Published in final edited form as:
Cell. 2013 June 6; 153(6): 1281–1295. doi:10.1016/j.cell.2013.04.053.

Architectural protein subclasses shape 3-D organization of genomes during lineage commitment

Jennifer E. Phillips-Cremins[1,6], **Michael E. G. Sauria**[1,2], **Amartya Sanyal**[6], **Tatiana I. Gerasimova**[3], **Bryan R. Lajoie**[6], **Joshua S. K. Bell**[1], **Chin-Tong Ong**[1], **Tracy A. Hookway**[5], **Changying Guo**[3], **Yuhua Sun**[4], **Michael J. Bland**[1], **William Wagstaff**[1], **Stephen Dalton**[4], **Todd C. McDevitt**[5], **Ranjan Sen**[3], **Job Dekker**[6,*], **James Taylor**[1,2,*], and **Victor G. Corces**[1,*]

[1]Department of Biology, Emory University, Atlanta, GA, 30322.

[2]Department of Mathematics and Computer Science, Emory University, Atlanta, GA, 30322.

[3]Laboratory of Molecular Biology and Immunology, National Institute of Aging, Baltimore, MD, 21224.

[4]Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602.

[5]Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA 30332.

[6]Program in Systems Biology, University of Massachusetts Medical School, Worcester, MA, 01605.

Summary

Understanding the topological configurations of chromatin may reveal valuable insights into how the genome and epigenome act in concert to control cell fate during development. Here we generate high-resolution architecture maps across seven genomic loci in embryonic stem cells and neural progenitor cells. We observe a hierarchy of 3-D interactions that undergo marked reorganization at the sub-Mb scale during differentiation. Distinct combinations of CTCF, Mediator, and cohesin show widespread enrichment in looping interactions at different length scales. CTCF/cohesin anchor long-range constitutive interactions that form the topological basis for invariant sub-domains. Conversely, Mediator/cohesin together with pioneer factors bridge shortrange enhancer-promoter interactions within and between larger sub-domains. Knockdown of Smc1 or Med12 in ES cells results in disruption of spatial architecture and down-regulation of genes found in cohesin-mediated interactions. We conclude that cell type-specific chromatin organization occurs at the sub-Mb scale and that architectural proteins shape the genome in hierarchical length scales.

Introduction

Genomes are organized at multiple length scales into sophisticated higher-order architectures (Misteli, 2007). Individual chromosomes occupy distinct spatial territories in

*Co-corresponding authors: vcorces@emory.edu; job.dekker@umassmed.edu; james.taylor@emory.edu.

Accession Numbers
Data has been deposited in GEO under accession number GSE36203.

interphase nuclei (Cremer and Cremer, 2001). Within each territory, at intermediate length scales of ~1–10 Megabase (Mb), compartments of transcriptionally-active euchromatin tend to group together, and independent from, compartments of inactive heterochromatin (Lieberman-Aiden et al., 2009). At the sub-compartment level, chromatin is further organized into Mb-sized topologically associating domains (TADs) that represent spatial neighborhoods of high-frequency chromatin interactions (Dixon et al., 2012; Hou et al., 2012; Nora et al., 2012; Sexton et al., 2012). Within TADs, however, the precise features of chromatin folding at the sub-Mb scale remain poorly understood.

Emerging evidence suggests that nuclear architecture is critically important for cellular function. Seminal microscopy studies have linked the spatial positioning of specific genomic loci to gene expression (Fraser and Bickmore, 2007; Kosak and Groudine, 2004; Lanctot et al., 2007), replication (Gilbert et al., 2010), X-chromosome inactivation (Erwin and Lee, 2008; Nora and Heard, 2010), DNA repair (Misteli and Soutoglou, 2009), and chromosome translocations (Roix et al., 2003). Moreover, molecular methods based on proximity ligation, such as Chromosome Conformation Capture (3C) or Circularized-3C (4C), have been used to detect functional long-range interactions between two specific genomic loci in a population of cells (Dekker et al., 2002; Simonis et al., 2006; Zhao et al., 2006). Principles from these studies have been difficult to generalize, however, because most previous reports focus on interrogation of 3-D interactions between specific pre-selected fragments (Kurukuti et al., 2006; Noordermeer et al., 2011; Schoenfelder et al., 2010; Vakoc et al., 2005). More recently, technologies for genome-wide mapping of chromatin architecture have been described, but comprehensive detection comes at the expense of resolution for mammalian genomes (Hi-C) (Dixon et al., 2012; Lieberman-Aiden et al., 2009) or is restricted to only interactions mediated by a pre-selected protein of interest (ChIA-PET) (Handoko et al., 2011; Li et al., 2012). Thus, there is a great need to elucidate principles of genome folding at the sub-Mb scale by mapping 3-D chromatin interactions in an unbiased manner at high-resolution.

Mb-scale TADs appear to be constant between mammalian cell types and conserved across species (Dixon et al., 2012; Nora et al., 2012). Therefore, we hypothesized that genome organization at the sub-Mb scale, e.g. within TADs, plays a critical role in the establishment and/or maintenance of cellular state. To test this hypothesis, we present an unbiased, large-scale, and high-resolution analysis of 3-D chromatin architecture in a continuous developmental system. We employed Chromosome Conformation Capture Carbon Copy (5C) (Dostie et al., 2006) in combination with high-throughput sequencing to map higher-order chromatin organization during differentiation of pluripotent mouse ES cells along the neuroectoderm lineage. An alternating 5C primer design was applied to query longrange chromatin interactions in a massively parallel manner across six 1–2 Mb-sized genomic regions around key developmentally regulated genes (*Oct4, Nanog, Nestin, Sox2, Klf4, Olig1-Olig2*). Our analyses reveal that distinct combinations of architectural proteins shape the 3-D organization of mammalian genomes at different length scales for unique functional purposes during lineage commitment.

Results

Generation of high-resolution chromatin interaction maps

To investigate cell type-specific patterns of higher-order chromatin organization, we first derived homogeneous populations of multipotent neural progenitor cells (NPCs) from mouse ES cells using a well-established, four-stage procedure (Mikkelsen et al., 2007; Okabe et al., 1996). qRT-PCR analysis and confocal microscopy coupled with immunofluorescence staining confirmed a >95% pure population of Nestin/Sox2-postitive, Oct4/Nanog-negative cells displaying morphological characteristic of NPCs (Figure S1).

We then employed 5C in combination with high-throughput sequencing to generate high-resolution long-range interaction maps for two biological replicates of ES cells and ES-derived NPCs (Dostie et al., 2006). 5C, a high-throughput derivative of 3C, involves the selective amplification of chromatin interactions within specific regions of interest. By pre-selecting regions to be queried, we were able to obtain insight into chromatin architecture at the resolution of single interrogated fragments (~4 kb), which is not yet feasible in a cost-effective manner with genome-wide Hi-C technologies in mammalian systems. Forward and reverse 5C primers were designed in an alternating scheme using tools from the publicly available my5C suite (Lajoie et al., 2009) (Figure S2E). The tiled, alternating design queried ~90,000 cis and ~500,000 trans interactions in parallel across seven genomic loci surrounding developmentally regulated genes (*Oct4, Nanog, Sox2, Klf4, Nestin, Olig1-Olig2*, gene-desert control) (Table S1).

We first evaluated the quality of our raw 5C data by (1) assessing consistency between biological replicates and (2) comparing our high-resolution 5C data to Hi-C data recently reported at 40 kb resolution for E14 ES cells and primary cells isolated from mouse cortex (Dixon et al., 2012). Raw 5C counts were highly correlated between replicates (ES 1 vs. ES 2, Pearson correlation coefficient: 0.91; NPC 1 vs. NPC 2, Pearson correlation coefficient: 0.89) and more correlated than between ES cells and NPCs (Pearson correlation coefficient: 0.70) when considering the log of all fragment combinations with > 100 reads. Furthermore, heatmaps of raw 5C counts showed high similarity between biological replicates at each individual locus (Figure S2F), suggesting that region-specific primers amplified each region in a robust and consistent manner. Importantly, we also observe a striking similarity between global topological features in 5C and Hi-C data for all regions queried, albeit with marked differences in resolution (Figure S3). These results indicate that our 5C libraries are high quality, consistent between replicates, and achieve notable similarity to data generated with an independent method (Hi-C) in an independent study with similar cells.

Unique hierarchy of topological sub-domains at each genomic locus

We next examined large-scale architectural features by visualizing heatmaps of 5C counts in ES cells and NPCs (Figure 1C,F). Comparison of high-resolution 5C maps to Hi-C maps at each region revealed a complex hierarchy of chromatin organization (Figure 1A,B,D,E). TADs are readily detected in both Hi-C and 5C datasets. Importantly, the higher resolution of 5C data revealed that TADs previously defined with Hi-C are further subdivided into sub-topologies (sub-TADs). We systematically identified sub-TADs in 5C data with a Hidden Markov Model-based approach (**Extended Experimental Procedures**). Using this method, we uncovered numerous distinct subtopologies arranged in a hierarchy within the larger TAD organization. Indeed, > 60 invariant and cell type-specific sub-TAD boundaries were identified with our 5C data at the sub-Mb scale, whereas only 7 TAD boundaries were called in our regions of interest with a previous Hi-C analysis (Dixon et al., 2012) (Figure 1 G–H). Topological features were unique to each region (Figure S3), suggesting that each genomic locus has an architectural signature that may reflect the functional activity of that region. Taken together, these data demonstrate that 5C achieves a marked increase in resolution compared to Hi-C, which enables mapping of finer-scale architectural features within TADs.

Constitutive and cell type-specific features of 3-D chromatin organization

Our observation that sub-Mb-scale architectural features undergo marked changes between cell types prompted us to systematically identify constitutive and cell type-specific looping interactions within and between larger-scale TADs. To account for bias intrinsic to all 3C-based methods, as well as to 5C specifically, we developed a probabilistic model that simultaneously captures the distance-dependent background level of non-specific chromatin interactions and the non-biological contribution from each primer (Imakaev et al., 2012;

Yaffe and Tanay, 2011) (**Extended Experimental Procedures**). Our model produces an interaction score that is comparable within and between experiments, and allows for robust detection of fragment-to-fragment looping interactions that are significant above the expected background signal (Figure S4).

3-D contacts with interaction scores greater than stringent, pre-established thresholds in both biological replicates were subjected to further analysis (Figure 2A). To rigorously minimize false positives, thresholds were selected so that the large majority of cell-type specific interactions were lost when randomly permuting data (Figure 2B). By applying these stringent thresholds, we identified 83 ES cell-specific interactions that are lost upon differentiation, 260 constitutive interactions that are constant between cell types, and 165 NPC-specific interactions that are absent in ES cells and acquired upon differentiation (Figure 2C). Thus, only cell type-specific architectural features corresponding to the top 0.096% and 0.190% of all queried interactions in ES and NPC libraries, respectively, were considered for downstream analysis.

We next integrated 5C data with other epigenomic data sets. We observed that a significant proportion of fragments engaged in 3-D interactions were occupied by specific histone modifications. For example, a series of ~80–120 kb-sized interactions connect the *Sox2* gene with a putative downstream enhancer in ES cells marked by H3K4me1, H3K27ac, and low levels of H3K4me3 (Creyghton et al., 2010; Heintzman et al., 2009; Rada-Iglesias et al., 2011) (Figure 2D). Loss of enhancer marks in NPCs occurs in parallel with loss of ES-specific looping interactions, suggesting that this particular chromatin conformation has important functional significance. Similarly, an ES-specific interaction connects the *Pou5f1/Oct4* gene to a putative enhancer ~25 kb upstream marked by H3K4me1, H3K27ac, and low levels of H3K4me3 (Figure 2E). By contrast, we detected a hierarchy of constitutive interactions around the pluripotent genes *Nanog* and *Slc2a3* that were constant between cell types despite changes in gene activity during differentiation (Figure 2F). These examples provide evidence that a notable proportion of looping interactions identified in this study may be involved in genome function.

Candidate architectural protein subclasses

To gain more insight into organizing principles governing genome folding, we integrated 5C data with genome-wide maps of protein occupancy. We first examined factors that have been reported as both essential for cellular functions and correlated with a specific looping interactions using 3C technology. The top three candidates fulfilling these criteria were CTCF, cohesin, and Mediator (Hadjur et al., 2009; Handoko et al., 2011; Kagey et al., 2010; Kurukuti et al., 2006; Splinter et al., 2006). Genomewide binding sites for CTCF, mediator subunit Med12, and cohesin subunit Smc1 have been previously identified in ES cells by ChIP-seq (Kagey et al., 2010; Stadler et al., 2011). We first used the ChIP-seq data to quantify unique and overlapping occupied sites in our regions of interest. As previously reported, high-confidence Smc1 binding sites significantly overlapped high-confidence CTCF and Med12 binding sites (Kagey et al., 2010). However, in addition to Med12+Smc1 and CTCF+Smc1 co-occupied sites, we also found notable subclasses of CTCF alone and Med12 alone sites both genomewide and in our regions of interest in ES cells (Figure 3A–B). Noteworthy, Med12 rarely overlaps CTCF in the absence of cohesin, but a sub-class with occupancy of all three proteins (i.e. Med12+Smc1+CTCF) does emerge as significant.

Architectural proteins organize the genome at different length scales

We next examined the enrichment of CTCF, Med12, and Smc1 in 5C looping interactions. Unsupervised cluster analysis demonstrated that >80% of significant interactions were anchored by some combination of CTCF, Med12, or Smc1 in ES cells, which is significantly

higher than the enrichment of these proteins in all queried background interactions (Figure 3C). By contrast, only ~40% of interactions were occupied by some combination of Oct4, Nanog, and/or Sox2, which is not significant compared to the expected background enrichment of these proteins (Figure 3D). The widespread occupancy of CTCF, Med12, and Smc1 in 3-D interactions led us to hypothesize that these three proteins might have important architectural roles in shaping 3-D genome organization.

To investigate the possibility that a class of proteins exist for solely architectural purposes, we explored the specific role for each candidate architectural protein subclass in genome organization. We observed a striking pattern in which multiple adjacent binding sites for the same architectural protein subclass were often found at the base of significant interactions. Indeed, enrichment for a particular architectural subclass in 3-D interactions showed a strong correlation with the number of binding sites (Figure 3E). Therefore, to explore only high-confidence interactions, we focused our analysis on only loops anchored by > 2 or >3 co-occupied sites (Figure 3F). CTCF+Smc1 and CTCF Alone subclasses were highly over-represented at the base of constitutive interactions compared to background non-loops (Figure 3F). By contrast, Med12+Smc1 and Med12 alone subclasses were predominantly enriched in only ES-specific looping interactions. Intriguingly, sites co-occupied by Med12+CTCF+Smc1 showed enrichment in both constitutive and ES cell-specific interactions, suggesting the potential of this subclass to mediate both types of chromatin interactions.

We also noticed that interactions mediated by each candidate architectural subclass displayed markedly different size distributions (Figure 4A). Med12+Smc1 cooccupied sites were predominately enriched at the smallest <100 kb length scale (Figure 4B), whereas Med12 alone sites, independent from cohesin, were enriched at intermediate length scales of 600–1000 kb (Figure 4D). The subclass with all three proteins (Med12+CTCF+Smc1) also displayed a loop size distribution shifted toward small to intermediate (<300 kb) length scales (Figure 4C). By contrast, the size distribution of loops connected by CTCF+Smc1 and CTCF Alone subclasses were significantly biased toward interactions greater than 1 Mb in size (Figure 4E–F). Together these results suggest that architectural protein subclasses function at different length scales to fulfill distinct roles in genome organization.

CTCF and cohesin anchor constitutive interactions

To further explore the molecular mechanisms regulating constitutive chromatin interactions, we mapped CTCF and Smc1 occupancy in NPCs using ChIP-seq. Genomic loci co-occupied by CTCF and Smc1 in both ES cells and NPCs represented the largest architectural subclass genome-wide and in our regions of interest (n=159) (Figure 5A–B). Moreover, the sites with constant occupancy of CTCF+Smc1 between cell types were highly enriched in constitutive interactions compared to background (Figure 5C). This result is illustrated with a series of loops around *Nanog* and *Slc2a3* (Figure 2F) and *Olig1* and *Olig2* (Figure 5D, Figure S6A). At both genomic loci, fragments anchoring the base of constitutive interactions contain CTCF+Smc1 co-occupied sites that remain constant between ES cells and NPCs. Thus, constitutive CTCF occupancy may be a critical mechanism regulating the establishment and/or maintenance of constitutive chromatin architecture.

We next used high-resolution 3D fluorescent in-situ hybridization (FISH) to assess the importance of CTCF/cohesin in configuring architecture. Two 10 kb probes (Figure 5D) corresponding to the base of the predicted constitutive looping interaction around *Olig1-Olig2* produced virtually super-imposable FISH signals in the majority of wild type ES cells (Figure 5E–F). Having validated this looping interaction with an independent assay, we then directly tested the role for CTCF and cohesin by knocking down these proteins in V6.5 ES cells. CTCF and Smc1 mRNAs were markedly depleted to <20% of their wild type

expression levels after transduction of ES cells with lentiviral shRNA constructs and subsequent puromycin selection (Figure S5). By contrast to observations in wild type nuclei, FISH probes were no longer co-localized in CTCF- and Smc1-KD ES cells (Figure 5E–F). These data indicate loop disruption and provide strong evidence that 3-D contacts identified by 5C represent bona fide chromatin interactions. We conclude that both CTCF and Smc1 are essential for maintaining this particular constitutive interaction, and propose that similar mechanisms will apply to constitutive interactions genome-wide.

Mediator and cohesin bridge proximal enhancer-promoter interactions

To understand the organizing principles regulating cell type-specific chromatin architecture, we first examined CTCF-independent Smc1 sites that were occupied only in ES cells and then lost upon differentiation (n=123) (Figure 5B, Figure 6A). Loss of ES-specific Smc1 in NPCs occurred in parallel with abrogation of ES-specific interactions (Figure 6B), supporting the idea that Smc1 can function in a CTCF-independent manner as an architectural protein essential for cell type-specific chromatin interactions.

We then set out to identify co-factors that partner with cohesin to bridge cell typespecific interactions. Genome-wide analysis of CTCF-independent, ES-specific Smc1 binding sites revealed a strong co-localization with Mediator and pioneer transcription factors Oct4, Sox2, and Nanog (OSN) (Soufi et al., 2012) (Figure 6A). Indeed, > 95% of all ES-specific Smc1 binding sites co-localize with Med12, whereas only ~55% of these sites co-localize with OSN in our regions of interest (Figure 6C). Importantly, ES-specific Smc1 binding sites were enriched in ES-specific interactions in both cases where these sites co-localized with OSN and also in cases where these sites did not colocalize with OSN (Figure 6D). These observations suggest that cohesin does not require pioneer factors OSN to serve an architectural role in the establishment and/or maintenance ES-specific chromatin interactions.

We further investigated the mechanistic link between pioneer transcription factors and chromatin architecture by parsing OSN subclasses genome-wide and in our regions of interest (Figure 6E, Figure S6D). We noticed a partial overlap between OSN occupied sites genome-wide and architectural proteins in ES cells. Indeed, in our regions of interest, ~50% of OSN binding sites co-localized with Med12+Smc1, whereas ~25% did not co-localize with any architectural protein subclass (Figure 6F). Importantly, OSN pioneer factors were only enriched in ES-specific looping interactions in cases where these proteins co-localized with architectural proteins (Figure 6G). Together, these data suggest that OSN do not have a specific role in long-range chromatin organization independent from architectural proteins.

To validate the roles for Mediator and cohesin in ES cell-specific looping interactions, we carried out high-resolution 3D-FISH in wild type and Med12- or Smc1-knockdown V6.5 ES cells. For this analysis we chose an interaction between *Olig1* and a putative downstream ES cell-specific enhancer (Figure S6B–C). Probes generated from these interacting regions (Figure S6C) co-localized in WT ES nuclei, but not in Med12- or Smc1-KD cells (Figure 6H–I). We conclude that Mediator and cohesin are essential for formation of an ES cell-specific loop at *Olig1*, and propose that similar mechanisms will apply to other ES-specific chromatin interactions.

Cohesin-mediated interactions are functionally linked to gene expression

To further test the hypothesis that Med12+Smc1-mediated interactions have functional significance during lineage commitment, we examined the expression of genes in looping interactions connected by these proteins. Analysis of microarray data generated in ES cells and NPCs (Creyghton et al., 2010) demonstrated that ES-specific, Smc1-mediated

interactions are biased toward connecting genes that are highly expressed in ES cells and turned off in NPCs (Figure 6J). Gene ontology analysis confirmed an over-representation of developmentally regulated pluripotent genes (e.g. *Pou5f1/Oct4*, *Sox2* and *Notch4)* in ES-specific interactions connected by ES-specific Smc1 compared to all genes in ES-specific interactions (Figure S6E). By contrast, the expression distribution of genes co-localized with CTCF+Smc1 in constitutive looping interactions was not significantly different from the expression distribution of all genes in constitutive looping interactions (Figure 6K).

We then examined the effect of knocking down Med12 and Smc1 on gene expression. After siRNA knock-down of either Smc1 or Med12 in ES cells (Kagey et al., 2010), expression of genes found in cohesin-mediated interactions was markedly reduced compared to expression of all genes found in ES-specific interactions (Figure 6L). Noteworthy, the reduction in gene expression after siRNA treatment was more severe for cohesin-co-localized genes found in looping interactions vs. non-looping background. These results expand upon previous reports at specific genomic loci (Kagey et al., 2010) by suggesting that the architectural roles for Mediator and cohesin might be a widespread mechanism linking gene expression and chromatin organization genome-wide.

Overall, these data are consistent with a model in which Mediator/cohesin connect ES cell-specific looping interactions between proximal regulatory elements and promoters of developmentally regulated pluripotent genes. This idea is illustrated at the Sox2 locus, where a series of ~80–120 kb-sized looping interactions connects the Sox2 TSS to a putative active enhancer (Figure 2E). OSN pioneer factors and the Med12+Smc1 architectural subclass co-localize at the fragments connecting these loops. Loss of architectural protein binding and ES-specific looping interactions in NPCs occurs in parallel with loss of pluripotent gene expression, suggesting that chromatin structure and function are intricately linked.

Architectural proteins facilitate looping of cell type-specific enhancers

The involvement of Mediator and cohesin in relatively short-range enhancerpromoter interactions prompted us to look more broadly at the relationship between architectural proteins and distal cell type-specific regulatory elements. Putative ES-specific enhancers were parsed as genomic loci at least 2 kb distal from TSSs with high-confidence signal for H3K4me1 and H3K27ac in ES cells and loss of these chromatin marks in NPCs (Figure 7A). Similarly, putative NPC-specific enhancers were parsed as genomic loci at least 2 kb distal from TSSs that do not display H3K27ac signal in ES cells, but acquire high-confidence signal for H3K4me1 and H3K27ac in NPCs (Figure 7F). We noticed that putative ES-specific and NPC-specific enhancers could be sorted from high to low intensity of H3K4me3 signal (Figure 7A, F). A strong correlation was observed between ES-specific enhancers displaying H3K4me3 signal and RNA Pol II occupancy, suggesting that enhancers marked by H3K4me3 might have active transcription of eRNAs (Kim et al., 2010) (Figure 7A).

To further explore the link between enhancers and chromatin architecture, we parsed ES-specific and NPC-specific enhancers in our regions of interest into high, intermediate, and low levels of H3K4me3 (Figure 7B, G). Noteworthy, only ES-specific enhancers with high levels of H3K4me3 were enriched in ES-specific looping interactions, whereas ES-specific enhancers with low levels of H3K4me3 were enriched only in constitutive interactions compared to background (Figure 7C). Similarly, NPC-specific enhancers with high and intermediate levels of H3K4me3 were enriched in NPC-specific looping interactions, whereas NPC-specific enhancers with low levels of H3K4me3 were not enriched in chromatin interactions compared to background (Figure 7H). Altogether these observations

provide support for the idea that eRNA transcription correlates with enhancer activity and subsequent activity-dependent looping of cell typespecific enhancers.

As a final step, we queried the potential co-localization of enhancers with architectural proteins. We noticed that the majority of ES-specific enhancers (~95%) colocalized with architectural proteins in ES cells (Figure 7D). Consistent with our previous analyses, ES-specific enhancers that co-localized with Smc1 alone were enriched in ES-specific looping interactions, whereas ES-specific enhancers colocalized with CTCF+Smc1 were enriched predominantly in constitutive interactions (Figure 7E). Intriguingly, only ~25% of NPC-specific enhancers co-localized with architectural proteins (Figure 7I). NPC-specific enhancers that co-localized with CTCF+Smc1 were enriched in NPC-specific looping interactions, whereas NPC-specific enhancers that did not co-localize with architectural proteins were not enriched in 3-D structure (Figure 7J). Because the CTCF+Smc1 architectural subclass is markedly enriched in constitutive interactions and also displays a slight enrichment in NPC-specific interactions (Figure 5C), we hypothesize that one functional purpose for constitutive sub-domains is to pre-mark specific locations in the genome that will acquire or lose enhancer activity during development.

An example of architectural proteins cooperating with enhancers is shown at the Sox2 locus, where the *Sox2* gene is highly expressed in both ES cells and NPCs (Figure 7K). Data indicate that this developmentally regulated gene is controlled by different regulatory elements even though expression levels remain high as cells switch fate. In ES cells, the *Sox2* TSS is connected to a proximal enhancer through a series of ~100kb-sized interactions, whereas in NPCs these smaller looping interactions break apart, and a larger (~450kb) sub-domain is present between the *Sox2* TSS and a more distal NPC enhancer. Mediator/cohesin mark the proximal ES-specific enhancer while CTCF/ cohesin pre-mark the sub-TAD boundary in ES cells that ultimately acquires a distal NPC-specific enhancer upon differentiation. These results support a prevalent role for architectural proteins in spatially connecting proximal and distal enhancer elements to the genes that they regulate. Although enhancers and insulators are traditionally thought to serve distinct mechanistic functions in gene regulation, our data suggest that enhancers and architectural proteins may work in collaboration to organize the genome much more than previously realized.

Discussion

Analyses presented here provide an important step toward understanding the link between higher-order chromatin architecture, epigenetic modifications, and cell typespecific gene expression. By analyzing the genome in 3-D, we now discover that three proteins thought to play more traditional roles in transcriptional activation and insulation might belong to a class of architectural proteins with primarily “chromatin organizing” function. It was originally suggested that vertebrate CTCF is an insulator protein based on transgene studies demonstrating that this protein blocks communication between adjacent regulatory elements in a position-dependent manner. However, data presented here are more consistent with recent reports suggesting that enhancer blocking or barrier insulation may only occur in a context-dependent manner as a consequence of CTCF’s primary role in connecting long-range interactions (Handoko et al., 2011; Sanyal et al., 2012). Although traditionally considered an adaptor protein with multiple subunits essential for transcriptional activation (Kornberg, 2005; Malik and Roeder, 2000), we suggest that the widespread enrichment in 3-D interactions predicts a similar architectural role for Mediator. In the case of cohesin, a protein with a well-known architectural function during mitosis, recent reports have suggested a critical mechanistic role in insulation based on the finding that CTCF and cohesin co-localize across the genome at thousands of loci (Parelho et al., 2008; Wendt et al., 2008). Our data predicts that cohesin functionally affects gene expression through its

architectural role during interphase instead of through classic insulator mechanisms (Hadjur et al., 2009).

The molecular mechanisms governing chromatin folding at the sub-Mb scale remain critical unanswered questions in nuclear biology. CTCF is present at most boundaries between conserved Mb-sized TADs (Dixon et al., 2012; Nora et al., 2012). However, this protein cannot be considered the sole determinant of topological organization because the majority of CTCF sites are found within TADs. Here we expand our understanding of chromosome organization at the sub-Mb scale by combining high-resolution 5C data and a detailed probabilistic model to computationally resolve individual fragment-to-fragment looping interactions within TADs. Our analyses reveal that larger, invariant TADs are hierarchically organized into constitutive and cell type-specific sub-topologies. By integrating 5C architecture maps with genome-wide maps of epigenetic modifications, we observed that a large proportion of sub-domains coincide with specific looping interactions between architectural proteins and other regulatory sequences. Thus, we favor the idea that architectural protein binding sites found within larger TADs could be responsible for connecting interactions that form the topological basis for sub- domains. For example, the Mb-sized region around the *Olig1* and *Olig2* genes spans a boundary between two larger Mb-sized TADs. Interactions identified in the present work with 5C reveal a further nested hierarchy of constitutive interactions anchored by CTCF+Smc1 and ES-specific interactions within and between sub-domains connected by Med12+Smc1 (Figure S7).

We suggest a refined model for genome organization in which architectural protein subclasses function at different length scales to fulfill distinct roles in genome organization (Figure 7L). Data presented here predict that Mb-sized TADs are constant throughout development and are demarcated by constitutive occupancy of CTCF/cohesin at their boundaries. Within TADs, at intermediate length scales of 100 kb - 1 Mb, CTCF/cohesin co-occupied sites create sub-domains by anchoring constitutive interactions around developmentally regulated or repressed tissue-specific genes. One possible functional role for CTCF/cohesin-mediated interactions at this intermediate length scale is to cooperate with distal regulatory elements to connect long-range enhancer-promoter interactions. Finally, at the smallest length scale (<100 kb), Mediator and cohesin cooperate to bridge ES-specific interactions between enhancers and core promoters of developmentally regulated genes. We note that a small but significant proportion of CTCF/Mediator/cohesin binding sites are not involved in chromatin interactions. It is possible that these sites are involved in interactions outside of the 5C regions queried in this study. Moreover, it is also possible that architectural proteins require additional layers of regulation that we do not yet understand, such as posttranslational modifications or additional binding partners that regulate protein-mediated chromatin organization in a spatiotemporal manner (Phillips and Corces, 2009).

Overall, this work provides new insights into the organizing principles governing higher-order chromatin architecture and is significant toward understanding how the genome and the epigenome act in concert to regulate the formation of a diverse array of cell types during development. Powerful insights into the link between the hierarchical organization of 3-D genomes and cellular function could be realized by future studies combining high-resolution chromatin architecture analyses with genetic experiments in developmentally-relevant model systems.

Experimental Procedures

Methods are available in the Extended Experimental Procedures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Research reported in this publication was supported by the National Institute of General Medical Sciences of the NIH under award number R01GM035463 (V.G.C), the National Human Genome Research Institute of the NIH under award number R01HG003143 (J.D.), the National Institute of Diabetes and Digestive and Kidney Diseases of the NIH under award number R01DK065806 (J.T.), a W.M. Keck Foundation Distinguished Young Scholar in Biomedical Research Award (J.D.), and award number RC2HG005542 (J.T.) provided by funds from the American Recovery and Reinvestment Act (ARRA). J.E.P.C. was supported by an NIH NRSA postdoctoral fellowship (5F32NS065603). Funds from NIH grant PO1GM85354 also contributed to this work (S.D.).

References

Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet. 2001; 2:292–301. [PubMed: 11283701]

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010; 107:21931–21936. [PubMed: 21106759]

Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science. 2002; 295:1306–1311. [PubMed: 11847345]

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012; 485:376–380. [PubMed: 22495300]

Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, Rubio ED, Krumm A, Lamb J, Nusbaum C, et al. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res. 2006; 16:1299–1309. [PubMed: 16954542]

Erwin JA, Lee JT. New twists in X-chromosome inactivation. Curr Opin Cell Biol. 2008; 20:349–355. [PubMed: 18508252]

Fraser P, Bickmore W. Nuclear organization of the genome and the potential for gene regulation. Nature. 2007; 447:413–417. [PubMed: 17522674]

Gilbert DM, Takebayashi SI, Ryba T, Lu J, Pope BD, Wilson KA, Hiratani I. Space and time in the nucleus: developmental control of replication timing and chromosome architecture. Cold Spring Harb Symp Quant Biol. 2010; 75:143–153. [PubMed: 21139067]

Hadjur S, Williams LM, Ryan NK, Cobb BS, Sexton T, Fraser P, Fisher AG, Merkenschlager M. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature. 2009; 460:410–413. [PubMed: 19458616]

Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, Lee CW, Ye C, Ping JL, Mulawadi F, et al. CTCF-mediated functional chromatin interactome in pluripotent cells. Nat Genet. 2011; 43:630–638. [PubMed: 21685913]

Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature. 2009; 459:108–112. [PubMed: 19295514]

Hou C, Li L, Qin ZS, Corces VG. Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. Mol Cell. 2012; 48:471–484. [PubMed: 23041285]

Imakaev M, Fudenberg G, McCord RP, Naumova N, Goloborodko A, Lajoie BR, Dekker J, Mirny LA. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat Methods. 2012; 9:999–1003. [PubMed: 22941365]

Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, et al. Mediator and cohesin connect gene expression and chromatin architecture. Nature. 2010; 467:430–435. [PubMed: 20720539]

Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, et al. Widespread transcription at neuronal activityregulated enhancers. Nature. 2010; 465:182–187. [PubMed: 20393465]

Kornberg RD. Mediator and the mechanism of transcriptional activation. Trends Biochem Sci. 2005; 30:235–239. [PubMed: 15896740]

Kosak ST, Groudine M. Form follows function: The genomic organization of cellular differentiation. Genes Dev. 2004; 18:1371–1384. [PubMed: 15198979]

Kurukuti S, Tiwari VK, Tavoosidana G, Pugacheva E, Murrell A, Zhao Z, Lobanenkov V, Reik W, Ohlsson R. CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. Proc Natl Acad Sci U S A. 2006; 103:10684–10689. [PubMed: 16815976]

Lajoie BR, van Berkum NL, Sanyal A, Dekker J. My5C: web tools for chromosome conformation capture studies. Nat Methods. 2009; 6:690–691. [PubMed: 19789528]

Lanctot C, Cheutin T, Cremer M, Cavalli G, Cremer T. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. Nat Rev Genet. 2007; 8:104–115. [PubMed: 17230197]

Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell. 2012; 148:84–98. [PubMed: 22265404]

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009; 326:289–293. [PubMed: 19815776]

Malik S, Roeder RG. Transcriptional regulation through Mediator-like coactivators in yeast and metazoan cells. Trends Biochem Sci. 2000; 25:277–283. [PubMed: 10838567]

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007; 448:553–560. [PubMed: 17603471]

Misteli T. Beyond the sequence: cellular organization of genome function. Cell. 2007; 128:787–800. [PubMed: 17320514]

Misteli T, Soutoglou E. The emerging role of nuclear architecture in DNA repair and genome maintenance. Nat Rev Mol Cell Biol. 2009; 10:243–254. [PubMed: 19277046]

Noordermeer D, Leleu M, Splinter E, Rougemont J, De Laat W, Duboule D. The dynamic architecture of Hox gene clusters. Science. 2011; 334:222–225. [PubMed: 21998387]

Nora EP, Heard E. Chromatin structure and nuclear organization dynamics during X-chromosome inactivation. Cold Spring Harb Symp Quant Biol. 2010; 75:333–344. [PubMed: 21447823]

Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 2012; 485:381–385. [PubMed: 22495304]

Okabe S, Forsberg-Nilsson K, Spiro AC, Segal M, McKay RD. Development of neuronal precursor cells and functional postmitotic neurons from embryonic stem cells in vitro. Mech Dev. 1996; 59:89–102. [PubMed: 8892235]

Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, Gregson HC, Jarmuz A, Canzonetta C, Webster Z, Nesterova T, et al. Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell. 2008; 132:422–433. [PubMed: 18237772]

Phillips JE, Corces VG. CTCF: master weaver of the genome. Cell. 2009; 137:1194–1211. [PubMed: 19563753]

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. Nature. 2011; 470:279–283. [PubMed: 21160473]

Roix JJ, McQueen PG, Munson PJ, Parada LA, Misteli T. Spatial proximity of translocation-prone gene loci in human lymphomas. Nat Genet. 2003; 34:287–291. [PubMed: 12808455]

Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. Nature. 2012; 489:109–113. [PubMed: 22955621]

Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, Kurukuti S, Mitchell JA, Umlauf D, Dimitrova DS, et al. Preferential associations between coregulated genes reveal a transcriptional interactome in erythroid cells. Nat Genet. 2010; 42:53–61. [PubMed: 20010836]

Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G. Three-dimensional folding and functional organization principles of the Drosophila genome. Cell. 2012; 148:458–472. [PubMed: 22265598]

Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet. 2006; 38:1348–1354. [PubMed: 17033623]

Soufi A, Donahue G, Zaret KS. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell. 2012; 151:994–1004. [PubMed: 23159369]

Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, Galjart N, de Laat W. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev. 2006; 20:2349–2354. [PubMed: 16951251]

Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature. 2011; 480:490–495. [PubMed: 22170606]

Vakoc CR, Letting DL, Gheldof N, Sawado T, Bender MA, Groudine M, Weiss MJ, Dekker J, Blobel GA. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. Mol Cell. 2005; 17:453–462. [PubMed: 15694345]

Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T, et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature. 2008; 451:796–801. [PubMed: 18235444]

Yaffe E, Tanay A. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. Nat Genet. 2011; 43:1059–1065. [PubMed: 22001755]

Zhao Z, Tavoosidana G, Sjolinder M, Gondor A, Mariano P, Wang S, Kanduri C, Lezcano M, Sandhu KS, Singh U, et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet. 2006; 38:1341–1347. [PubMed: 17033624]

Highlights

- Genome architecture is reorganized at the sub-Mb scale during differentiation
- CTCF, Mediator, and cohesin have widespread and unique roles in genome organization
- CTCF and cohesin anchor long-range, constitutive interactions
- Mediator and cohesin bridge short-range, cell type-specific enhancerpromoter interactions

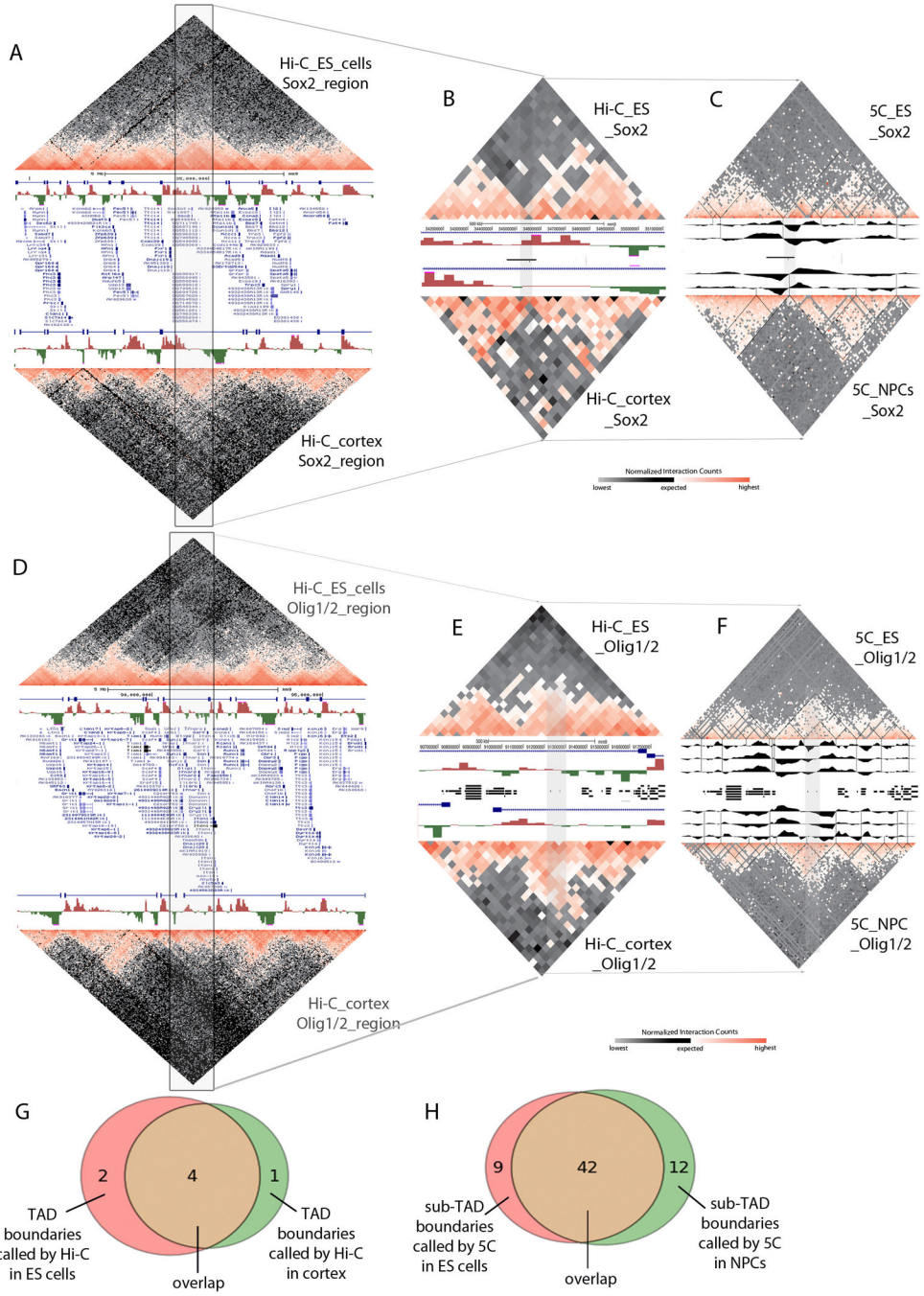

Figure 1. High-resolution mapping reveals a hierarchy of architectural subdomains within larger topological domains

(**A–F**) 5C and Hi-C interaction frequencies represented as normalized two-dimensional heat maps. (**A,B,D,E**) Hi-C data (adapted from (Dixon et al., 2012)) displayed for (**A,D**) 10 Mb and (**B,E**) 1 Mb regions around (**A,B**)*Sox2* and (**D,E**)*Olig1-Olig2* for mouse E14 ES cells (top) and mouse cortex (bottom). TADs reported in (Dixon et al., 2012) are represented as tracks for domain calls (blue bars) and directionality index (downstream bias (green), upstream bias (red)). (**C, F**) 5C data displayed for 1 Mb regions around (**C**)*Sox2* and (**F**) *Olig1-Olig2* for mouse V6.5 ES cells (top) and ES-derived NPCs (bottom). Constitutive and cell type-specific sub-domains called with our Hidden Markov Model (see **Extended**

Experimental Procedures) are represented as black lines overlaid on 5C heatmaps and a directionality index displayed as a hierarchy of black wiggle tracks. (**G, H**) Overlap between cell types for (**G**) TAD boundaries called from Hi-C data in (Dixon et al., 2012) and (**H**) sub-domain boundaries called from 5C data.

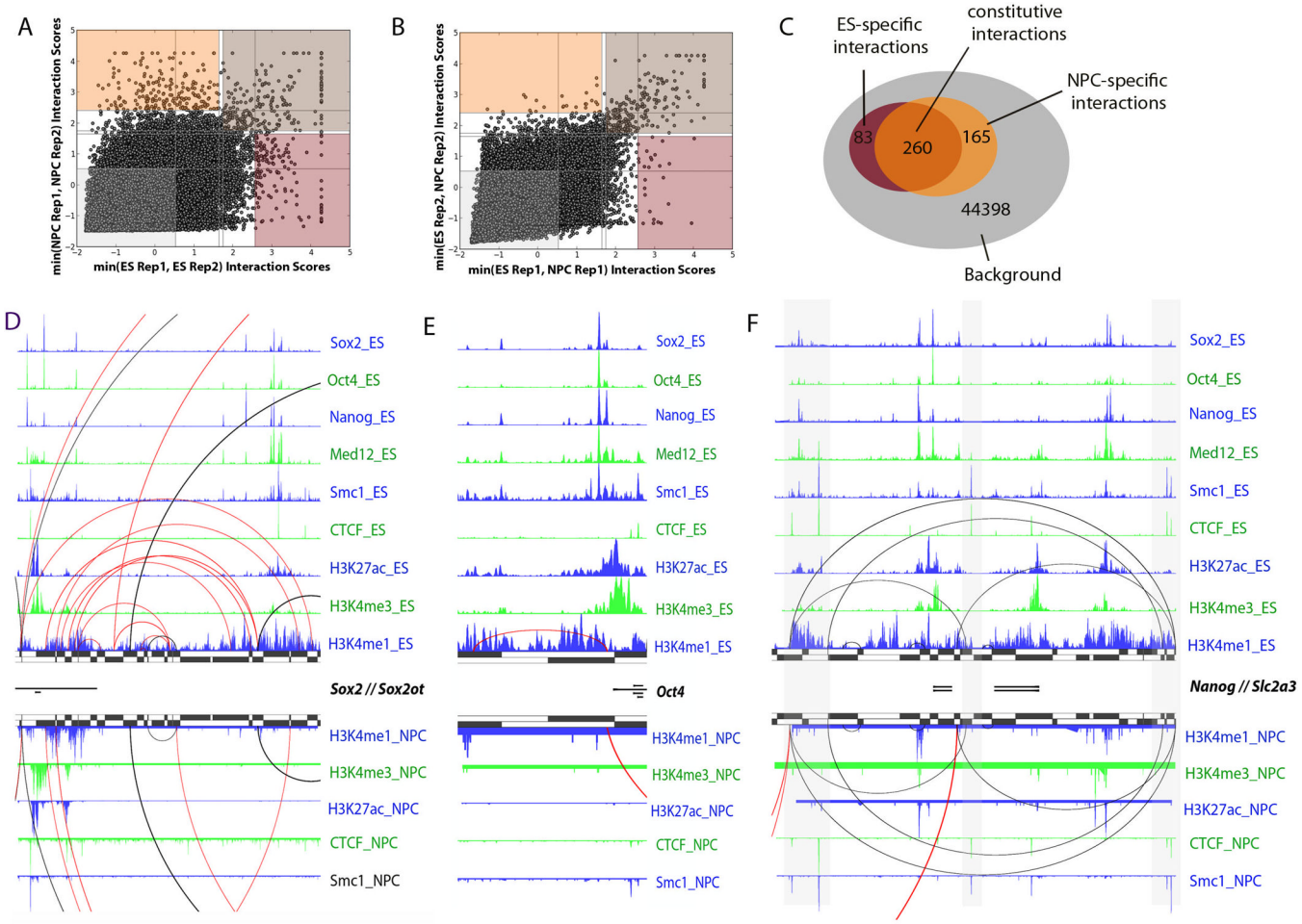

Figure 2. Genome architecture undergoes marked reorganization at the sub-Mb scale upon differentiation

(**A**) Scatterplot comparison of interaction scores between ES cells and NPCs. Thresholds for constitutive and cell type-specific looping interactions are represented as colored boxes (brown, constitutive; red, ES-specific; orange, NPC-specific; grey, background). (**B**) Scatterplot comparison of interaction scores after randomly permuting replicates. (**C**) Interactions called significant in ES cells and NPCs. (**D–F**) Chromatin interactions and epigenetic modifications at specific genomic loci in ES cells and NPCs. ChIP-seq reads are displayed for CTCF, Med12, Smc1, Sox2, Oct4, Nanog, H3K27Ac, H3K4me1, and H3K4me3 in ES cells (above gene track) and CTCF, Smc1, H3K27Ac, H3K4me1, and H3K4me3 in NPCs (below gene track). 3-D interactions are represented as mirror image arcplots for ES cells (above gene track) and NPCs (below gene track), with constitutive and cell type-specific interactions displayed in black and red, respectively. Black bars represent HindIII restriction fragments. (**D**) ES-specific interactions between *Sox2* and a putative enhancer. (**E**) ES-specific interactions between *Oct4* and a putative enhancer. (**F**) Constitutive interactions around *Nanog and Slc2a3*.

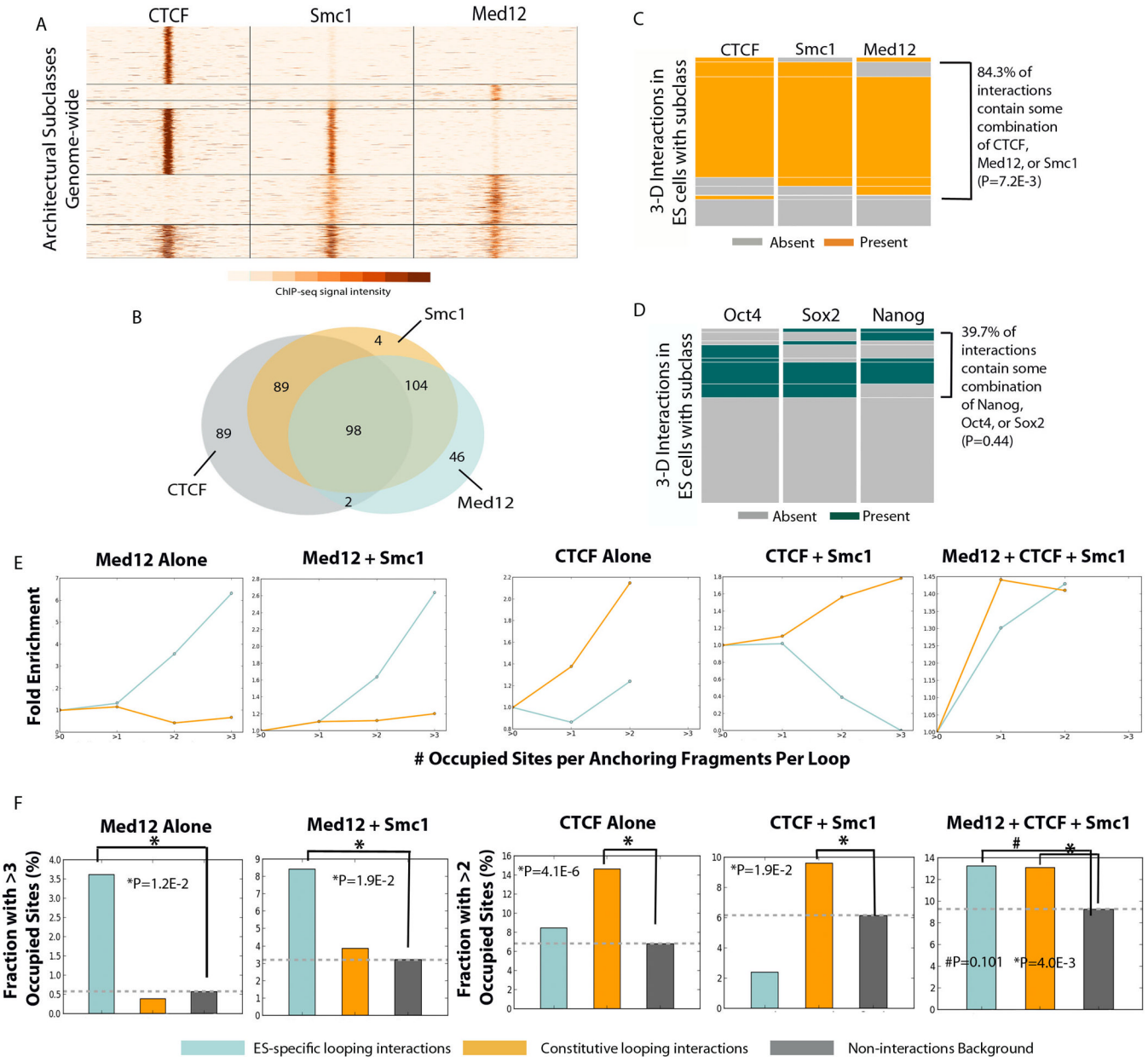

Figure 3. Architectural protein subclasses have distinct roles in genome organization

(**A**) Heatmap representation of ChIP-seq signal for seven distinct architectural protein subclasses genome-wide. (**B**) Venn diagram comparing binding patterns for high-confidence ($P<1\times10^{-8}$) CTCF, Med12, and Smc1 occupied sites in 5C regions. (**C–D**) Unsupervised hierarchical clustering for significant interactions in ES cells enriched for (**C**) CTCF, Med12, Smc1 or (**D**) Oct4, Nanog, Sox2. (**E**) Fold enrichment of architectural protein subclasses in looping interactions vs. the number of occupied sites per anchoring fragments. (**F**) Fraction of constitutive or ES-specific looping interactions enriched with architectural protein occupied sites compared to the expected enrichment in background. Fisher's Exact test: *, P<=0.05.

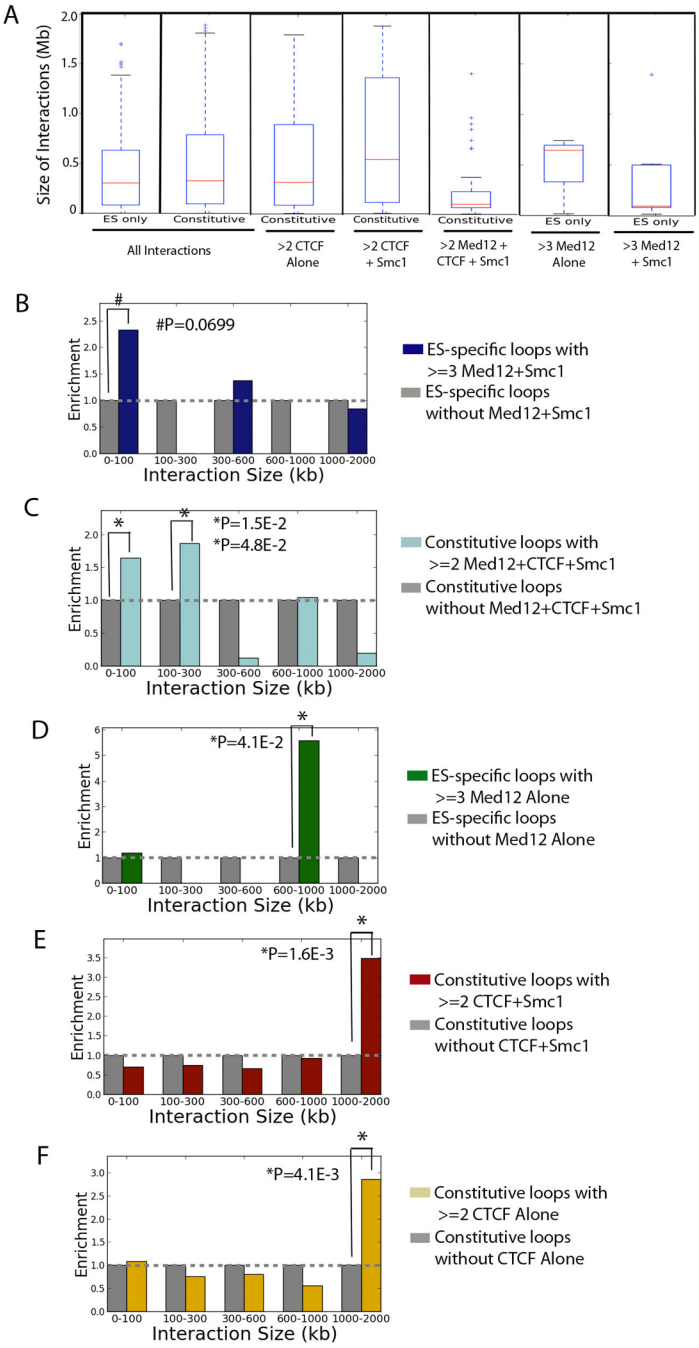

Figure 4. Architectural protein subclasses function at different length scales

(A) Size distributions of looping interactions mediated by distinct subclasses of architectural proteins. (**B–F**) Histograms binned by loop size displaying fold enrichment of interactions connected by (**B**) Med12+Smc1 (navy), (**C**) Med12+CTCF+Smc1 (light blue), (**D**) Med12 alone (green), (**E**) CTCF+Smc1 (red), or (**F**) CTCF alone (orange) compared to all interactions not containing the occupied sites in ES cells (gray). Fisher's Exact test: *, P<=0.05.

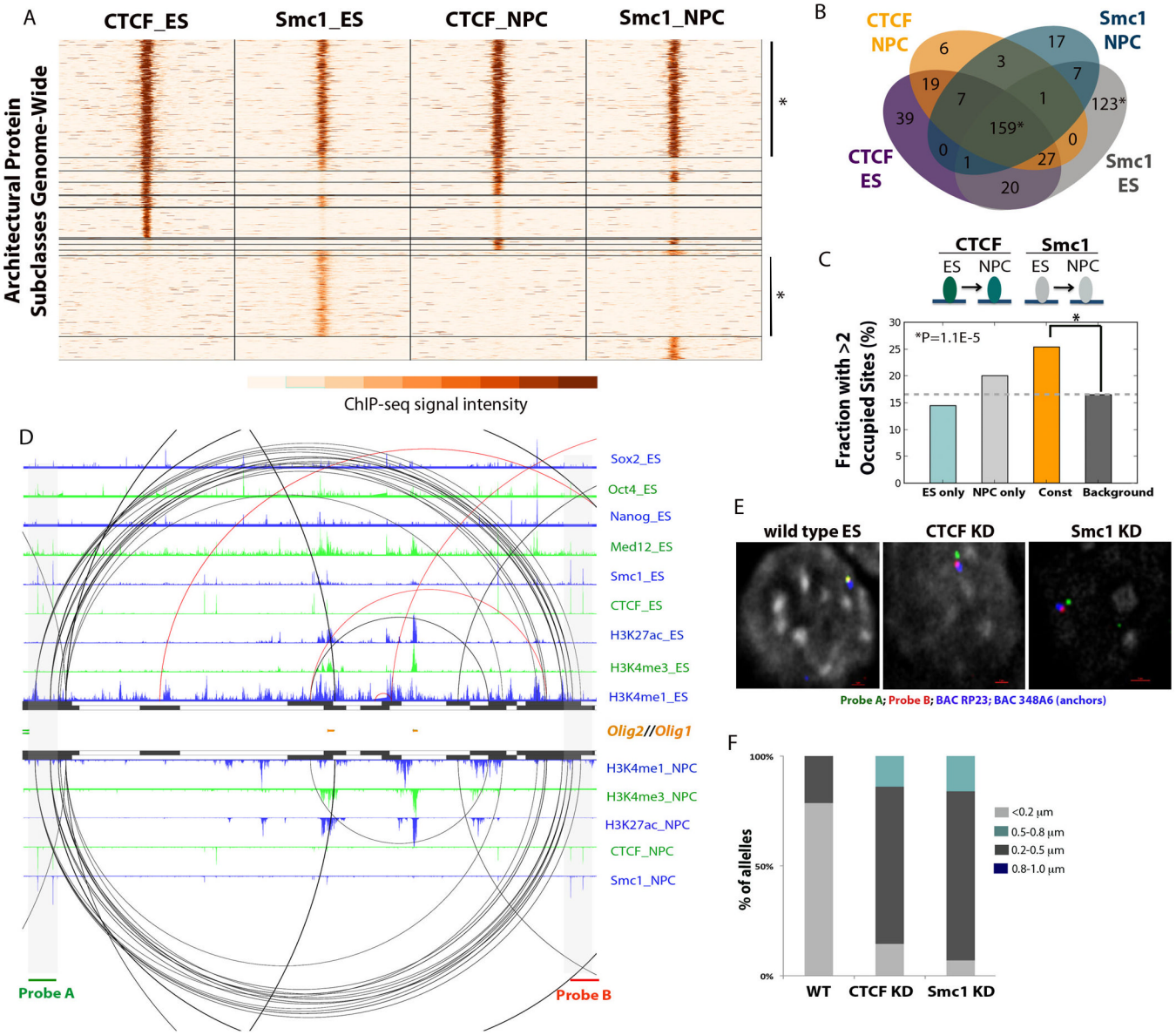

Figure 5. Constitutive looping interactions are anchored by constitutive binding of CTCF and cohesin

(**A**) Heatmap representation of distinct subclasses of architectural protein occupancy between cell types genome-wide. (**B**) Venn diagram representing unique and overlapping high-confidence ($P<1\times10^{-8}$) CTCF and Smc1 occupied sites in ES cells and ES-derived NPCs in 5C regions. (**C**) Fraction of constitutive or cell type-specific looping interactions enriched with constitutive occupancy of CTCF+Smc1 compared to the expected background enrichment. Fisher's Exact test: *, P<=0.05. (**D–F**) DNA FISH analysis of chromatin interactions connected by sites constitutively bound by CTCF+Smc1. (**D**) Arcplot of constitutive interactions anchored by constitutive CTCF+Smc1 occupied sites (black) and cell type-specific interactions anchored by ES-specific Smc1 occupied sites (red) compared to epigenetic marks around *Olig1* and *Olig2* genes. Shaded grey bars highlight genomic fragments constitutively bound by dual CTCF+Smc1 sites anchoring the base of a series of constitutive looping interactions. (**E**) Probes specific for fragment A (green) and fragment B (red) were used to perform DNA FISH in wild type V6.5 ES cells and ES cells treated with

lentiviral shRNA for CTCF or Smc1. Scale bar, 1 μm. (**F**) Quantification of spatial distances separating FISH probes (mean ± s.d.). Wild type V6.5 ES cells (0.144 ± 0.05 μm, n=126), CTCF knock-down (0.421 ± 0.21 μm, n=130), and Smc1 knock-down (0.385 ± 0.13 μm, n=113).

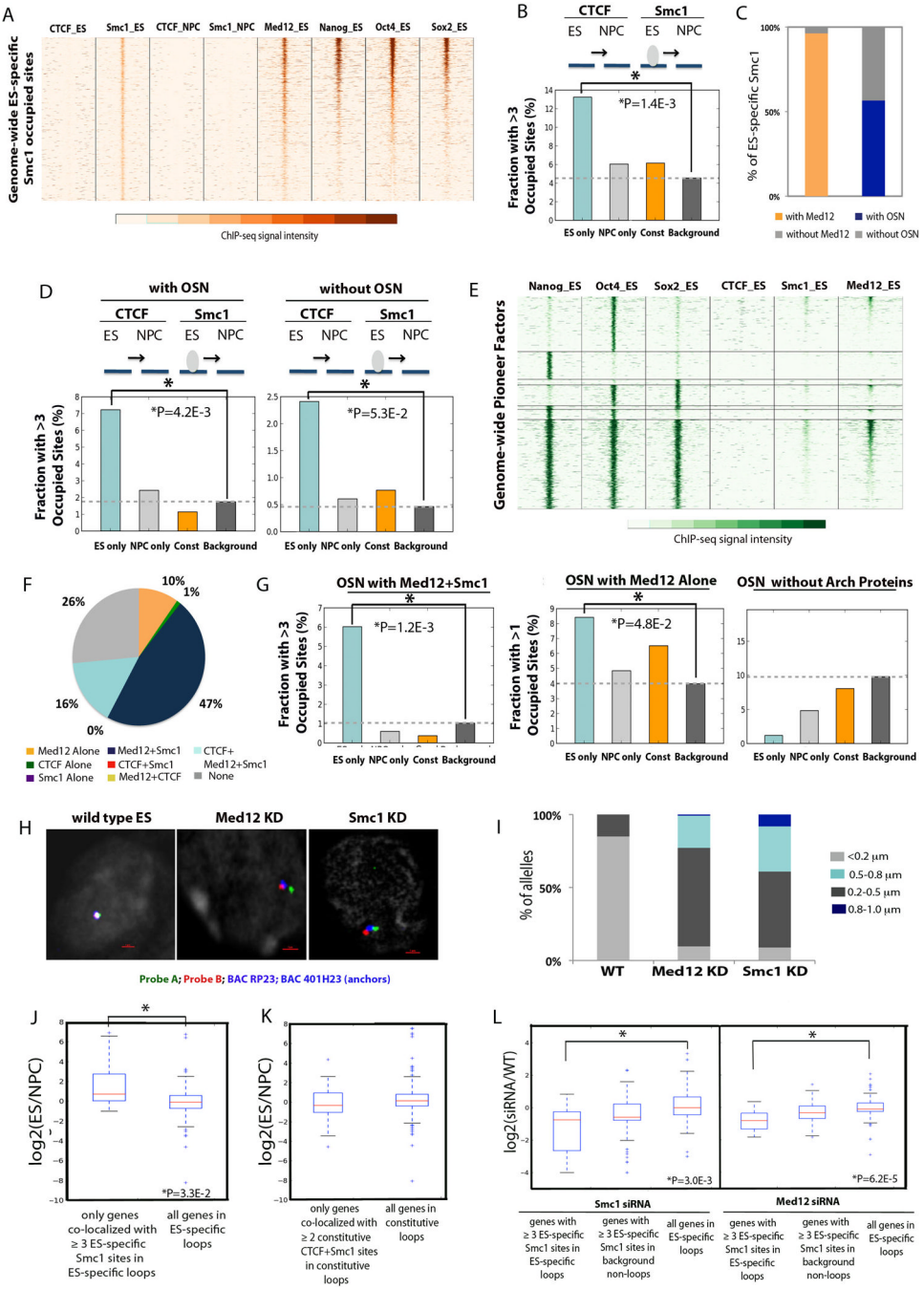

Figure 6. Mediator and cohesin bridge ES-specific enhancer-promoter interactions

(**A**) Heatmap representation of all ES-specific Smc1 occupied sites compared to Med12, Oct4, Sox2, and Nanog occupied sites genome-wide. (**B**) Fraction of constitutive or cell type-specific looping interactions enriched with ES-specific Smc1 occupied sites compared to the expected background enrichment. Fisher's Exact test: *, P<=0.05. (**C**) Fraction of ES-specific Smc1 occupied sites co-localized with Med12 or Oct4/Sox2/Nanog in 5C regions. (**D**) Fraction of constitutive or cell type-specific looping interactions enriched with ES-specific Smc1 occupied sites with or without Oct4/Sox2/Nanog compared to the expected enrichment in background. Fisher's Exact test: *, P<=0.05. (**E**) Heatmap representation of

all Oct4/Sox2/Nanog subclasses compared to architectural proteins sorted by Med12 occupancy genome-wide. (**F**) Pie chart showing percentages of Oct4/Sox2/Nanog occupied sites co-localized with architectural proteins in 5C regions (n=102). (**G**) Fraction of constitutive or cell type-specific looping interactions enriched with Oct/Sox2/Nanog with or without architectural proteins compared to the expected enrichment in background. Fisher's Exact test: *, P<=0.05. (**H**) Probes specific for fragment A (green) and fragment B (red) anchoring an ES-specific looping interaction connected by an ES-specific cohesin site were used to perform DNA FISH in wild type V6.5 ES cells and ES cells treated with shRNA for Med12 or Smc1. Scale bar, 1 μm. (**I**) Quantification of spatial distances separating FISH probes (mean ± s.d.). Wild type V6.5 ES cells (0.139 ± 0.04 μm, n=114), Med12 knockdown (0.390 ± 0.16 μm, n=123), and Smc1 knock-down (0.462 ± 0.21 μm, n=123). (**J–K**) Gene expression ratio between ES cells and NPCs for (**J**) genes in ES-specific interactions co-localized with ES-specific Smc1 compared to all genes in ES-specific interactions or (**K**) genes in constitutive interactions co-localized with constitutive CTCF +cohesin compared to all genes in constitutive interactions. (**L**) Gene expression ratio between siRNA-treatment for Med12 or Smc1 and wild type ES cells for genes in ES-specific interactions co-localized with ES-specific Smc1 compared to either all genes in ES-specific interactions or all genes co-localized with Smc1 in background non-interactions. Kolmogorov-Smirnov test: *, P<=0.05.

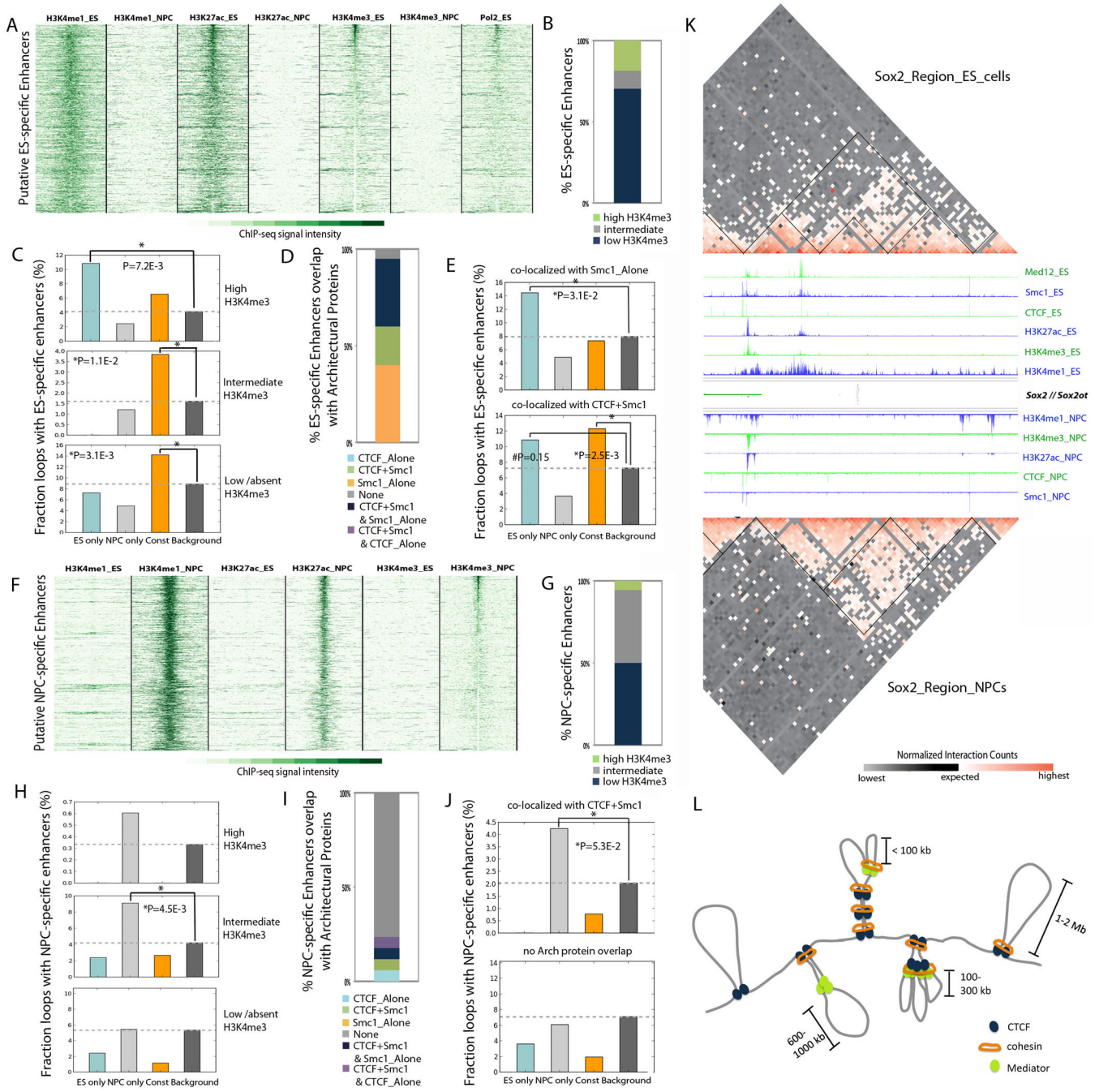

Figure 7. Architectural proteins cooperate with cell type-specific enhancers to form cell type-specific interactions

(**A**) Heatmap representation of chromatin modifications demarcating putative ES-specific enhancers genome-wide. (**B**) Fraction of ES-specific enhancers co-localized with high, intermediate, or low levels of H3K4me3 in 5C regions. (**C**) Fraction of constitutive or cell type-specific looping interactions enriched with ES-specific enhancers with high, intermediate, or low levels of H3K4me3 compared to the expected background enrichment. Fisher's Exact test: *, P<=0.05. (**D**) Fraction of ES-specific enhancers co-localized with architectural proteins in 5C regions. (**E**) Fraction of constitutive or cell type-specific looping interactions enriched with ES-specific enhancers co-localized with Smc1 Alone or CTCF

+Smc1 compared to the expected enrichment in background. Fisher's Exact test: *, P<=0.05. (**F**) Heatmap representation of chromatin modifications demarcating putative NPC-specific enhancers genome-wide. (**G**) Fraction of NPC-specific enhancers co-localized with high, intermediate, or low levels of H3K4me3 in 5C regions. (**H**) Fraction of constitutive or cell type-specific looping interactions enriched with NPC-specific enhancers co-localized with high, intermediate, or low levels of H3K4me3 compared to the expected background enrichment. Fisher's Exact test: *, P<=0.05. (**I**) Fraction of NPC-specific enhancers co-localized with architectural proteins in 5C regions. (**J**) Fraction of constitutive or cell type-specific looping interactions enriched with NPC-specific enhancers co-localized with CTCF +Smc1 or without architectural proteins compared to the expected enrichment in background. Fisher's Exact test: *, P<=0.05. (**K**) 5C interaction frequencies and epigenetic modifications at the *Sox2* locus in ES cells (above gene track) and NPCs (below gene track). (**L**) Architectural length scale model for developmentally regulated chromatin organization.