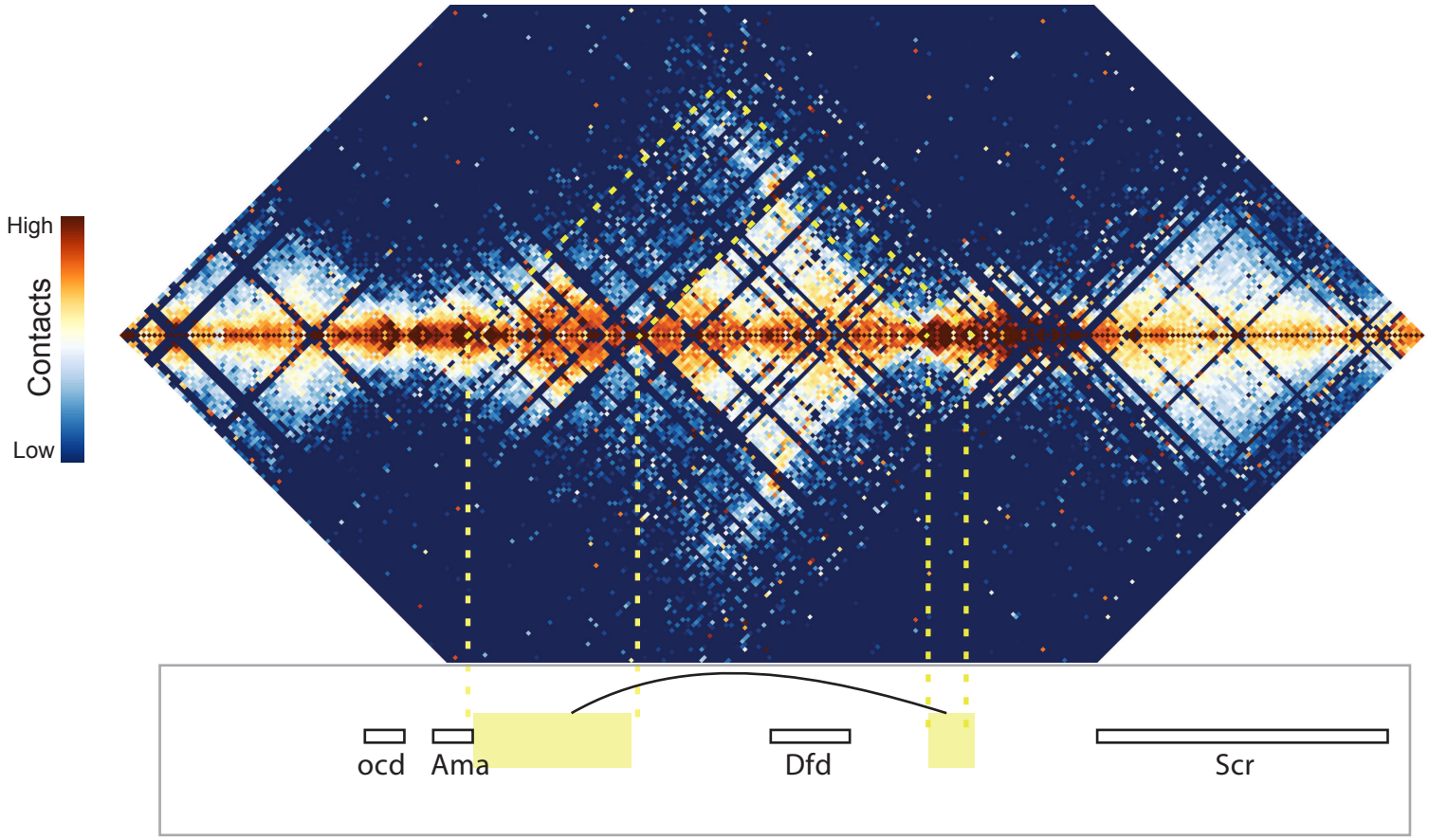
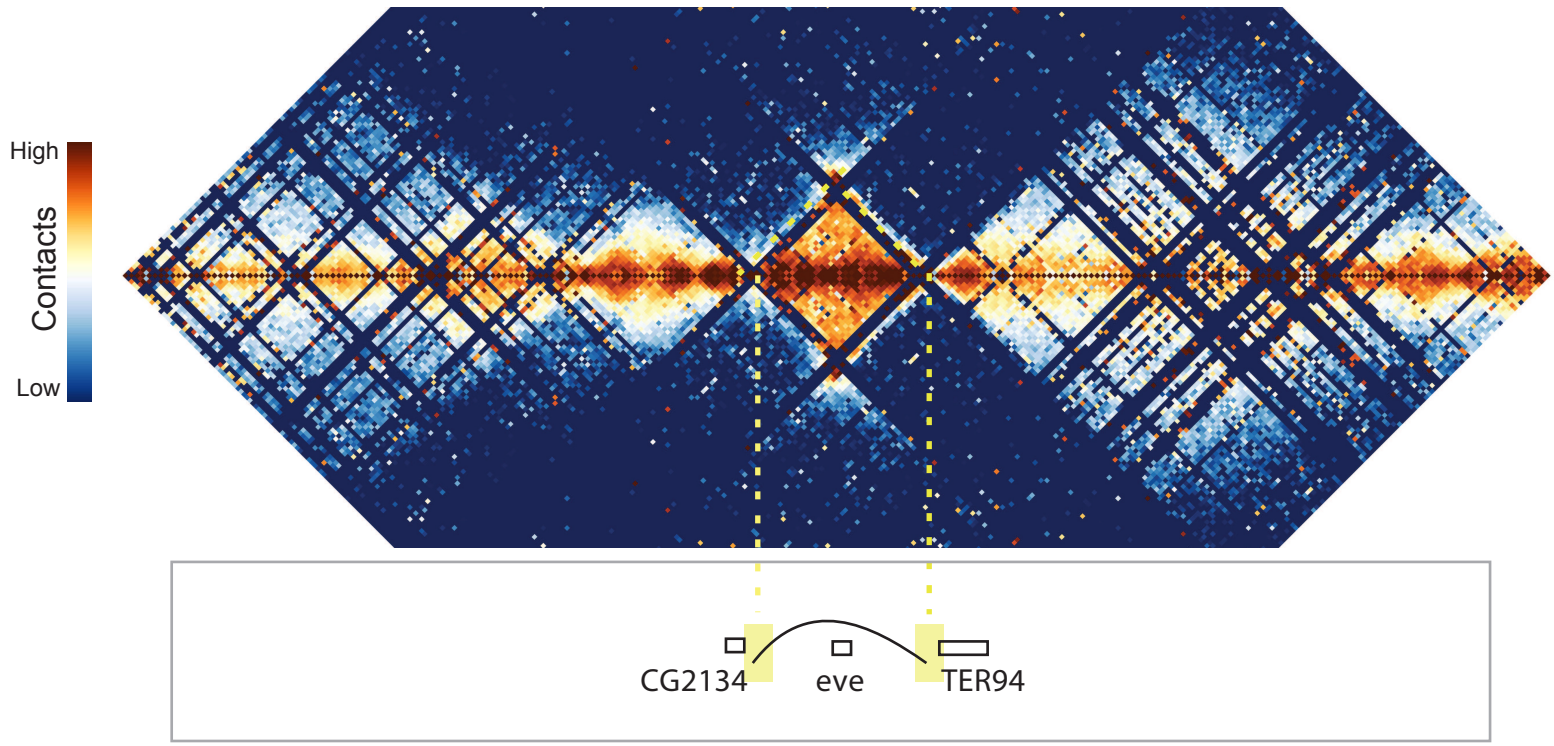

S5
S9
S10
S11
S14
Hi-C Boundaries
Rep. 1
Rep. 2
High
Low
H3K3me1 Peaks
Rep. 1
Rep. 2
High
Low

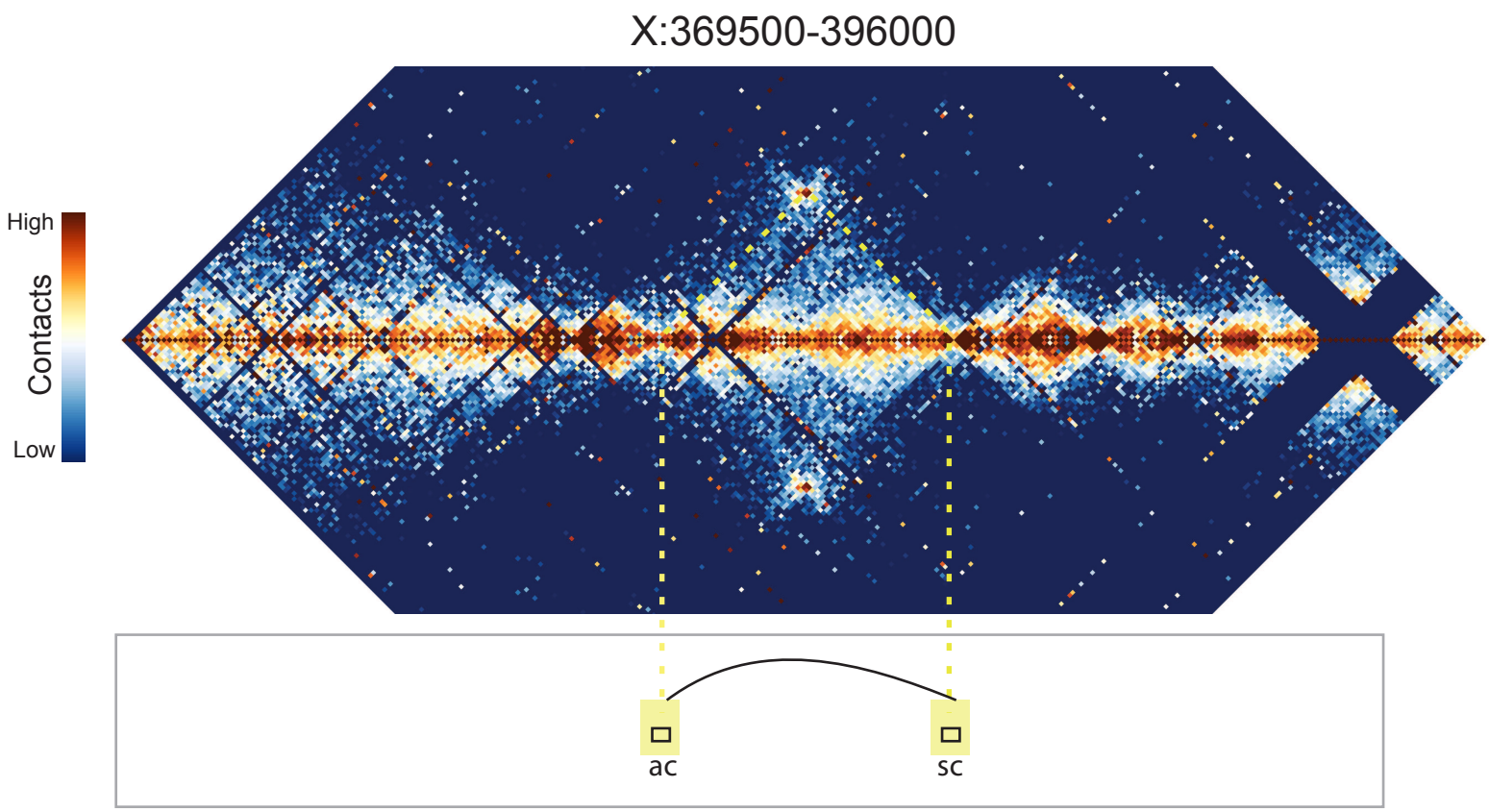
Figure 8

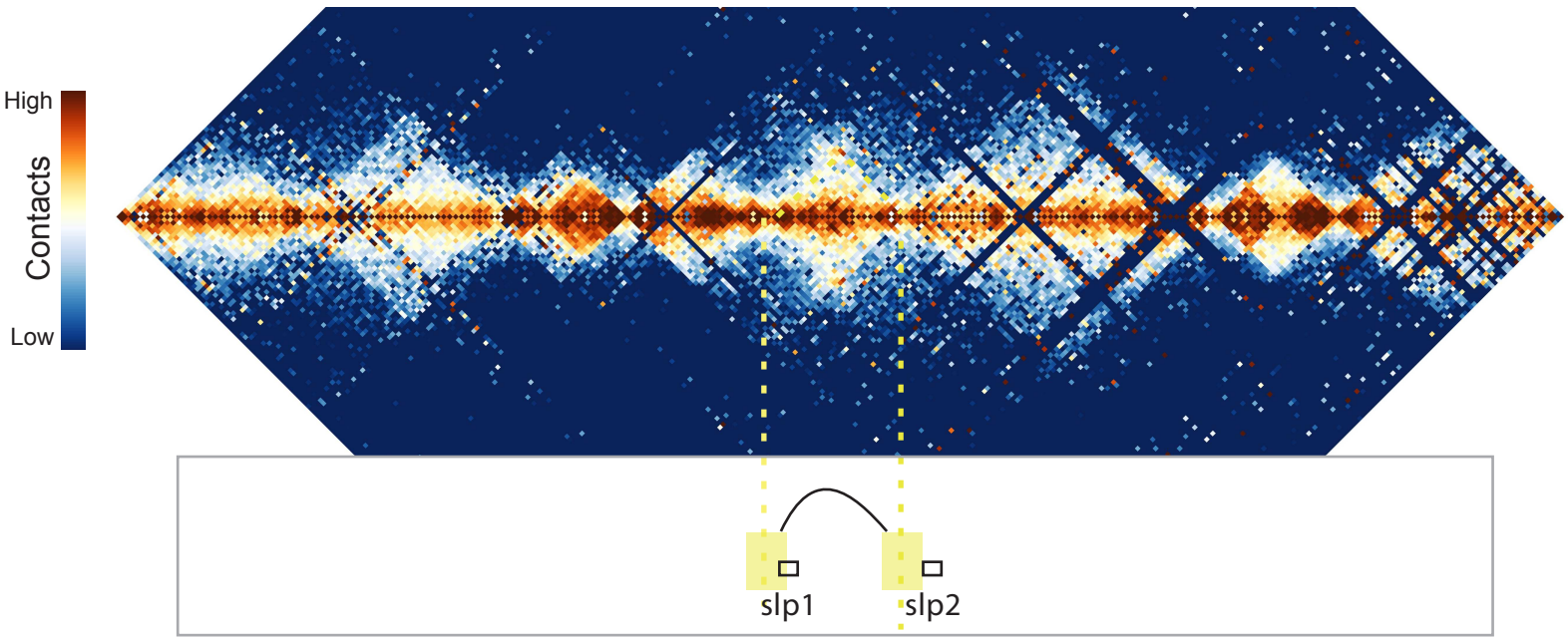

A
3R:2678969-2696797
High
Contacts
Low
1
2
3
4
Dfd
Scr
ftz
Antp
B
3L:20614828-20689547
knrl
CG13251
kni
CG34260/
CG13253


3R:6780000-6807000
High
Contacts
Low
ocd
Ama
Dfd
Scr

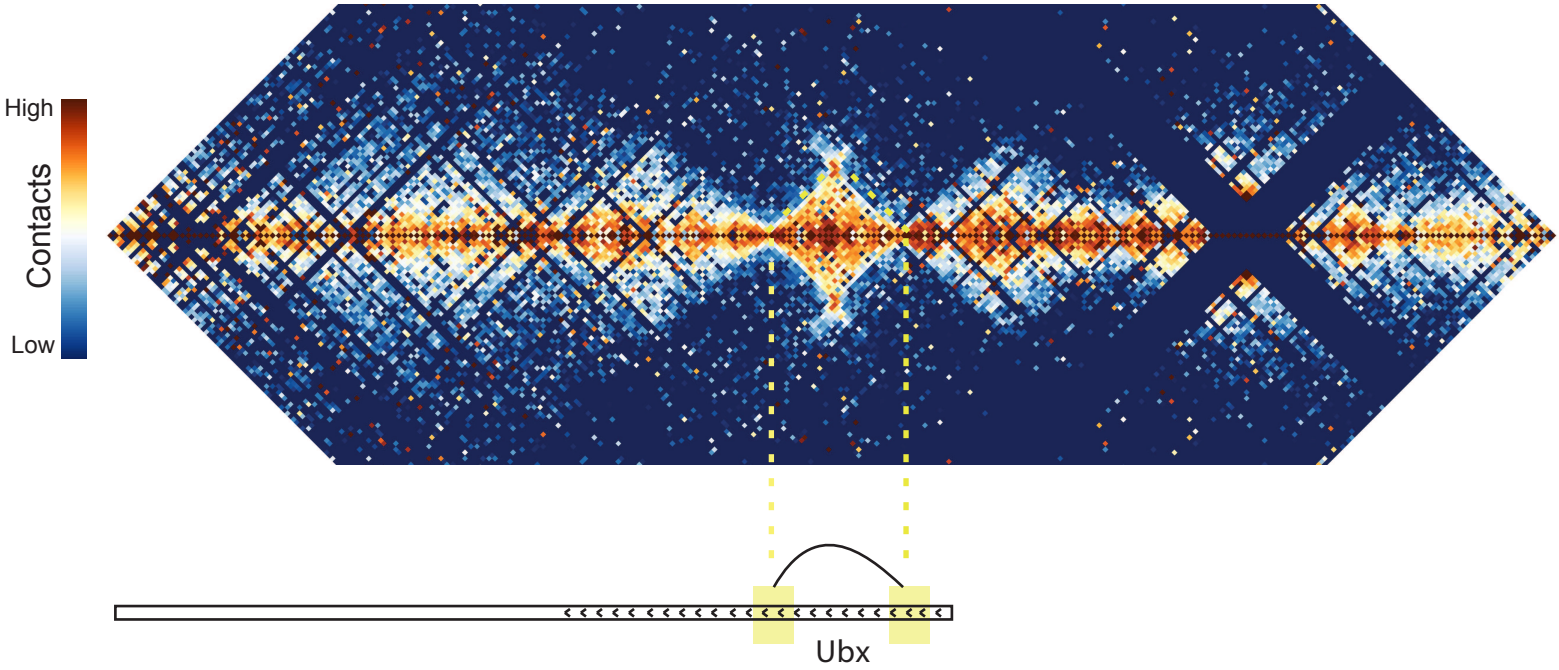

2R:9973000-9988500
High
Contacts
Low
CG2134
eve
TER94

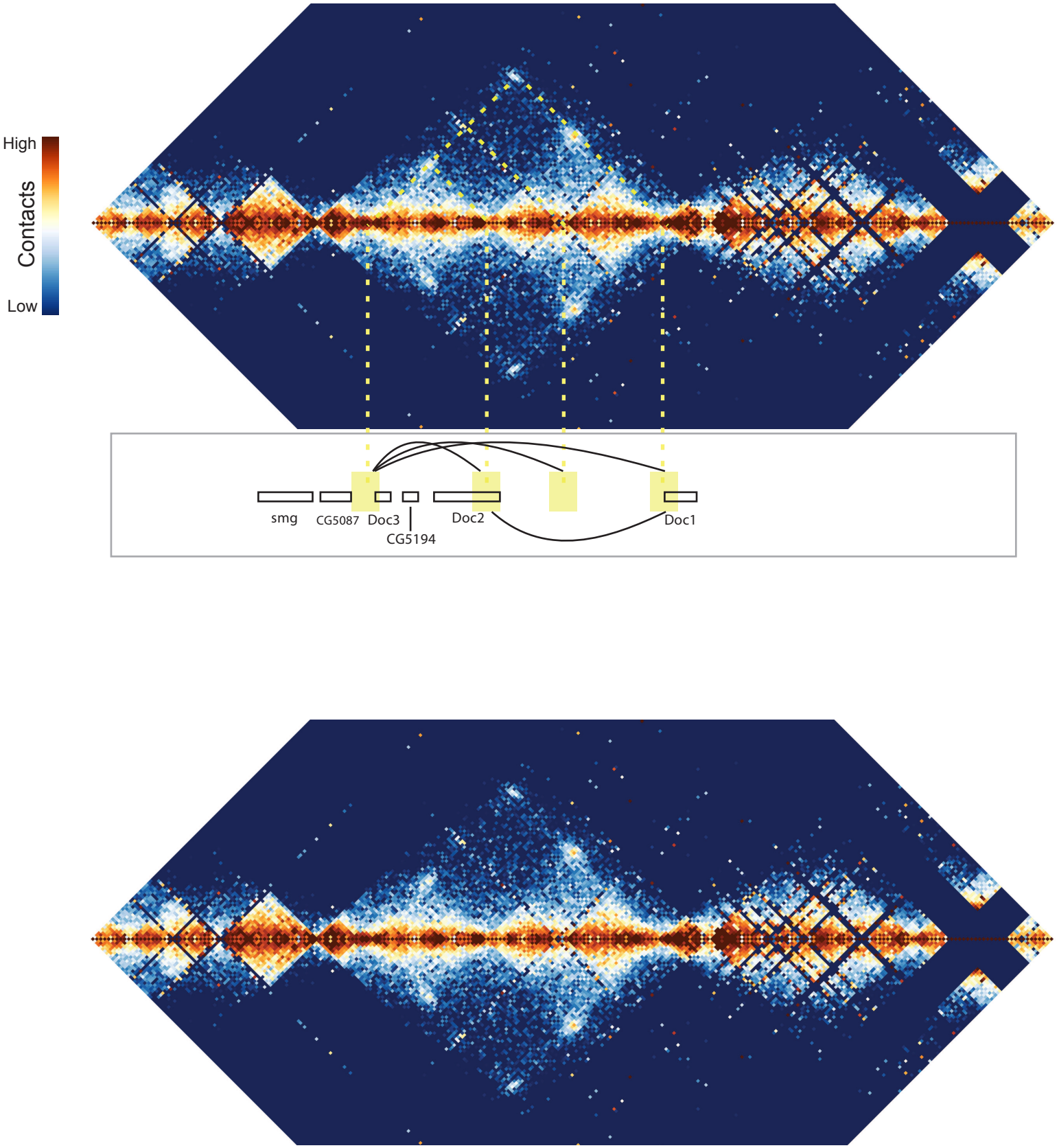

knrl
CG13251
kni

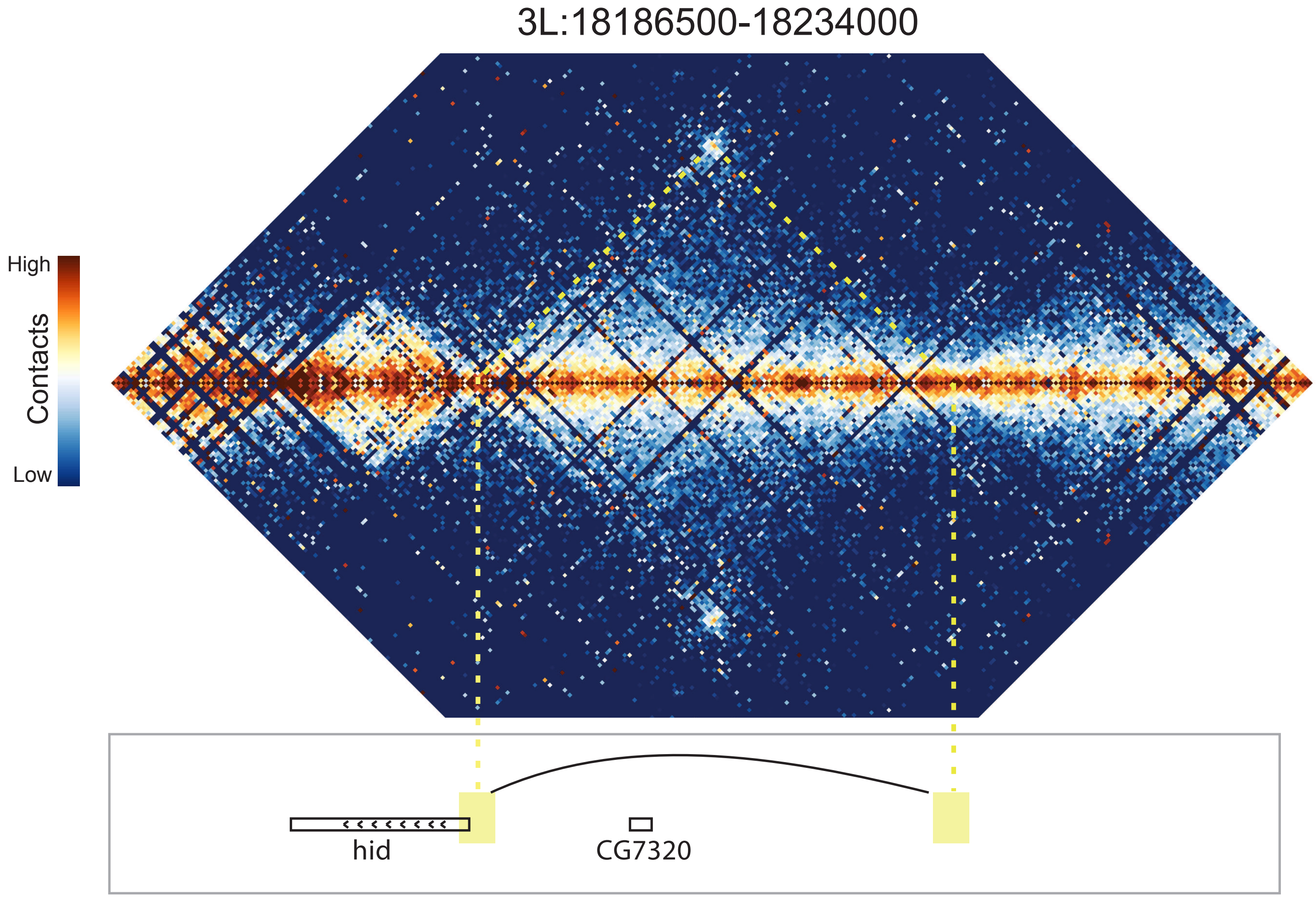

X:369500-396000
High
Contacts
Low
ac
sc

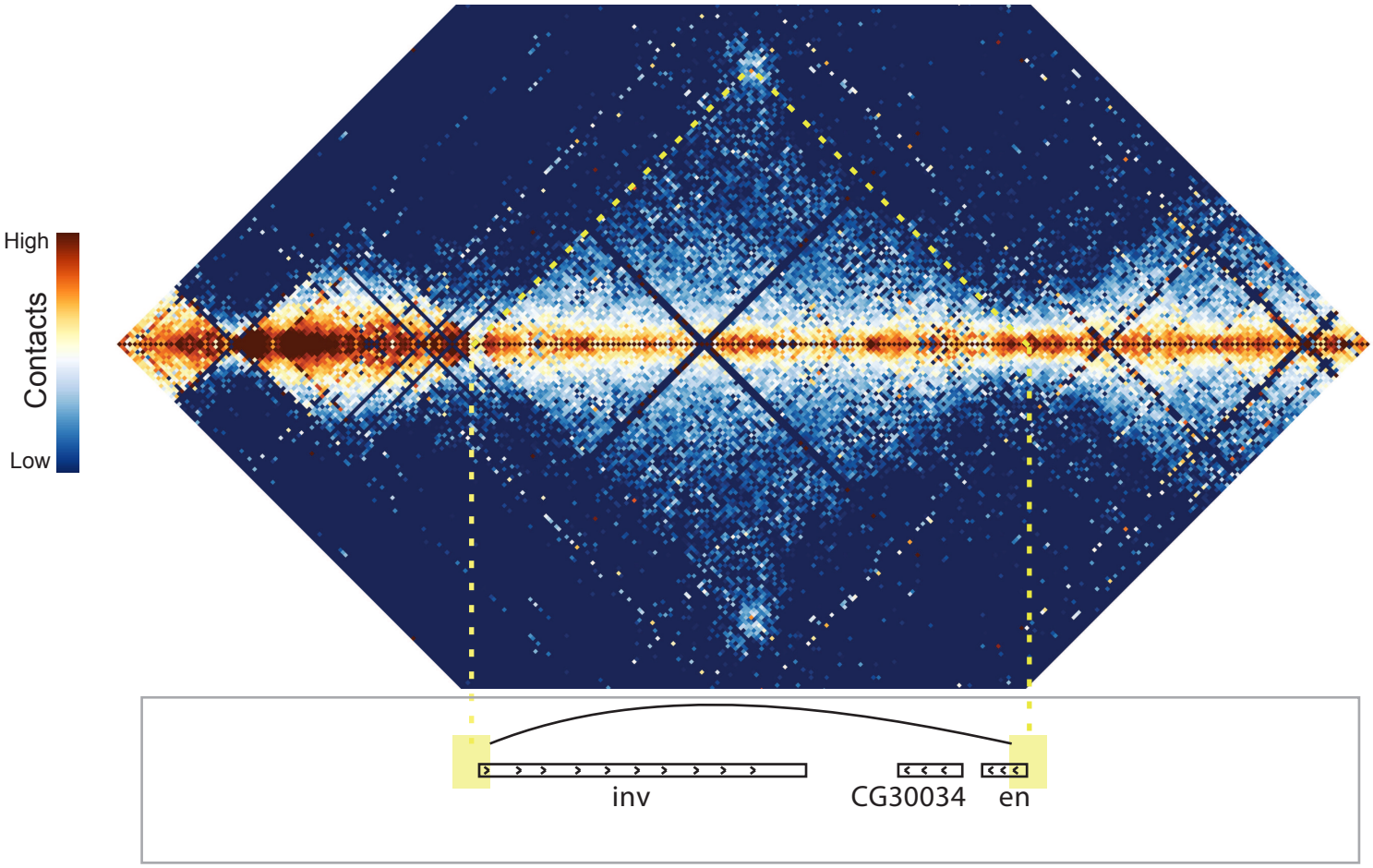

2L:3825500-3837000
High
Contacts
Low
slp1
slp2

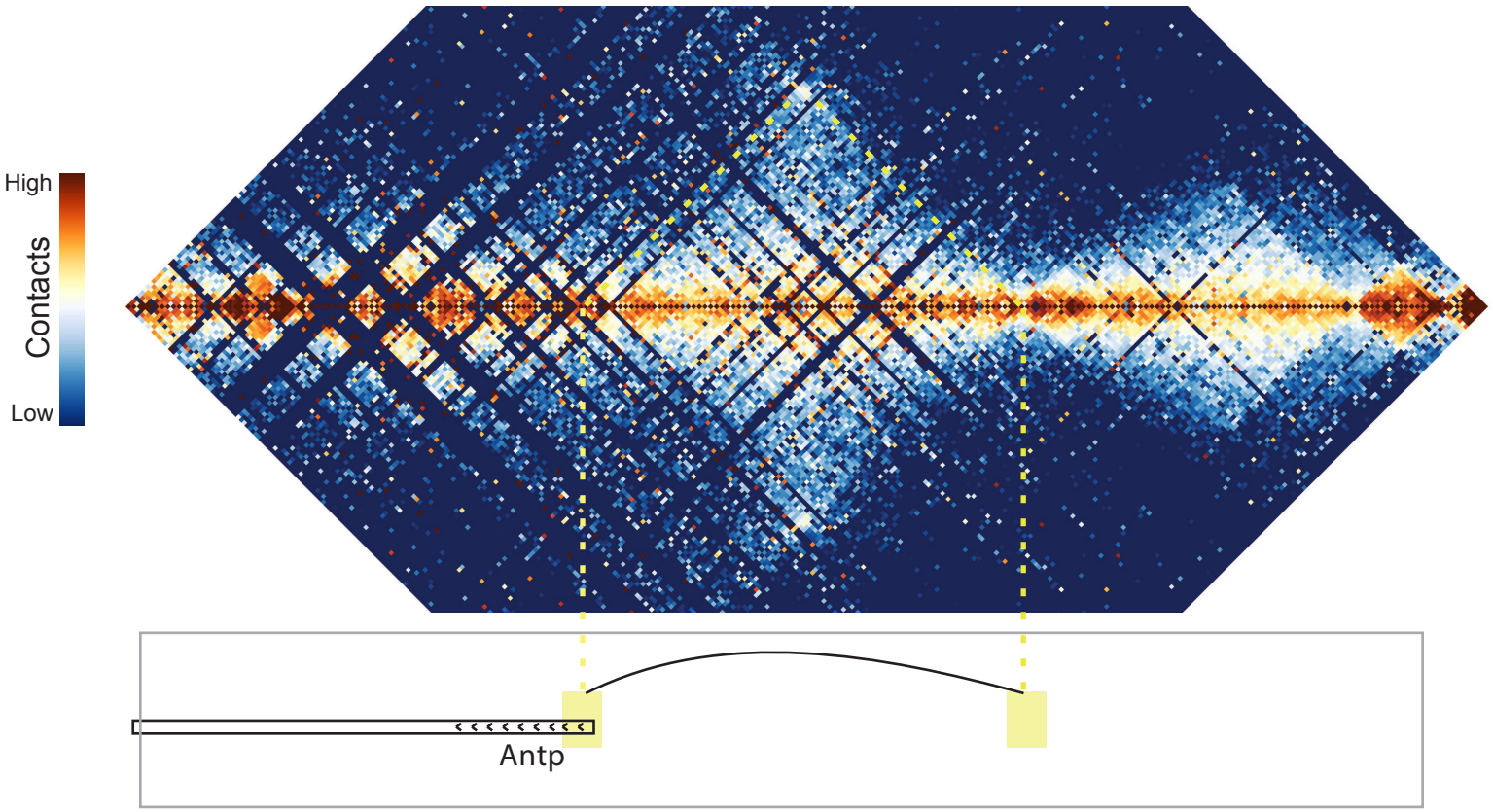
3R:16720000-16730500




3L:9003500-9040500
High
Contacts
Low
smg
CG5087
Doc3
CG5194
Doc2
Doc1


3L:18186500-18234000
High
Contacts
Low
hid
CG7320


2R:11474500-11528500
High
Contacts
Low
inv
CG30034
en


3R:6999000-7038000
High
Contacts
Low
Antp

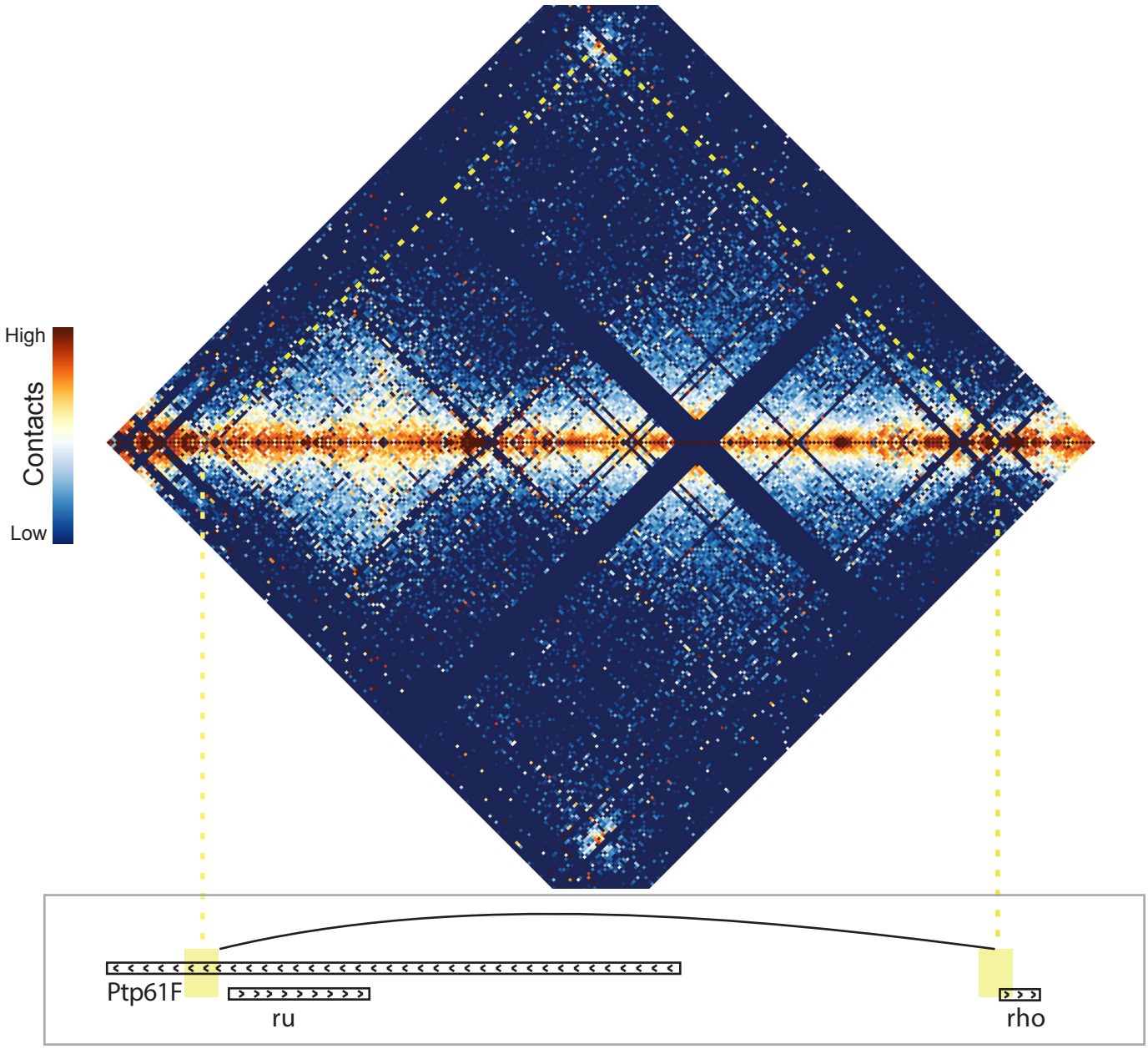
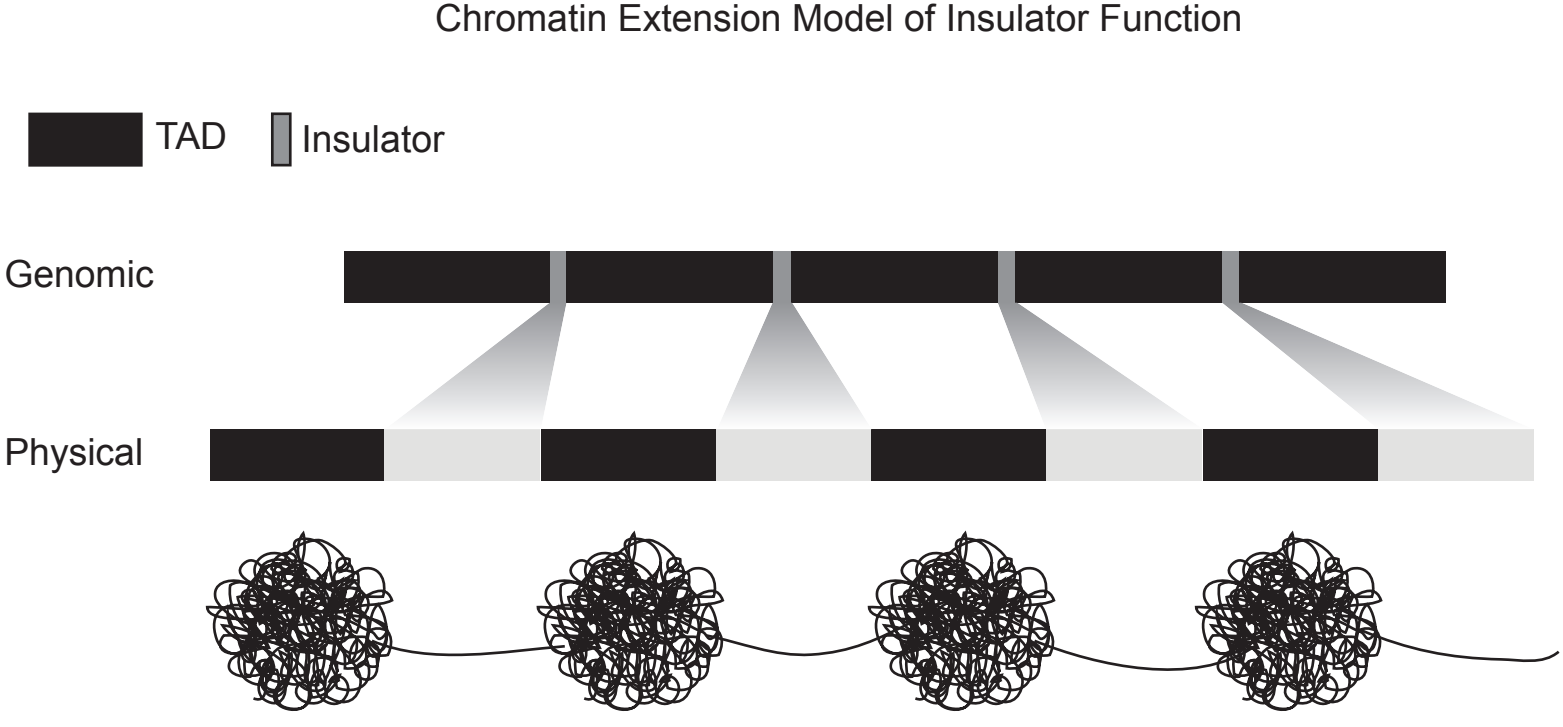

3L:1367500-1464000
High
Contacts
Low
Ptp61F
ru
rho

Figure 9


Chromatin Extension Model of Insulator Function
TAD
Insulator
Genomic
Physical