Topology-informed regulatory element collections coordinate cell identity gene expression programs

Jie Lv[1*], Kelsey A. Maher[1*], Alaguraj Veluchamy[1], Yuna Kim[1], Li Dong[1], Bensheng Ju[1], Virginia Valentine[2], Marc Valentine[2], Steven Burden[2], John Easton[1], Stanley B. Pounds[3], and Brian J Abraham[1]

[1]St. Jude Children's Research Hospital, Department of Computational Biology, Memphis, TN
[2]St. Jude Children's Research Hospital, Cytogenetics, Memphis, TN
[3]St. Jude Children's Research Hospital, Department of Biostatistics, Memphis, TN

*These authors contributed equally to this work.

ABSTRACT

Cis-regulatory elements (CREs) contact target genes to regulate their transcription, but CRE interactions are complex and predicting their function is challenging. Contemporary approaches to predict CRE targets generally assume one-to-one associations and linear genomic proximity. These methods fail to account for long-range CRE contacts caused by folding of the genome, for circumstances in which a single CRE regulates multiple genes, or, conversely, when a single gene is regulated by multiple CREs. To address these challenges and investigate three-dimensional (3D) CRE communities, we developed BOUQUET, an integrative, graph-theory-based approach that uses multiple aspects of genome topology to probe communities of CREs, their bound apparatus, and their target genes. Using this approach, we uncover communities of CREs that are undetectable by existing unbiased methods, such as those with interactions that span insulating loop boundaries, and regulatory communities that comprise clustered promoter elements. A subset of these communities, which we term "3D super-enhancer CRE communities" (3DSE CRE communities), accrues exceptional amounts of regulatory apparatus at cell identity-defining genes and evolves the super-enhancer concept. In imaging assays, we observe co-localization and co-expression of in-community genes within puncta of transcriptional co-factors, consistent with microcompartment biology. Genome-wide, pairs of genes within the same 3D CRE community are co-expressed in single cells, and pairs of CREs are co-accessible, suggesting both chromatin activity and transcription within a topologically defined community is coordinated. Our model more accurately reflects the complexity of CRE interaction networks, underscores the need for approaches that incorporate spatial organization in the study of enhancer biology, and provides a framework for the reinterpretation of SEs, their condensates, and their target genes through the lens of 3D chromatin structure.

INTRODUCTION

Gene transcription is controlled by DNA *cis*-regulatory elements (CREs), such as enhancers and promoters, but accurately identifying which CREs regulate the transcription of a given gene remains challenging despite extensive efforts (Banerji et al., 1981; Bulger & Groudine, 2011; Daniel et al., 2014; Maston et al., 2006; Shen et al., 2012; Spitz & Furlong, 2012). Historically, enhancers have been predicted to regulate their nearest neighboring gene, so proximity is often used for unbiased computational enhancer-gene assignment (Anders & Huber, 2010; Creyghton et al., 2010; Heintzman et al., 2009). While this strategy is supported by experimental evidence, its conservative approach fails

to capture the breadth of observed interactions. Enhancers have been observed to loop to and regulate distal genes, in some cases over a megabase away (Carey, 1998; Levine & Tjian, 2003; Pombo & Dillon, 2015; Shen et al., 2012; The ENCODE Project Consortium, 2012; Visel et al., 2009; Williamson et al., 2011). Further, CRE-CRE interactions defy a one-to-one interaction model, as they can interact with dozens of other CREs, and can regulate the expression of multiple genes simultaneously (Madsen et al., 2020; Talukder et al., 2021; Tsai et al., 2019; Uyehara & Apostolou, 2023). Promoters can also function as enhancers for other genes (Nguyen et al., 2016; Serfling et al., 1985). Overall, the conventional approach of proximity-based, one-to-one assignment of enhancers to genes underestimates the observed complexity of transcriptional regulation.

Activating CREs, i.e. enhancers and promoters, contribute to RNA Pol II-driven transcription by forming chromatin loops and coming into physical proximity with promoters of target genes. The contact frequencies between pairs of DNA regions are quantified with Chromosome Conformation Capture (3C) and related technologies and used to discover recurrent and regulated loop-formation (Quinodoz et al., 2018, 2022). Previous work built on 3C technologies revealed topological chromatin structures, including topologically associating domains (TADs) and sub-TAD structures. TADs reflect relatively high contact frequencies between genomic bins within the domain, and some TADs have been shown to restrict CRE interactions (Chang et al., 2023a; Dixon et al., 2012; Ji et al., 2016; Narendra et al., 2015). A subset of TADs, as well as smaller sub-TADs, are insulated neighborhoods (INs), chromatin loops formed by contacts between CTCF-bound regions whose interaction is reinforced by cohesin. INs have also been shown experimentally to constrain CRE-CRE interactions. With the topological folding of the genome, linearly distal genes and CREs can be brought into physical proximity; the interaction of neighboring regions can also be blocked by a topological boundary. It is therefore necessary to better characterize the organization of the genome to identify the specific gene or genes regulated by each CRE.

Specialized cell identities are established and maintained by the subsets of genes that are expressed in each cell (Vaquerizas et al., 2009). The expression of cell-identity-defining genes is frequently regulated by clusters of enhancers called “super-enhancers” or “stretch enhancers” (SEs) (Parker et al., 2013; Whyte et al., 2013). These clusters of heterogeneous, active enhancers correlate with elevated proximal target gene expression, and have been hypothesized to enable the tight regulation of potent cell-identity-defining genes in response to environmental cues. SEs are identified quantitatively as large, linear, high-signal clusters of markers of transcriptional enhancer activity, such as BRD4, MED1, or H3K27ac. Recently, live-cell imaging of endogenously tagged BRD4 and MED1 revealed that these transcriptional co-factors form discrete nuclear puncta encompassing at least some SE gene loci and are argued to act as transcriptional “microcompartments” (Goel et al., 2023; Sabari et al., 2018). For both BRD4 and MED1, approximately 1,000 nuclear puncta per cell were noted in mESCs, including the loci encoding *Sox2* and its associated CREs; however, in the same cell type, only 231 SEs were identified using the current gold-standard SE-calling method (Whyte et al., 2013). We hypothesized that other assemblies of CREs associated with these nuclear transcription co-factor puncta or microcompartments are overlooked by current identification approaches, particularly assemblies that are formed through complex and/or long-range CRE interactions.

Here, we propose a model using topological structures to systematically identify networks of cis-regulatory elements (CREs) responsible for gene regulation genome-wide. We developed a software

suite called Building Optimized Units of Quantified Enhancer Topologies (BOUQUET), which deploys an approach informed by graph theory to parse genome organization observations into discrete network structures we call "3D CRE communities". 11,050 3D CRE communities of diverse sizes and complexities were identified in mouse embryonic stem cells (mESCs). CRE communities contain multiple genes whose expression change in response to loss of a single community enhancer. High-signal 3D CRE communities associate with genes important for cell identity, consistent with SE biology, and recapture many linear SE-associated genes. Additionally, our approach refines many of the SE constituent enhancer assignments and identifies novel "3DSE CRE communities" in an evolution of the SE concept. CRE community CREs share co-localization, co-accessibility, and co-expression of their target genes. CRE activities are most highly coordinated when associated with co-factor nuclear puncta, consistent with the microcompartment model(Goel et al., 2023). Thus, our strategy for the systematic integration of genome topology maps, enhancers, and promoters enables the discovery of novel biological principles of transcriptional regulation and the control of specific cellular identities.

RESULTS

Identifying three-dimensional cis-regulatory element communities using topological DNA structures

Activating DNA *cis*-regulatory elements (CREs) are marked by the characteristic histone modification H3K27ac and regulate their target genes' expression through topological looping. As such, we began our investigation of three-dimensional (3D) CRE groups by examining the genomic loops involving active CREs using H3K27ac HiChIP. Here, CREs are defined as the collapsed union of high-confidence H3K27ac ChIP-seq peaks and the promoters of transcriptionally active genes (**Methods**). We developed a computational pipeline (Hi-LOW) for the rapid, reproducible, one-click analysis of Hi-C-derived data, e.g. HiChIP, that identifies high-confidence (FDR < 0.01) loops (see **Methods**, **Fig. 1A**, **Supp. Fig. 1A**). We found that, in mouse embryonic stem cells (mESCs), only 74.8% of active enhancers and 81.3% of active promoters were involved in high-confidence HiChIP loops (**Supp. Fig. 1B**), and not every loop identified connected pairs of activation-associated CREs (**Supp. Fig. 1C**). This analysis suggests that, on their own, high-confidence HiChIP loops are not sufficient to study the suite of interactions involving active CREs within a given cell type.

To explore the nearly ~25% of active CREs that are not involved in high-confidence CRE loops, we used insulating genomic loops to infer connectivity. Insulated neighborhoods (INs) have been shown experimentally to constrain CRE-CRE interactions, and their size, on the order of tens of kilobases, permits higher resolution predictions of CRE interaction targets than TADs (Dowen et al., 2014). We modernized a computational protocol for identifying INs (Dowen et al., 2014) from SMC1A ChIA-PET and CTCF ChIP-Seq data in mESCs (**Methods**, **Supp. Fig. 1D, 1E**). We identified 25,581 high-confidence SMC1A-anchored loops, of which 13,777 formed CTCF/SMC1A-anchored INs. 66.6% and 59.6% of active enhancers and promoters, respectively, fell within INs. This is an increase over the overall percentage of the mouse genome that falls within INs, 42%, suggesting that this type of genomic structure is enriched at regions containing active CREs (**Supp. Fig. 1F**). Using high-confidence H3K27ac HiChIP loops and INs in combination enabled us to assess the connectivity of 89.9% of enhancers and 91.9% of promoters (**Supp. Fig. 1B**). To capture the remaining ~10% of active CREs, we used a similar process to incorporate topologically associating domains (TADs), a related category

of putatively insulating genomic loop. TADs are on the order of several hundreds of kilobases in size, and previous work built on 3C technologies demonstrated that TADs reflect relatively high contact frequencies between genomic bins within the same TAD region (Dixon et al., 2012). The size of TADs and the size of the bins used to identify them decrease interaction resolution, but many TADs have been shown to restrict CRE interactions, and they have been used widely to make predictions on CRE connections (Dixon et al., 2012; Rao et al., 2014; Sexton et al., 2012; Ulianov et al., 2016). In the absence of direct HiChIP loops or INs, TADs can provide useful information on the potential connectivity of CREs in a given region of the genome. Altogether, our approach collectively allows for the connectivity assessment for the vast majority of active CREs across the mESC genome, covering 99.1% of enhancers and 98.6% of promoters in total (**Supp. Fig. 1B**).

Having comprehensively identified activating CREs, we identified connections between them. We considered CREs connected if they met at least one of two criteria: 1) if they were joined by a high-confidence HiChIP loop; or 2) if they were contained within the same insulated neighborhood (IN) (see **Methods**). CRE pairs that were connected by high-confidence loops showed similar HiChIP read counts between them as CRE pairs that were exclusively connected using INs, a result that is consistent with the known challenge in identifying statistically significant (i.e., recurrent and regulated) loops using Chromosome Conformation Capture (3C) techniques (**Supp. Fig. 1G**) (McCord et al., 2020). Notably, a large fraction of CRE pairs that were connected by high-confidence HiChIP loops do not fall within the same IN, seemingly violating insulating boundaries (**Supp. Fig. 1H**). While insulating boundaries have been considered as a mechanism to constrain CRE in the nucleus, recent studies have also noted CRE-CRE interactions that cross these boundaries, validating our strategy (Chang et al., 2023b). Connections between pairs of CREs are therefore intrinsically complex, include within-neighborhood and cross-neighborhood interactions, and support our integrative computational strategy.

This set of CREs and their connections were used as inputs to identify 3D CRE communities. We developed BOUQUET (Building Optimized Units of QUantified Enhancer Topologies), a community-detection algorithm that uses concepts from graph theory to convert undirected networks of CREs and their connections ("nodes" and "edges", respectively) into discrete communities of CREs associated with each expressed gene (**Methods, Fig. 1A**, **Supp. Fig. 1A**). Briefly, we developed a modified label-propagation strategy wherein highly interconnected sets of nodes are unbiasedly assembled into communities. Any additional CREs connected directly to a given promoter either by high-confidence HiChIP loops or shared INs are then incorporated into that gene's community. Using the criteria outlined above, we were able to assign 69% of enhancers and 96% of promoters to regulatory "pre-communities" of more than one node. Upon adding the final level of TAD assignment criteria, we were able to increase those community assignment rates to 92% and 98%, respectively (**Supp. Fig. 1I**).

For 23,083 active genes in mESCs we identified 11,050 3D CRE communities, with a median of 7 CREs per community (**Fig. 1B, Supp. Fig. 1J**). The expression of genes within a given community was significantly, though modestly, positively correlated with the number of CREs within its 3D community (**Fig. 1C**). The numbers of promoters and enhancers within a given 3D CRE community were positively correlated, with relatively few communities heavily biased toward containing one type of element over the other (**Supp. Fig. 1K**). 3D CRE communities showed a range of complexity. Some, such as the community for *Zfp160* and *Ppp2r1a*, were relatively simple and contained few CRE nodes

(**Fig. 1D, 1E**). Others, such as the 3D CRE community for *Sox2*, were more complex, with many CRE nodes and interactions, including some that span INs (**Fig. 1D, 1F**). *Sox2* is a master transcription factor and is responsible for the maintenance of pluripotency in ESCs (Boyer et al., 2005; Buganim et al., 2014; Takahashi & Yamanaka, 2006; S. Zhang, 2014; H. Y. Zhou et al., 2014a). Previous efforts to interrogate the cis-regulatory environment of *Sox2* revealed that is it regulated both by a gene-proximal linear SE and a gene-distal linear SE approximately 100 kb downstream called the "Sox2 Control Region" (SCR) (Li et al., 2014; H. Y. Zhou et al., 2014b). Genetic deletion of the SCR in mESCs affects only the expression of *Sox2* itself, and not the surrounding genes (H. Y. Zhou et al., 2014b). Both SEs are within the 3D CRE community we identify for *Sox2*. Our prediction that *Sox2* is the only protein-coding gene in its community is consistent with the previously observed specific effect of SCR-deletion on *Sox2*. (**Fig. 1D**). Our CRE communities thus capture the known complexity of transcriptional cis-regulation important to stem cell identity.

The 3D CRE community model accurately predicts effects of enhancer perturbation

If CRE communities contain elements that control the expression of genes throughout a community, then loss of a community CRE should affect genes within that community and not beyond it. To determine the effects of CRE perturbation across a community, we examined mESCs in which a constituent enhancer of a CRE community was deleted (**Fig. 2A, 2B, 2C**). Previously, a CRE within an SE associated with the polycistronic pri-miRNA encoding *miR290 - miR295 (D7Ertd143e*) was deleted via CRISPR-Cas9 (Hnisz et al., 2015) (a gift from the Young Lab, Acknowledgements) (**Fig. 2A**). *D7Ertd143e* is a mammalian-specific microRNA cluster that regulates embryonic stem cell pluripotency (Medeiros et al., 2011). We performed RNA-seq in deletion-containing and wild-type cells. Consistent with the previous publication, we observed significant downregulation of the *D7Ertd143e* transcript (**Fig. 2D, 2E**).

We predicted that the deletion of a constituent enhancer of the *D7Ertd143e* SE would cause downregulation of genes within the same regulatory community (marked in red) while leaving neighboring non-community genes unaffected (**Fig. 2B, 2C**). Collectively, we observe that the genes of this community are significantly more downregulated than anticipated ($p=5.8x10^{-4}$, Wilcoxon Rank Sum Test). Indeed, the expression of seven out of the eight genes in the *D7Ertd143e* community, including *D7Ertd143e* itself ($p=2.77x10^{-107}$, adjusted p-value from DESeq2)*, Myadm* ($p=6.04x10^{-146}$)*, Prkcg* ($p=2.81x10^{-2}$)*, Cacng7* ($p=2.09x10^{-14}$)*, Cacng8* ($p=3.10x10^{-11}$), *AU018091* ($p=4.68x10^{-45}$), and *Cacng6* ($p=6.47x10^{-4}$), was significantly decreased following genetic ablation of a single constituent enhancer (**Fig. 2D, Fig. 2E, Supp. Fig. 2A, Supp. Fig. 2B**). The expression of genes immediately beyond the linear boundaries of the community, including *Tfpt* (p=0.35)*, Leng1* (p=0.74), *Tmc4* (p=0.92)*,* and *Mboat7* (p=0.63), was not diminished (**Fig. 2D, 2F**, **Supp. Fig. 2A, Supp. Fig. 2C**). This is true even though the proximity of these genes to the deleted enhancer is comparable to that of other genes within the community. The substantial and specific effects on genes throughout a community caused by perturbation of a single enhancer demonstrate the degree of regulatory interconnectedness found within 3D CRE communities.

3DSE CRE communities recapitulate and expand upon linear SE assignment approaches

Several 3D CRE communities involve linear SEs and their associated genes, including *Sox2* and *D7Ertd143* (*miR-290-295*, Fig **1F, 2B**). In the classic SE model, multiple linearly clustered enhancers concentrate high amounts and varieties of transcriptional protein apparatus at genes that define cell identity. Two algorithms have been used most commonly to identify SEs: ROSE (Lovén et al., 2013; Whyte et al., 2013) and LILY (Boeva et al., 2017). Both approaches identify domains containing sets of ChIP-Seq-defined enhancer loci and rank these domains by their enhancer ChIP-seq signal quantification (**Supp. Fig. 3A**). This results in a plot quantifying co-factor MED1 signal accumulation that displays a characteristic curve where most clustered enhancers have a modest amount of signal ("typical enhancers"), and a small minority of clustered enhancers have an exceptional amount of signal ("super-enhancers (SEs)"). Consistent with linear SEs, we see a small subset of 3D CRE communities that contain exceptionally high collective signal across their constituent CREs (**Fig. 3A**). We recapitulated the SE analysis previously performed in mESCs using OCT4, SOX2, NANOG, and MED1 ChIP-Seq (**Methods**); our results are consistent with the previously published findings, identifying the cluster of microRNA genes, *D7Ertd143e*, and the master TF genes *Sox2*, *Nanog*, *Klf4*, and *Esrrb* as associated with linear SEs in mESCs (**Fig. 3B**) (Hnisz et al., 2013).

To directly compare our high-signal mESC 3D CRE communities with linear SEs, we ranked each community by the collective MED1 signal at their enhancer CREs. We observed a characteristic signal curve and similarly defined a cutoff above which 3D CRE communities contained exceptional enhancer signal (**Fig. 3C**). We termed these 3D CRE communities that surpass this threshold based only on their co-factor signal at enhancers "3DSE E-CRE communities". We identified 675, or 6%, of 3D E-CRE communities as having exceptional signal from transcriptional co-factor loading. We noted that most genes associated with a traditional linear SE were also associated with a 3DSE E-CRE community (115 out of 214), including *D7Ertd143e*, *Sox2*, *Esrrb*, and *Klf4*, reiterating that cell identity-defining genes are associated with an exceptional accumulation of transcription apparatus not only at their linearly proximal enhancers but also at linearly distal enhancers. (**Fig. 3B**, **3C, Supp. Fig. 3B**).

Examining topology data and identifying 3DSE E-CRE communities revealed that some linear SEs that were previously assigned based on linear proximity were more reasonably associated with a different gene or genes. As an example, the linear SE associated with *Rpap3* was reassigned to *Pced1b* based on direct H3K27ac HiChIP loops (**Supp. Fig. 3C**). In other cases, the linear SE retained its original gene assignment, but the genome topology data uncovered additional constituent enhancers, resulting in a shift in the signal ranking of the community (**Fig. 3D**). This situation was observed in the *H2-M5* SE community; in addition to retaining its original linear SE, multiple typical enhancers were assigned to the 3DSE E-CRE community of *H2-M5*, as well as the SE originally associated with *Olfr90* (**Supp. Fig. 3D**). The incorporation of additional CREs into a regulatory community is significant because gene expression was found to be positively correlated with overall community enhancer signal (**Fig. 3E**).

The CRE community model is consistent with known CRE biology, including the ability of promoters to regulate the expression of their own and other genes (Cho et al., 2018; Mikhaylichenko et al., 2018). To capture this phenomenon, we designed our approach to also determine the MED1 signal loaded across all CREs, i.e., both promoters and enhancers, in each community. This strategy improves upon current state-of-the-art approaches by enabling the identification of promoter-biased communities, which have also been observed to form transcriptional condensates (Dejosez et al., 2010)

similar to linear SE transcriptional condensates (Sabari et al., 2018). Using this approach, a subset of 3D CRE communities was found to be associated with exceptional MED1 signal when combining their enhancer and promoter CREs, which we call “3DSE EP-CRE communities” (**Fig. 3F**).

Promoter-rich and enhancer-rich 3DSE communities are associated with a partially overlapping set of genes, but genes specifically associated with linear, E-CRE, and EP-CRE SEs had noteworthy differences. Consistent with prior studies, the most enriched ontologies of genes associated with traditional linear SEs are relevant to cell identity (**Supp. Fig. 3E**). Regardless of whether enhancer-only (E-CRE) or both enhancer and promoter (EP-CRE) MED1 signal was considered in calling 3DSE CRE communities, stem cell-defining genes were highly enriched in the small number of 3DSE CRE communities, similar to linear SEs (**Supp. Fig. 3E**). Compared to 3DSE E-CRE communities or 3DSE CRE communities that contained a balanced ratio of enhancer and promoter signal, promoter-biased 3DSE CRE communities were rare (**Fig. 3A**). The genes associated with 3DSE EP-CRE communities tended to be less cell-identity-defining and more related to housekeeping cellular processes, as has been previously observed for promoter-driven genes in embryonic stem cells (Dejosez et al., 2010). An example promoter-rich CRE community included the histone-encoding gene loci (**Fig. 3G, 3H**). Histone genes are known to assemble into a transcriptional condensate called the histone locus body (Nizami et al., 2010). Indeed, we identified a single network component that contains the histone genes and their associated CREs. The histone genes and their CREs are assembled into four communities, suggesting the potential for an internal substructure within this complex cluster. Thus, using the model of 3DSE CRE communities, we were able to identify a new category of SE in which both promoters and enhancers contribute signal and have the potential to form nuclear bodies.

It has been hypothesized that SEs evolved to enable tight control of cell-identity-defining genes via signaling pathways (Blobel et al., 2021; Di Micco et al., 2014; Hnisz et al., 2013, 2015; Li et al., 2014; Lovén et al., 2013; Whyte et al., 2013). We therefore sought to determine if 3DSE CRE communities might also enable this utility of SE biology. We acquired ChIP-seq data from three terminal TFs of distinct signaling pathways, STAT3, SMAD3, and TCF3. We assessed whether the CREs within 3DSE CRE communities, typical 3D CRE communities, or randomized communities of CREs contained observed binding sites for zero, one, two, or all three TFs (Methods; **Fig. 3I**). We reasoned that the ability to bind TFs from a greater number of signaling pathways would permit the 3D CRE community to integrate information from the extracellular environment and regulate expression in response to a larger variety of cues. Indeed, 60% of 3DSE EP-CRE communities contained all three types of TFs, far more than typical 3D CRE communities (9%) and randomized communities (42%, Chi-squared test P-value=1.4×10^{-18}). These findings suggest that most 3DSE CRE communities contain sets of CREs likely to be responsive to multiple signaling pathways, potentially allowing them to facilitate the tight regulation of target gene expression in a manner analogous to linear SEs.

Genes within communities show co-localization and co-expression

3C-based genome-wide topological data suggest that the CREs within CRE communities interact with high frequency, so we examined whether we could quantify their physical association frequency and regulatory activities on the single-cell level using fluorescence microscopy. Recent studies have reported that linear-SE-bound cofactors, such as BRD4 and MED1, form transcriptional condensates and microcompartments, which are visible in imaging assays (Sabari et al., 2018). We

performed sequential RNA-DNA fluorescence in situ hybridization (FISH) and BRD4 immunofluorescence (IF) to track the co-localization patterns of two linearly distal components of a 3DSE CRE community, two nascent RNA transcripts from community genes, and BRD4 in the same mESCs (**Fig. 4A**). We selected the largest 3DSE CRE community, both in terms of the number of constituent CREs (**Fig. 1B**) and by cumulative MED1 signal (**Fig. 3C**, **3F**). This 3DSE CRE community spans ~1.05 linear Mb and includes *Irf2bpl*, *Esrrb*, and *Sptlc2* and their associated CREs (**Fig. 4B**). While both *Irf2bpl* and *Esrrb* have been separately associated with linear SEs, we observe that these SEs and genes are members of a larger regulatory community. H3K27ac HiChIP data suggest numerous, high-confidence interactions among the *Irf2bpl*, *Esrrb*, and *Sptlc2* loci despite spanning three non-contiguous sets of INs, a finding that is consistent with emerging data that insulating chromatin loops incompletely restrict genetic element activity (Chang et al., 2023c).

We designed FISH probes for both the DNA and the pre-spliced nascent mRNAs of *Esrrb* and *Sptlc2*, as they were the gene-pair separated by the greatest linear distance in this community and therefore maximized visual resolution (**Fig. 4B**). In the same fixed cells, we visualized the distribution of nuclear BRD4 with IF, imaged the locations of nascent *Esrrb* and *Sptlc2* mRNA using RNAFISH, digested RNAs using RNaseA, and imaged the locations of *Esrrb* and *Sptlc2* DNA using DNAFISH (**Methods, Supp. Fig. 4A**).

We performed multi-color FISH/IF in mESCs and identified 655 nuclear regions of interest (ROI) defined by the presence of puncta for one or both gene loci. In these ROIs, we visualized the frequency of co-localization between *Esrrb* and *Sptlc2* gene DNAs, their nascent mRNAs, and BRD4 co-factor puncta (**Fig. 4C**). As an expected control, mRNA production was highly correlated with presence of the gene encoding that RNA: we detected the corresponding gene in >90% of ROIs that contained mRNA transcribed from that gene (93% *Esrrb*; 95% *Sptlc2*). Conversely, a minority of ROIs that contained a gene showed mRNA production from that gene (13% *Esrrb*; 32% *Sptlc2*). Of ROIs containing both genes, 13% produce both mRNAs, 19% produce one mRNA, and 67% produce neither mRNA. In comparison, when both mRNAs were produced in the same ROI, both genes were co-localized within that ROI 87% of the time. Therefore, gene co-localization will not guarantee but is largely required for mRNA co-expression.

Coordinated CRE activity and gene expression are associated with co-factor puncta

As expected (McSwiggen et al., 2019; Sabari et al., 2018), BRD4 IF signal showed discrete nuclear puncta within a dilute nuclear background, with 49% of our ROIs overlapping BRD4 puncta (364/748). When gene co-localization and mRNA co-expression occurred in the same ROI, it was overwhelmingly also associated with the presence of BRD4 (96%, 52/54 ROIs with both mRNAs and genes detected, **Fig. 4D**). In addition, mRNA co-expression was overall positively associated with BRD4 presence in individual puncta (97%, 60/62 ROIs with BRD4 presence have both mRNAs detected). Curiously, co-localization of genes also correlated modestly with presence of BRD4, as 46% of ROIs containing neither mRNA showed BRD4 presence when the genes were co-localized, but only 20% of similarly mRNA-negative ROIs showed BRD4 presence when only a single gene was detected. These results indicate that co-expression of co-localized genes is tightly correlated with the presence of co-factor puncta (**Fig. 4D**).

When a gene was expressed in an ROI, there tended to be BRD4 co-localization with it. Of ROIs that contained a gene and its transcribed mRNA, the vast majority overlapped BRD4 puncta (84%, 69/82 ROIs for *Esrrb*, 95%, 160/168 ROIs for *Sptlc2*, **Fig. 4D**). However, of ROIs that contained a gene and overlapped BRD4, the likelihood that it also contained the corresponding mRNA was only 23% for *Esrrb* (69/298 ROIs), and no better than chance for *Sptlc2* (51%, 160/316 ROIs). This suggests that presence of BRD4 may be necessary for but not sufficient to cause transcription. These observations demonstrate that the constituents of 3DSE CRE communities associate with substantial amounts of co-factor within the nucleus. In addition, we show that these CREs are both preferentially transcriptionally active and co-localized with each other when they do so. As such, our results suggest multiple genes and regulatory elements within the same 3DSE CRE community are present and active within a single transcriptional condensate.

Coordinated CRE activity and gene expression genome-wide

While fluorescence microscopy can demonstrate association between pre-selected components, it is limited in its capacity to measure and quantify functionality genome-wide. Our results suggest frequent co-expression of *Esrrb* and *Sptlc2*; we therefore sought to determine if co-expression was a general property of the genes in 3D CRE communities. We performed single-cell RNA-seq on mESCs, and 847 cells survived our stringent quality control metrics (**Methods**). UMAP clustering of their expression profiles demonstrated high expression homogeneity, consistent with their culturing conditions (**Supp. Fig. 4B**). The observed expression homogeneity enabled us to investigate whether co-expression is linked to co-localization, and not to subpopulations of cells within a heterogeneous sample. For each pair of expressed genes (gene-pair), we calculated Spearman's correlation coefficients across cells (**Fig. 4E**). The median baseline expression correlation level for a given gene-pair not within the same 3D CRE community was r=0.01. Gene-pairs from the same 3D CRE community displayed slightly but significantly higher levels of expression correlation (median r=0.025, $p=3.48\times10^{-226}$, Wilcoxon Rank Sum Test). Expression correlation was also slightly but significantly elevated in 3DSE CRE communities (median r=0.026, $p=2.89\times10^{-95}$, Wilcoxon Rank Sum Test). We repeated the same correlation analysis using 8 other commonly used metrics and consistently found that gene-pairs from the same 3D CRE community were significantly more coordinately expressed than expected (**Supp. Fig. 4C**).

If transcriptional regulation is also coordinated among CREs of the same community, we would expect that they show correlated chromatin activity. To assess the coordinated activity of CREs within communities, we performed single-cell ATAC-Seq in mESCs, with 3,171 cells passing our quality control metrics. ATAC-seq measures one aspect of enhancer activity, i.e., chromatin accessibility to the Tn5 transposase, which correlates both with that locus' binding to transcription-regulating proteins and with gene regulation. UMAP clustering analysis of the single-cell ATAC-seq reiterated that our cell population is highly homogeneous at the chromatin level (**Supp. Fig. 4D**). In contrast to single-cell RNA-seq, there is an upper limit on the number of ATAC-seq integration sites across two copies of the same CRE. To overcome this obstacle, we developed an algorithm, Loci ATAC-Seq Enrichment Resampling or "Laser," which calculates odds-ratio-based measurements that assess co-accessibility between locus pairs across single cells. Laser demonstrated that all CRE pairs across two control chromosomes are generally positively correlated and established a background distribution of co-accessibility. In contrast, pairs of CREs within a 3D CRE community showed significantly elevated

correlation of accessibility across single cells than across background ($p=3.6x10^{-97}$, bootstrapping procedure assuming normal distribution) (**Fig. 4F**). Among the constituent enhancers of linear SEs, co-accessibility was elevated, as has been previously reported (Hnisz et al., 2015; Schuijers et al., 2018). In line with the co-accessibility profiles of linear SEs, CREs within 3DSE CRE communities also showed elevated co-accessibility. We therefore believe that constituent CREs of both typical and high-signal 3D CRE communities have highly correlated activity, consistent with what might be expected if they occupied a transcriptionally active microcompartment.

DISCUSSION

Understanding which cis-regulatory elements (CREs) control a gene's transcription is critical to decoding gene regulation. Established methods for unbiased CRE-gene assignment are limited in their ability to incorporate the roles of chromatin topology, and as such these approaches fail to capture the full complexity of CRE-gene interactions. Contemporary attempts to use individual conformational features of DNA in isolation, e.g., direct loops or TADs, also fall short by either overlooking a substantial fraction of regulatory elements or by lacking specificity in CRE-gene assignment. Here, we describe a novel computational framework, Building Optimized Units of Quantified Enhancer Topologies (BOUQUET), which incorporates multiple well-characterized categories of topological interactions to perform minimally biased, genome-wide CRE-gene assignment. BOUQUET has many advantages over previous tools for CRE-gene assignment. It optimizes connections among CREs by prioritizing higher-confidence/higher-specificity evidence of interaction (i.e. direct loops) while also leveraging information from lower-confidence/lower-specificity evidence of interaction (i.e. TADs). This strategy is necessary because high-confidence loop calls are sparse, while a complete set of loops is difficult to determine with high confidence for largely technical reasons. In this study, we uncover three-dimensional (3D) cis-regulatory element (CRE) communities in mouse embryonic stem cells (mESCs) using this computational approach, but our pipeline is portable across any species or cell type. BOUQUET was also designed to be agnostic of datatype, as parallel improvements in 3C-derived technologies offer CRE-assigning capabilities conceptually resembling those offered by H3K27ac HiChIP (McCord et al., 2020).

The coordinated transcriptional activity of co-localized DNA regions is an area of recent interest. It has been hypothesized that co-localized genes are co-expressed and that co-localized enhancers are coordinately active. Our 3D CRE communities enable genome-wide assessment of the transcriptional activities of CREs that show evidence of co-localization, including gene co-expression and co-accessibility to transposase, across individual cells. A co-localized gene pair is more co-expressed than expected in single-cell RNA-seq and fluorescence microscopy. Despite our experimental design, we were unable to confidently distinguish the enhancer CREs and promoter CREs proximal to these genes' transcription start sites. CRE transposase-accessibility is an established proxy for CRE activity, so our observation that pairs of CREs within the same community are more highly correlated in accessibility than random CRE pairs across a chromosome suggests that their activity is correlated. However, these correlations between co-localization, co-expression, and co-accessibility are modest, perhaps owing to the relatively low number of cells where these DNA loci are co-localized. This latter observation suggests that HiC derivatives are sufficiently sensitive to detect interactions that exist in the minority of cells.

We found 11,050 3D CRE communities in mESCs, each with distinctive properties. Whereas most communities have only a few CREs, some are extremely CRE-rich and, correspondingly, extremely rich in co-factor signal. Both community CRE number and community co-factor signal correlate positively with the expression of genes in that community; however, this correlation is relatively weak and suggests that many other factors contribute to expression level. We observe that gene promoters interact with CREs predicted to act as activators, silencers, and/or insulators (**Supp. Fig 1C**). This suggests that integration of these types of regulators is not only possible with a related approach, but also likely essential for studying each gene's regulation. Of note, our results showcase that the most widely used strategy of assigning single enhancers to their single most proximal gene frequently misrepresents the elements' likely target or targets, as enhancers commonly interact with multiple genes, often skipping over those that are unexpressed. The nature of 3D CRE communities further highlights that CTCF/cohesin-based insulation is incomplete, and many high-confidence CRE-CRE interactions span across the supposed boundaries of insulated neighborhoods or TADs. This observation is consistent with previous work that SE-containing TADs were co-localized in laser microdissection studies (Beagrie et al., 2017). The sophistication of our community assignment approach was further supported by the CRISPR-Cas9-mediated deletion of a constituent enhancer of the *D7Ertd143e* community, which revealed that the expression of genes within the *D7Ertd143e* community was substantially decreased following enhancer ablation, while genes outside the *D7Ertd143e* community, but linearly proximal, were unaffected.

The assembly of CREs and transcription apparatus at SE-associated genes and the histone loci along with observations that SE-bound proteins form nuclear condensates suggests that our communities also form similar nuclear condensates or microcompartments (Goel et al., 2023). Broadly speaking, membraneless organelles arise from high-valency/low-affinity interactions among transcription-regulating proteins. The unusually high concentration of apparatus at 3DSE CRE communities is consistent with their existence at or in condensates. Indeed, we observe that the largest 3DSE CRE community by co-factor signal (*Irf2bpl; Esrrb; Sptlc2;* et al.*)* co-localizes with protein puncta when multiple genes are expressed and the genes themselves are in proximity. This finding is consistent with an in-depth examination of the *Sox2* 3DSE, which found that *Sox2*, a distal enhancer locus, and transcriptional condensates all dynamically associate with one another, potentially in the recently proposed "3-way kissing model" (Du et al., 2024). We believe the 3DSE model approximately equalizes the number of observed co-factor puncta per cell and the number of CRE sets comprising them. Collectively, our microscopy observations suggest that gene expression might be coordinated in 3DSE CRE community condensates by the activities of collaborating CREs. Given the technical challenges of measuring co-localization via immunofluorescent imaging, including probe photobleaching and relocation misalignment, these counts are likely underestimates, but nonetheless provide a valuable direct comparison between the relative frequency of loci associations as measured by two commonly used orthogonal approaches. Future efforts will need to disentangle the roles of each CRE and regulatory proteins in governing gene expression at these complex networks, as well as the kinetics and frequency of each event, including condensate lifecycles.

Integrating genome topology into CRE community-identification overcomes challenges with CRE assignment and enables identification of second-generation or "3DSE CRE communities". As with linear SEs, these 3DSE CRE communities are enriched for genes critical for cell identity and function,

including *Esrrb*, *Oct4*, and *Nanog* in mESCs. Many 3DSE genes were associated with traditional linear SEs, similarly linking 3DSE CRE communities to cell-identity-control but expanding the set of CREs that control expression of cell identity genes. This finding suggests that study of mutant enhancers in cancer genomes should similarly expand if topology is considered and improve prediction of the effects of intergenic mutation. Genes that are associated with 3DSE CRE communities are enriched for binding sites for STAT3, SMAD3, TCF3, all transcription factors at the top of signaling cascades. These same factors have been shown to be enriched at genes associated with traditionally defined SEs, suggesting that 3DSE CRE communities may play a similar role in allowing cellular genetic programs to be responsive to extracellular stimuli. Ultimately, by accounting for the various layers of physical arrangement of the genome, we can more accurately reflect biological reality and offer new insight into the fundamentals of gene regulation, the nature of super-enhancers, and the mechanisms that drive cellular identity.

ACKNOWLEDGEMENTS

We thank Richard A. Young (Whitehead Institute) for providing parental mESC V6.5 and deletion (dMir290e) mESC V6.5 lines. We thank Charles H. Li, Abraham S. Weintraub, and Tong Ihn Lee for early conceptual discussions. We thank Sarah August for the editing support. This work is supported by the Transcription Collaborative of St. Jude Children's Research Hospital and the American Lebanese Syrian Associated Charities (ALSAC).

METHODS

Data availability

Data generated in this study have been deposited to the Gene Expression Omnibus (GEO) under accession number GSE254728.

Publicly available datasets used

Data Type	Accession Number
H3K27ac ChIP-seq, mESCs	GSE62380
polyA-selected RNA-seq, V6.5 mESCs	GSM1246869
SMC1A ChIA-PET, mESCs	GSE57911
CTCF ChIP-seq, mESCs	GSE30203
H3K27ac HiChIP, mESCs?	GSE99521
STAT3 ChIP-seq and input Control, mESCs	GSM686673,GSM686677
SMAD3 ChIP-seq and input Control, mESCs	GSM539542,GSM539541, GSM560357
TCF3 ChIP-seq and igG Control, mESCs	GSM307143,GSM307142, GSM979717

Inferring enhancers from H3K27ac ChIP-seq

We defined enhancers as H3K27ac ChIP-seq peaks from mESCs. H3K27ac ChIP-Seq of wild type V6.5 mouse Embryonic Stem Cells (mESCs) was downloaded from Gene Expression Omnibus

(GEO) database (GSE62380)(Ji et al., 2015). We downloaded the mouse reference genome sequence version mm10 (GRCm38) from the UCSC website, https://hgdownload.soe.ucsc.edu/goldenPath/mm10/bigZips/chromFa.tar.gz. We identified genome-wide H3K27ac ChIP-seq peaks using SEASEQ(Adetunji & Abraham, 2022).

Inferring active genes/promoters

For our reference gene annotation, we downloaded RefSeq Release 108 for GRCm38.p6 from NCBI ftp site, https://ftp.ncbi.nlm.nih.gov/refseq/M_musculus/annotation_releases/108/GCF_000001635.26_GRCm38.p6/GCF_000001635.26_GRCm38.p6_genomic.gtf.gz. To focus on high confidence gene annotations, we filtered (1) against genes annotated as "unknown_transcript," (2) against MGI-predicted genes, i.e. those named "Gm," and (3) for genes whose identifiers start with "NM" or "NR", which are curated protein-coding and non-protein-coding genes respectively, resulting in 40,618 genes for downstream analysis.

We determined the expression level of each gene by analyzing polyA-selected RNA-seq of wild type V6.5 mESCs (GSM1246869) (Hnisz et al., 2013). We used hisat2 (v2.1.0) (Kim et al., 2019) with default parameters to map RNA-seq reads to the mm10 reference genome. We used htseq-count (v0.11.2)(Anders et al., 2015) with parameters "--stranded=reverse -m intersection-strict –gene_id" to count reads mapped to any transcript of each gene. For each gene, we extracted its total non-redundant, i.e. collapsed, exon length using the merge function of bedtools v2.30.0 and used this collective exon length to calculate transcript per million (TPM) values.

We defined the promoter set for each gene as 4kb regions centered on each of its transcripts' transcription start sites (TSS). We measured the activity of each promoter by the number of H3K27ac ChIP-Seq (GSE62380) reads mapped to the 4kb promoter regions with at least one base pair overlap using bedtools intersect. Then we defined active genes/promoters as those with TPM>=1 and ranking in the top two-thirds promoter H3K27ac coverage. For expressed genes/promoters with multiple transcript isoforms, we collapsed all promoters using bedtools merge.

HiChIP/HiC Looping Organization Wrapper (Hi-LOW): identifying high-confidence HiChIP loops

<u>Step1: Initial processing of HiChIP data</u>

H3K27ac HiChIP data of wild-type V6.5 mouse Embryonic Stem Cells (mESCs) were downloaded from Gene Expression Omnibus (GEO) database (GSE99519) (Weintraub et al., 2017). We applied HiC-Pro (v 2.11.1) (Servant et al., 2015) on the pooled paired-end reads (FASTQ files) of three replicates to perform read alignment, read filtering, quality check, and contact matrix building as described with default parameters unless otherwise specified below. To achieve improved performance, we split the initial FASTQ files into multiple files, each with 10 million read pairs, which were used as input for HiC-Pro in parallel mode. HiC-Pro was configured so that each end of the read pairs was mapped independently to the mm10 reference genome using bowtie2 (v2.3.5.1) (Langmead et al., 2019). For global and local alignment, we used the options "bowtie2 --very-sensitive -L 30 --score-min L,-0.6,-0.2 --end-to-end –reorder" and "bowtie2 --very-sensitive -L 20 --score-min L,-0.6,-0.2 --end-to-end –reorder" respectively. For read mapping, we used MAPQ threshold=10 and removed multi-mapped reads (RM_MULTI=1) and duplicated read pairs (RM_DUP=1). Taking the final valid read pairs generated from HiC-Pro as input, we used the "hicpro2juicebox.sh" utility script to create a HiChIP contact map in ".hic" format.

<u>Step2: Loop calling using HiChIP data</u>

Predefined loop anchors were required for downstream loop calling steps. When a map of active enhancers is available, as the case for mESCs in this study, a combined list of active enhancers and promoters (active CREs), generated by the merge function of bedtools v2.30.0, were used as potential loop anchors.

In case an enhancer map is not available, due to lack of cell-type matching data, an alternative method is available. We use HiChIP-Peaks (v0.1.2) to call enriched peak regions from HiChIP datasets and use the resulted peak set as anchor input to FitHiChIP for loop calling.

We used FitHiChIP (v9.0) (Bhattacharyya et al., 2019) to determine statistically significant pairwise interactions/loops in HiChIP datasets. The valid HiChIP read pairs generated by HiC-Pro were used as FitHiChIP input. Briefly, the genome was divided into 5kb bins, for which read coverage was quantified. For the identification of loops, 5kb bins that overlapped with enhancers or active promoters as defined above were determined by FitHiChIP. In this study, we required each significant loop to involve an active CRE (enhancer or active promoter) bin on at least one end (Peak-To-All mode). In addition, we used a loose background estimation (FitHiChIP(L) mode) to infer expected (null) contact from HiChIP read pairs. The binomial distribution is employed in FitHiChIP on the observed contact frequencies to estimate the *P*- values, which are then corrected for multiple testing. Interactions having false discovery rate (FDR) < 0.01 were considered significant and are reported as loop calls. Subsequently, the loops were converted to bedpe format for downstream visualization and enhancer-gene association.

Insulated neighborhood calling

Step 1: ChIA-PET data processing

Raw mouse embryonic stem cell SMC1A ChIA-PET datasets (GEO accession: GSE57911) were processed using an in-house bioinformatics pipeline derived from the original report (Dowen et al., 2014). Each read-pair from each replicate of ChIA-PET data was examined for the occurrence(s) of one or both linker sequences (Linker A=CTGCTGTCCG; Linker B=CTGCTGTCAT) at any position, and the distribution of linker position in reads was examined. Consistent with the prior report and the activity of ECOP151, most of the linkers in end-reads that contained them were located at around 27 bp from the 5'-end. For end-reads with a single perfect linker match, each end-read was trimmed at the 5' position of this perfect linker sequence hit, i.e. no mismatches and no indels (cutadapt v4.3 with parameters --no-indels --pair-adapters --pair-filter any -m 27:27 -M 36:36 -n 1 -O 10 -e 0). Read-pairs comprised of trimmed end-reads between 27bp-36bp were separated into chimeric, i.e. read-pairs comprised of end-reads where an end-read has one A linker and the other has one B linker; non-chimeric, i.e. read-pairs comprised of end-reads that both contain the same A or B linker; or other read-pairs with other linker configurations. Non-chimeric reads are preferred, as they are thought to represent interactions that occurred in the cells and not during the experimental preparation.

Using Bowtie (v 1.2.2), non-chimeric read-pairs were mapped onto the mouse genome reference version mm10 in single-end mode with parameters –k 1, -v 1, -m 1 –e 70 –best –strata. For read-pairs where each end-read successfully aligned, samtools (v 1.15.1) merge (-@ 16), sort (-@ 16), and fixmate were used to re-mate and check proper pairing.

For peak-calling, the remaining, aligned read-pairs were separated into a set of end-reads and further processed. For connectivity analysis, presumed PCR duplicate read-pairs, i.e. with identical chromosome, start, and end coordinates for each end-read, were identified and de-duplicated and further processed as described below.

Surviving reads from both replicates were merged for peak-calling. Peaks were called using MACS (v1.4.1) (Y. Zhang et al., 2008) with parameters -p 1e-9, --keep-dup=2, nomodel and nolambda, and these peaks were considered candidate loop anchors. PETs thought to represent self-ligation events were removed if their end-reads were separated by less than 5kb. Connectivity between pairs of candidate loop anchors was quantified by overlapping PETs to the anchor peaks defined earlier with MACS by using bedtools (v 2.30.0) (pairToBed -f 0.75). As used previously (Dowen et al., 2014), significant interactions between two candidate loop anchors were required to meet two simultaneous cutoffs: that they are connected by >=3 PETs in the combined replicate data and that the FDR <=0.01 from a hypergeometric test (dhyper module in R, comparing the number of PETs connecting this anchor-pair against the total number of PETs connecting any anchor-pair and the number of PETs overlapping each anchor) derived from Dowen et al. The above ChIA-PET processing pipeline is available at https://github.com/stjude/Dowen_Fan.

Step 2: ChIP-Seq data processing for insulated neighborhoods

Quality control, mapping and peak calling for CTCF ChIP-seq data (GEO accession: GSE30203) in the mm10 reference genome were performed using SEASEQ (Adetunji & Abraham, 2022). Parameters for peak detection in SEASEQ are as follows: --keep-dup=1, -p 1e-09.

Step 3: Insulated Neighborhood calling

Insulated neighborhoods were defined as the significantly interacting SMC1 ChIA-PET loops whose anchors overlapped with CTCF ChIP-seq peaks and contained at least one gene. To identify the subset of SMC1 ChIA-PET loops that represent CTCF-CTCF interactions, we selected loops whose anchor regions both overlapped with CTCF ChIP-seq peaks by at least 1bp using PairToBed (Quinlan & Hall, 2010). CTCF-CTCF interactions that contained at least one gene (Refseq 108) were derived using PairToBed -ospan in bedtools, selecting loops that spanned at least one entire gene i.e, start1 of anchor1 < start of gene and end2 of anchor2 > end of gene.

Positions of insulated neighborhoods identified in mouse ESCs relative to the mm9 reference genome were obtained from the literature (Hnisz et al., 2016). INs whose anchors overlapped a promoter were removed. Remaining INs were lifted over from mm9 to mm10 using the UCSC liftover tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver) with parameter -ends=1000. For display, these IN coordinates were converted into the BEDPE format by assuming each anchor was 1kb.

Step 4: Comparison of Insulated Neighborhoods (INs) called *de novo* with lifted over INs

The overlap between *de novo* SMC1A ChIA-PET interactions and SMC1A ChIA-PET interactions lifted over from mm9 from the same data (Dowen et al., 2014; Hnisz et al., 2016) was used, resulting in 22,589 significant SMC1A ChIA-PET interactions in common between the approaches. INs are the subset of SMC1-anchored loops that overlap CTCF peaks on both anchors and contain at least one gene. This resulted in a total of 13,777 insulated neighborhoods in mESCs for mm10. See Supplementary Figure 1D.

Building Optimized Units of QUantified Enhancer Topologies (BOUQUET): Inferring structure-informed gene regulatory communities

To accurately represent and analyze interactions among CREs on a genome-wide scale, we built enhancer/promoter association networks in mESCs by combining a reference gene list, an active enhancer map, an active promoter map, H3K27ac HiChIP loops, Insulated Neighborhoods (INs), and topologically associating domains (TADs) (Dixon et al., 2012).

Step 1: Connecting pairs of CREs

Nodes of the network, i.e. CREs, were defined using ChIP-seq and reference genes. CREs were the collapsed union of ChIP-seq-defined enhancers and active gene promoters as defined above. We added an edge between two nodes if they fit either or both of two criteria: 1) both nodes overlapped with the two anchors of the same HiChIP loop, defined above; 2) an enhancer node and a promoter node were both fully contained within the smallest IN, i.e. minimal number of bps, that contains the promoter.

Step 2: Identifying communities of CREs

With the enhancer/promoter networks formatted as lists of nodes and edges, we employed the R package igraph (v.1.2.6) to infer communities. Communities are defined in graph theory as sets of nodes that are highly interconnected, i.e., many edges connect nodes within a community, and few edges connect two nodes from distinct communities. We define each gene's regulatory landscape as the sets of CREs associated with a given gene through this process of community detection. Thus, the enhancers/promoters within the same regulatory landscape of a given gene are predicted as a group of CREs that are proximal to each other in 3D nuclear space, implying their potential contributions to the given gene's regulation. We used a Label Propagation (LP) algorithm with back-filling to detect such gene regulatory landscape. Specifically, we first identified subsets of enhancers/promoters which are completely disconnected from others, which we called "components". For each component containing a given gene, every enhancer/promoter was initialized with a unique community label (an identifier). At every iteration of propagation, each enhancer/promoter updates its community label to the one that the maximum numbers of its neighbors belong to (the majority label). As labels propagate through the entire component, a single label can quickly become dominant in a densely connected group of enhancers/promoters but will have trouble crossing a sparsely connected region. As a result, densely connected groups of enhancers/promoters quickly reach a consensus on a unique label and those enhancers/promoters that end up with the same label of the given gene is considered part of the same regulatory community of the given gene. To ensure that directly linked promoters/enhancers were retained in a gene's community, we used a back filling method to rescue any directly linked promoters/enhancers that did not share the same community label of the given gene after LP process. We refer to the output of our graph-theory-informed method as "pre-communities."

Step 3: Use TADs to assign remaining CREs

Some enhancer-promoter association can be missed by the above community detection method. As a paradoxical example, some enhancers that are linearly close to promoters are likely to be missed by loops due to the limitations of HiChIP data bin resolution (5kb) and the size of detectable loops. Therefore, we developed an enhancer-promoter assignment strategy to rescue potential enhancers employing Topological Associated Domains (TADs). We rescued an unassigned enhancer by assigning it to either its upstream or downstream closest promoter if the enhancer and its target promoter are both fully contained in the same TAD. We refer to the networks of CREs formed by the combination of pre-communities and TAD-assigned CREs as "communities."

Data Visualization

We used the USCS Browser (Kent et al., 2002) (http://genome.ucsc.edu) to display various types of genome sequence data using the following Track Formats: hic for HiChIP interactions; bigInteract for HiChIP loops; bigGenePhred for gene annotation; bigwig (Kent et al., 2010) for HiChIP, ChIP-Seq, scATAC and RNA-Seq coverage; bigBed (Kent et al., 2010) for genomic location of enhancer, promoter, IN, and TAD. We used ggplot2 for box-and-whisker plots, violin plots and scatterplots; R package "Vennerable" for Venn diagrams; R package "igraph" for network diagrams; R package "Seurat" and "Signac" for scRNA-Seq and scATAC-Seq data visualization; Adobe Illustrator and BioRender.com for cartoon generation.

Gene Ontology and Gene Set Enrichment Analysis

Gene ontology (GO) analysis was performed using the PanglaoDB_Augmented_2021 database on Enrichr (Chen et al., 2013; Kuleshov et al., 2016; Xie et al., 2021) (https://maayanlab.cloud/Enrichr/). The heatmap generated from the GO analysis, the model of the microscopy results, and the model of the FISH/IF protocol was made using BioRender.com.

Cell culture for single-cell sequencing libraries

Mouse embryonic stem cell line V6.5 (a gift from the Young lab) was maintained in 2i/LIF medium at 37C with 5% CO2 in a humidified incubator. The culture medium DMEM-F12 (Gibco) was supplemented with 0.5X B27 (Gibco), 0.5X N2 (Gibco), 0.5mM L-glutamine (Gibco), 0.1 mM beta-mercaptoethanol (Sigma), 1X Pen-Strep (Gibco), 0.5X nonessential amino acids (Gibco), 1000U/ml ESGro recombinant LIF (Millipore), 1uM PD0325901 (R&D Systems), 3uM CHIR99021 (R&D Systems). Cells were cultured on vessel treated with 0.1% gelatin. For passaging, cells dissociated using Accutase (Millipore) and plated at a density of 30,000/cm2.

Single-cell RNA-seq

<u>scRNA-seq library preparation</u>

Cell cultures at the log phase of growth were collected and checked for cell number and viability using TC20 automated counter (Bio-Rad). Samples with >90 % viability were processed in order to prepare single cell suspensions and diluted into the target numbers plus the multiplet rate (1000 cells plus 0.8 % per library). Single cell RNA sequencing libraries were generated using the Chromium Next GEM Single Cell 5' Kit v2 (10x Genomics), following the manufacturer's manual. Briefly, single cell suspensions were loaded on the chromium controller to capture single cell gel beads and then used to perform in-droplet barcoding reactions followed by the amplification of full-length cDNAs. Sequencing libraries were constructed by sequential ligation reactions of kit-provided adaptor oligos and then dual indexed before enzymatically fragmenting the cDNAs. The resulting cDNA libraries were purified with solid-phase reversible immobilization SPRI beads (Beckman) and checked on a BioAnalyzer High sensitivity chip (Agilent) and a Qubit fluorometer (InVitrogen) for DNA concentrations and size distributions, respectively.

<u>scRNA-seq sequencing</u>

The libraries were sequenced on an Illumina NextSeq500 using the parameters recommended by 10x Genomics. For the run, the parameters "read1 26 base pair (bp) x i7 10 bp x i5 10 bp x read2 90 bp, 150-cycle High Output kit" were used to achieve reads depths of >50K per cell.

<u>scRNA-seq raw reads pre-processing</u>

The raw sequencing data from the Illumina sequencer was de-multiplexed and processed using the Cell Ranger v6.0.0 (10X Genomics) pipeline. FASTQ files were generated with a command 'cellranger mkfastq'. The mapping of raw reads to the mm10 mouse genome and generating count matrices of cell and unique molecular identifier (UMI) barcodes was performed with 'cellranger count'.

<u>scRNA-seq data analysis</u>

Cell Ranger v6.0.0 (10X Genomics) was used to process raw scRNA-seq sequencing data. Briefly, after converting Illumina raw basecall files (BCL) to FASTQ format, we used this pipeline to align sequencing reads to the mm10 transcriptome using the STAR aligner (Dobin et al., 2013). The expression of each gene in each cell was quantified as the number of Unique Molecular Identifiers

(UMI) for each gene per cell-specific cellular barcode. With ~61.8 million reads sequenced from mESCs, we obtained data from 849 cells that passed quality control steps implemented in Cell Ranger. In total, 67.2% of the total reads confidently mapped to the transcriptome. Reads that were confidently mapped to the transcriptome were used to generate gene-barcode matrices based on Chromium cellular barcodes. Filtered gene-barcode matrices containing cellular barcodes only were used for downstream analysis.

We carried out analyses of processed scRNA-seq data in R v4.0.2 with Seurat v4.1.0. As a further step of quality control, we filtered out genes expressed in <3 cells and cells with <200 or >8,000 detected genes. Additionally, cells with the percentage of mitochondrial genes >5% were filtered out. At the end, 847 cells survived this stringent quality control process. To normalize the raw read counts from the gene-barcode matrices, we employed a global-scaling normalization method "LogNormalize" that normalizes the gene expression measurements for each cell by the total expression, multiplies this by a scale factor of 10000, and log-transforms the result.

To examine transcriptome heterogeneity and find distinct cell clusters, we performed principal component analysis (PCA) to reduce data dimensionality using RunPCA command of Seurat. We first identified genes with highly variable expression across single cells using "VST" method. The normalized single-cell expression values of the top 2000 most variable genes were used. We selected the top 30 significant principal components using a permutation-based test and heuristic methods implemented in Seurat. The selected PCA loadings were used as input for graph-based cell clustering and as input for Uniform Manifold Approximation and Projection (UMAP) for reduction to two dimensions for visualization purposes.

scRNA-seq based gene-expression correlation

To measure pairwise gene expression association, we calculated Spearman correlation values of gene expression value across the 847 high-quality cells from scRNA-seq data. Before correlation calculation, raw read counts of genes were normalized using the "LogNormalize" method implemented in Seurat.

Due to the abundance of expression dropouts and overdispersion, the optimal measure of gene expression association based on scRNA-seq data is unclear (Kharchenko et al., 2014). Therefore, we calculated another eight measures of gene expression associated using Dismay R package (Skinnider et al., 2019): Pearson correlation, Kendall rank correlation, Weighted rank correlation, Biweight midcorrelation, Canberra distance, Jaccard index, Cosine similarity, Dice coefficient.

Single-cell ATAC-seq

scATAC-seq library preparation

Single cell ATAC sequencing libraries were generated using the Chromium Next GEM Single Cell ATAC kit v2 (10x Genomics), following the manufacturer's manual. Briefly, nuclei suspensions were prepared from single cells diluted in nuclei buffer. The transposition reaction of the nuclei was performed in the mixture of ATAC buffer and enzyme using a 30-minute incubation time. The transposed nuclei and barcoding reagents were loaded on the chromium controller to capture single nucleus gel beads and then to perform in-droplet barcoding PCR. Sequencing libraries were constructed by the index PCR of the barcoded free DNA fragments. The resulting ATAC libraries were purified with solid-phase reversible immobilization SPRI beads (Beckman) and checked on a BioAnalyzer High sensitivity chip (Agilent) and a Qubit fluorometer (InVitrogen) for DNA concentrations and size distributions, respectively.

scATAC-seq Sequencing

The libraries were sequenced on an Illumina NextSeq500 using the parameters recommended by 10x Genomics. For the run, the parameters “read1 50 base pair (bp) x i7 8 bp x i5 16 bp x read2 60 bp, 150-cycle High Output kit” were used to achieve reads depths of >50K per nuclei.

scATAC-seq raw reads pre-processing

The raw sequencing data from the Illumina sequencer was de-multiplexed and processed using Cell Ranger ATAC v2.1 (10X Genomics) pipeline. FASTQ files were generated with the command ‘cellranger-atac mkfastq’. The mapping of raw reads to the mm10 mouse genome and generating count matrices of cell and unique molecular identifier (UMI) barcodes was performed with ‘cellranger-atac count’.

scATAC-seq data analysis

We used the Cell Ranger ATAC pipeline v2.1.0 (10x Genomics) with the default parameters to analyze raw reads from mESC scATAC-seq library including reads filtering and alignment, barcode counting, identification of transposase cut sites, peak and cell calling, and count matrix generation. Reads are aligned to mm10 mouse reference genome (version 2020-A-2.0.0) provided by 10X Genomics. The resulting peak-by-cell count matrices, fragment file and single cell metadata were further processed using Signac R package v1.11.0 (Stuart et al., 2021) under R v4.3.1. To filter out low quality cells, we computed the following quality control metrics using Signac: 1) number of reads in peaks, 2) percentage of reads in peaks, 3) transcription start site (TSS) enrichment score (TSSEnrichment()), and 4) nucleosome signal (NucleosomeSignal()). Cells were filtered based on the following metrics: number of reads in peaks within the range of 3000-60000; percentage of reads in peaks > 40%; nucleosome signal < 3; TSS enrichment score > 2. We normalized peak counts by term frequency–inverse document frequency (TF-IDF) method using the Signac functions RunTFIDF(). We kept all features for downstream dimensional deduction by using FindTopFeatures with min.cutoff = ‘q0’. To reduce the dimensionality of data, we performed singular value decomposition (SVD) on the TD-IDF matrix using function RunSVD(). The resulted top latent semantic indexing (LSI) component was discarded as it correlated strongly with sequencing depth, and components 2-30 were used for downstream analysis, including further dimensionality reduction by uniform manifold approximation (UMAP) for data visualization (RunUMAP()) and neighborhood graph-based clustering (FindNeighbors() and FindClusters)

CRE accessibility correlation

To calculate CRE accessibility correlation, we created a customized CRE-by-cell matrix. We quantified the accessibility of each CRE in each cell using function FeatureMatrix() with the fragment file. CRE read counts were normalized by 'LogNormalize’ method using the function NormalizeData().

To measure correlation of pairwise CRE accessibility, we binarized (set CRE read counts to 0 or 1) CRE-by-cell matrix and calculated Odd Ratio (OR) using the 2 x 2 read count table for each pair of CREs. To test the statistical significance of CRE ORs between different groups we used a bootstrapping technic to generate simulated CRE-by-cell matrix by randomly resampling cells with replacement and computed a statistic from each simulated matrix: the difference of median ORs between two comparing groups, e.g., community vs background. Since this statistic largely follow a binomial distribution, we calculated a P-value when assuming the non-hypotheses that there is no difference of OR between the two comparing groups. We developed a method, Loci ATAC-Seq Enrichment Resampling (LASER), to perform the process described above. The LASER method is estimated to run in time O(n2) (with n be the number of total CREs), making the method computationally intractable for calculating all genome wide CRE pairs at once. We devised a strategy in which we split the genome into pairs of chromosomes (190 pairs) and called LASER on all pairs of

chromosomes in parallel. This method significantly cut down on the runtime and made it feasible to calculate pairwise CRE accessibility correlation genome wide.

Enhancer deletion analysis

Cell culture

A parental mouse Embryonic Stem Cell line (mESC) V6.5 and super-enhancer deleted line V6.5_dMir290e are gifts provided by Dr. Young at the Whitehead Institute (Hnisz et al., 2015). V6.5 cells are male cells derived from a C57BL/6(F) x 129/sv(M) cross. All cell lines were cultured in 2i + LIF medium as described in the previous articles (Weintraub et al., 2017). Briefly, 967.5 mL DMEM/F12 (GIBCO 11320) was supplemented with 5 mL N2 supplement (GIBCO 17502048), 10 mL B27 supplement (GIBCO 17504044), 0.5 mM L-glutamine (GIBCO 25030), 0.5X non-essential amino acids (GIBCO 11140), 100 U/mL Penicillin-Streptomycin (GIBCO 15140), 0.1 mM b-mercaptoethanol (Sigma), 1 uM PD0325901 (Stemgent 04- 0006), 3 uM CHIR99021 (Stemgent 04-0004), and 1000 U/mL recombinant LIF (ESGRO ESG1107). Cells were negative for mycoplasma tested through the beginning to the end of culture, using the MycoAlert Mycoplasma Detection Kit (Lonza LT07-318).

RNA-seq

RNA isolation and library preparation. RNA samples were isolated using AllPrep DNA/RNA Mini Kit (QIAGEN 80204). RNA libraries were prepared using KAPA RNA HyperPrep Kit with RiboErase (HRM) (Roche KK8651). Briefly, 1 µg of RNA of each sample was mixed with 2 µl of ERCC ExFold spike-in RNA mixes based on the RNA ratio recommended by the protocol (Invitrogen 4456739), rRNA were depleted by hybridization of complementary DNA oligonucleotides and treatment with RNase and DNase. RNA samples were fragmented, and double strand cDNA were generated by two steps of cDNA synthesis. Illumina TruSeq DNA UD Indexes (IDT 20040870) were used to make libraries. Libraries were amplified using KAPA HiFi HotStart ReadyMix (2X) and Library Amplification Primer Mix (10X), followed by Qubit dsDNA quantification (Invitrogen Q32853), and 2100 BioAnalyzer D1000 kit (Agilent 50671504) and Illumina MiSeq Nano v2 QC.

RNA/DNA FISH and BRD4 Immunofluorescence

Probe design, probe labeling, and antibodies used

Purified mouse BAC DNAs Sptlc2 (RP23-460P20 / 12D2 / chr12:87239078-87411743 GRCm39) and Sptlc2 (RP23-285E19 / 12D2 / chr12:87395112-87476186 GRCm39) were labeled with a red-dUTP (AF594, Molecular Probes) by nick translation. Purified mouse BAC DNA *Esrrb* (RP24-95M20 / 12D2 / chr12:86340459-86543664 GRCm39) was labeled with a green-dUTP (SeeBright 496, Enzo) by nick translation. All probes were used in the RNA and DNA assays. BRD4 was tagged using AF647 Anti-Brd4 antibody (Abcam).

Cell culture, slide preparation, and RNA probe hybridization

250,000 single-cell mouse ESC cells V6.5 (4-5-23) were applied to glass slides by cytocentrifugation. Slides were fixed in 1% PFA in PBS for 5 minutes followed by 1% PFA in PBS plus 0.05% Igepal for 5 minutes, and stored in 70% ETOH at -20°C. The slides were rinsed in PBS and blocked in 200µl blocking buffer (2X SSC and 1% BSA) and then incubated with an AF647 Anti-Brd4 antibody (Abcam) diluted at 1:100 in antibody diluent (2X SSC and 1% BSA) for 45 minutes at RT in the dark. Slides were then washed in PBS, placed in 4% PFA for 2 minutes at RT, rinsed in PBS, and dehydrated in 70%, 80%, 100% ETOH for 2 minutes each, and then hybridized overnight at 37°C for RNA using denatured probes labeled with AF594 and Seebright 496 in a solution containing sheared

mouse cot DNA with 50% formamide, 2X SSC, and 10% dextran sulfate. The slides were then washed in 50% formamide and 2X SSC and mounted in Vectashield containing DAPI and imaged for IF and RNA.

RNA digestion, DNA probe hybridization, and imaging

Slides were then washed in PBS and digested with RNaseA for 45 minutes at 37°C. Cross linking with 4% PFA for 8 minutes was then performed to stabilize antibodies on proteins that were added in the first step, followed by treatment in 0.2N HCL for 10 minutes on ice. Slides were then denatured in 70% formamide 2X SSC at 80°C for 10 minutes followed by ETOH dehydration series. Denatured DNA probes were then added in a solution containing mouse sheared cot DNA, 50% formamide, 2X SSC, and 10% dextran and hybridized at 37 °C overnight. Slides were then washed in 50% formamide, 2X SSC at 37 °C and mounted in Vectashield containing DAPI and imaged again using the same coordinates as had been imaged in the first set of IF+RNA images.

Image analysis and scoring

The RNA and DNA image fields were coordinate-matched and then examined side by side. The images are scored for the presence or absence of RNA, then whether the DNAs are interacting, and then how this relates to the AF647 Anti-Brd4 antibody immunofluorescence. Imaging was performed in 3D using a widefield fluorescence microscope equipped with Nikon Nis Elements version 4.2 software and Autoquant #D deconvolution. Images were acquired using 0.15-micron plane spacing and a 60X planapochromatic objective with a numerical aperture of 1.4. FIJI analysis was conducted to map and quantify fluorescence intensity in Regions of Interest (ROI). All acquired images underwent deconvolution using Nikon's imaging software and were then loaded into FIJI, where a max projection image was generated.

The process of ROI identification and mapping involved marking the first ROI (typically in the lower-left corner) by drawing a circle around the DNA punctum and labeling it as ROI #1. Subsequently, this ROI was copied and pasted sequentially across the image, with all generated ROI coordinates saved and specifically labeled for DNA to avoid confusion with subsequent RNA ROIs. Lastly, two ROI were marked near a cell but distinctly outside of a cell to measure background fluorescence intensity.

Due to potential misalignment between RNA and DNA images, the XY frame shift was identified by overlaying the DNA image with the RNA image and recording the necessary pixel shift for alignment. The DNA ROI coordinates were translated in FIJI using appropriate translation values, and the translated RNA ROI coordinates were saved accordingly.

DNA fluorescent intensities were measured by splitting the image using the Image>color>split channels function. The green and red channels were individually measured by applying the DNA ROI coordinates to each channel and utilizing the measure function in the ROI manager, providing pixel intensity values for each channel within the corresponding DNA-ROI locations.

RNA fluorescent intensities were measured by splitting the image using the Image>color>split channels function. Magenta (IF), green, and red channels were individually measured by applying the RNA ROI to each channel and using the measure function in the ROI manager, yielding pixel intensity values for each channel in the corresponding RNA-ROI locations.

To quantitatively assess the presence of single color or colocalized fluorescent signals within ROIs, cutoff values were assigned to each channel and then applied across all DNA and RNA images. Counts of green, red, and magenta (for RNA only) pixel intensities above threshold were determined to be positive signals DNA and RNA ROI. A composite scoring matrix was constructed, incorporating

the DNA and RNA pixel quantification results, which was used to evaluate various combinations of fluorescent signals.

For improved identification and viewing of RNA/DNA colocalized ROIs across the entire image, adjustments in DNA brightness and contrast were made using the image>adjust>brightness and contrast function. ROI-dependent images were then cropped in the Z-axis around the puncta maximum intensities to produce a 10-slice Z-stack for individual channel and composite images, which were subsequently captured. After image processing, red fluorescent channels were false-colored to magenta and magenta fluorescent channels were false-colored to cyan, to accommodate red-green colorblind audiences.

Computational analysis of ChIP-seq data

Published ChIP-seq datasets from NCBI GEO for H3K27ac (GSM1526287), H3K4me3 (GSM1526288), CTCF (GSM747534), H3K27me3 (GSM1526291), TCF3 (GSM307143, GSM307142), STAT3 (GSM686673) and SMAD3 (GSM539542, GSM539541) in mESCs were processed for finding corresponding enriched regions. Quality control, mapping, and peak calling were performed with the SEA-seq pipeline (v3.0) (Adetunji & Abraham, 2022). Reads were mapped to the mouse reference genome sequence version mm10 (GRCm38). To identify narrow enriched regions (peaks) for H3K27ac, CTCF and H3K4me3, MACS (v1.4.2) as utilized through SEA-seq with the following parameters: keep-dup= auto, band width = 300, model fold = 10,30, p-value cutoff = 1.00e-09. Visualization and analysis of the coverage and peaks was performed using ProteinPaint (X. Zhou et al., 2016). To identify significantly enriched regions of H3K27me3, we used broad peak finding within SEA-seq. Broad peak finding in SEA-seq utilizes SICER (v1.1) with the following parameters: W=200; G=400 and FDR=0.01.

References

Adetunji, M. O., & Abraham, B. J. (2022). SEAseq: a portable and cloud-based chromatin occupancy analysis suite. *BMC Bioinformatics*, *23*(1), 77. https://doi.org/10.1186/s12859-022-04588-z

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*(10), R106. https://doi.org/10.1186/gb-2010-11-10-r106

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Banerji, J., Rusconi, S., & Schaffner, W. (1981). Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, *27*(2), 299–308. https://doi.org/10.1016/0092-8674(81)90413-X

Beagrie, R. A., Scialdone, A., Schueler, M., Kraemer, D. C. A., Chotalia, M., Xie, S. Q., Barbieri, M., de Santiago, I., Lavitas, L.-M., Branco, M. R., Fraser, J., Dostie, J., Game, L., Dillon, N., Edwards, P. A. W., Nicodemi, M., & Pombo, A. (2017). Complex multi-enhancer contacts captured by genome architecture mapping. *Nature*, *543*(7646), 519–524. https://doi.org/10.1038/nature21411

Bhattacharyya, S., Chandra, V., Vijayanand, P., & Ay, F. (2019). Identification of significant chromatin contacts from HiChIP data by FitHiChIP. *Nature Communications*, *10*(1), 4221. https://doi.org/10.1038/s41467-019-11950-y

Blobel, G. A., Higgs, D. R., Mitchell, J. A., Notani, D., & Young, R. A. (2021). Testing the super-enhancer concept. *Nature Reviews Genetics*, *22*(12), 749–755. https://doi.org/10.1038/s41576-021-00398-w

Boeva, V., Louis-Brennetot, C., Peltier, A., Durand, S., Pierre-Eugène, C., Raynal, V., Etchevers, H. C., Thomas, S., Lermine, A., Daudigeos-Dubus, E., Geoerger, B., Orth, M. F., Grünewald, T. G. P., Diaz, E., Ducos, B., Surdez, D., Carcaboso, A. M., Medvedeva, I., Deller, T., … Janoueix-Lerosey, I. (2017). Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nature Genetics*, *49*(9), 1408–1413. https://doi.org/10.1038/ng.3921

Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., Gifford, D. K., Melton, D. A., Jaenisch, R., & Young, R. A. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, *122*(6), 947–956. https://doi.org/10.1016/j.cell.2005.08.020

Buganim, Y., Markoulaki, S., van Wietmarschen, N., Hoke, H., Wu, T., Ganz, K., Akhtar-Zaidi, B., He, Y., Abraham, B. J., Porubsky, D., Kulenkampff, E., Faddah, D. A., Shi, L., Gao, Q., Sarkar, S., Cohen, M., Goldmann, J., Nery, J. R., Schultz, M. D., … Jaenisch, R. (2014). The developmental potential of iPSCs is greatly influenced by reprogramming factor selection. *Cell Stem Cell*, *15*(3), 295–309. https://doi.org/10.1016/j.stem.2014.07.003

Bulger, M., & Groudine, M. (2011). Functional and Mechanistic Diversity of Distal Transcription Enhancers. *Cell*, *144*(3), 327–339. https://doi.org/10.1016/j.cell.2011.01.024

Carey, M. (1998). The Enhanceosome and Transcriptional Synergy. *Cell*, *92*(1), 5–8. https://doi.org/10.1016/S0092-8674(00)80893-4

Chang, L.-H., Ghosh, S., Papale, A., Luppino, J. M., Miranda, M., Piras, V., Degrouard, J., Edouard, J., Poncelet, M., Lecouvreur, N., Bloyer, S., Leforestier, A., Joyce, E. F., Holcman, D., & Noordermeer, D. (2023a). Multi-feature clustering of CTCF binding creates robustness for loop extrusion blocking and Topologically Associating Domain boundaries. *Nature Communications*, *14*(1), 5615. https://doi.org/10.1038/s41467-023-41265-y

Chang, L.-H., Ghosh, S., Papale, A., Luppino, J. M., Miranda, M., Piras, V., Degrouard, J., Edouard, J., Poncelet, M., Lecouvreur, N., Bloyer, S., Leforestier, A., Joyce, E. F., Holcman, D., & Noordermeer, D. (2023b). Multi-feature clustering of CTCF binding creates robustness for loop extrusion blocking and Topologically Associating Domain boundaries. *Nature Communications*, *14*(1), 5615. https://doi.org/10.1038/s41467-023-41265-y

Chang, L.-H., Ghosh, S., Papale, A., Luppino, J. M., Miranda, M., Piras, V., Degrouard, J., Edouard, J., Poncelet, M., Lecouvreur, N., Bloyer, S., Leforestier, A., Joyce, E. F., Holcman, D., & Noordermeer, D. (2023c). Multi-feature clustering of CTCF binding creates robustness for loop extrusion blocking and Topologically Associating Domain boundaries. *Nature Communications*, *14*(1), 5615. https://doi.org/10.1038/s41467-023-41265-y

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., & Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, *14*(1), 128. https://doi.org/10.1186/1471-2105-14-128

Cho, S. W., Xu, J., Sun, R., Mumbach, M. R., Carter, A. C., Chen, Y. G., Yost, K. E., Kim, J., He, J., Nevins, S. A., Chin, S.-F., Caldas, C., Liu, S. J., Horlbeck, M. A., Lim, D. A., Weissman, J. S., Curtis, C., & Chang, H. Y. (2018). Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. *Cell*, *173*(6), 1398-1412.e22. https://doi.org/10.1016/j.cell.2018.03.068

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A., & Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences*, *107*(50), 21931–21936. https://doi.org/10.1073/pnas.1016071107

Daniel, B., Nagy, G., & Nagy, L. (2014). The intriguing complexities of mammalian gene regulation: how to link enhancers to regulated genes. Are we there yet? *FEBS Letters*, *588*(15), 2379–2391. https://doi.org/10.1016/j.febslet.2014.05.041

Dejosez, M., Levine, S. S., Frampton, G. M., Whyte, W. A., Stratton, S. A., Barton, M. C., Gunaratne, P. H., Young, R. A., & Zwaka, T. P. (2010). Ronin/Hcf-1 binds to a hyperconserved enhancer element and regulates genes involved in the growth of embryonic stem cells. *Genes & Development*, *24*(14), 1479–1484. https://doi.org/10.1101/gad.1935210

Di Micco, R., Fontanals-Cirera, B., Low, V., Ntziachristos, P., Yuen, S. K., Lovell, C. D., Dolgalev, I., Yonekubo, Y., Zhang, G., Rusinova, E., Gerona-Navarro, G., Cañamero, M., Ohlmeyer, M., Aifantis, I., Zhou, M.-M., Tsirigos, A., & Hernando, E. (2014). Control of embryonic stem cell identity by BRD4-dependent transcriptional elongation of super-enhancer-associated pluripotency genes. *Cell Reports*, *9*(1), 234–247. https://doi.org/10.1016/j.celrep.2014.08.055

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, *485*(7398), 376–380. https://doi.org/10.1038/nature11082

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dowen, J. M., Fan, Z. P., Hnisz, D., Ren, G., Abraham, B. J., Zhang, L. N., Weintraub, A. S., Schuijers, J., Lee, T. I., Zhao, K., & Young, R. A. (2014). Control of Cell Identity Genes Occurs in Insulated Neighborhoods in Mammalian Chromosomes. *Cell*, *159*(2), 374–387. https://doi.org/10.1016/j.cell.2014.09.030

Du, M., Stitzinger, S. H., Spille, J.-H., Cho, W.-K., Lee, C., Hijaz, M., Quintana, A., & Cissé, I. I. (2024). Direct observation of a condensate effect on super-enhancer controlled gene bursting. *Cell*. https://doi.org/10.1016/j.cell.2023.12.005

Goel, V. Y., Huseyin, M. K., & Hansen, A. S. (2023). Region Capture Micro-C reveals coalescence of enhancers and promoters into nested microcompartments. *Nature Genetics*, *55*(6), 1048–1056. https://doi.org/10.1038/s41588-023-01391-1

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W., Ching, K. A., Antosiewicz-Bourget, J. E., Liu, H., Zhang, X., Green, R. D., Lobanenkov, V. V., Stewart, R., Thomson, J. A., Crawford, G. E., … Ren, B. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, *459*(7243), 108–112. https://doi.org/10.1038/nature07829

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., Hoke, H. A., & Young, R. A. (2013). Super-enhancers in the control of cell identity and disease. *Cell*, *155*(4), 934–947. https://doi.org/10.1016/j.cell.2013.09.053

Hnisz, D., Day, D. S., & Young, R. A. (2016). Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. *Cell*, *167*(5), 1188–1200. https://doi.org/10.1016/j.cell.2016.10.024

Hnisz, D., Schuijers, J., Lin, C. Y., Weintraub, A. S., Abraham, B. J., Lee, T. I., Bradner, J. E., & Young, R. A. (2015). Convergence of Developmental and Oncogenic Signaling Pathways at Transcriptional Super-Enhancers. *Molecular Cell*, *58*(2), 362–370. https://doi.org/10.1016/j.molcel.2015.02.014

Ji, X., Dadon, D. B., Powell, B. E., Fan, Z. P., Borges-Rivera, D., Shachar, S., Weintraub, A. S., Hnisz, D., Pegoraro, G., Lee, T. I., Misteli, T., Jaenisch, R., & Young, R. A. (2016). 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. *Cell Stem Cell*, *18*(2), 262–275. https://doi.org/10.1016/j.stem.2015.11.007

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, and D. (2002). The Human Genome Browser at UCSC. *Genome Research*, *12*(6), 996–1006. https://doi.org/10.1101/gr.229102

Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S., & Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics*, *26*(17), 2204–2207. https://doi.org/10.1093/bioinformatics/btq351

Kharchenko, P. V, Silberstein, L., & Scadden, D. T. (2014). Bayesian approach to single-cell differential expression analysis. *Nature Methods*, *11*(7), 740–742. https://doi.org/10.1038/nmeth.2967

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W., & Ma'ayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, *44*(W1), W90–W97. https://doi.org/10.1093/nar/gkw377

Langmead, B., Wilks, C., Antonescu, V., & Charles, R. (2019). Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics*, *35*(3), 421–432. https://doi.org/10.1093/bioinformatics/bty648

Levine, M., & Tjian, R. (2003). Transcription regulation and animal diversity. *Nature*, *424*(6945), 147–151. https://doi.org/10.1038/nature01763

Li, Y., Rivera, C. M., Ishii, H., Jin, F., Selvaraj, S., Lee, A. Y., Dixon, J. R., & Ren, B. (2014). CRISPR Reveals a Distal Super-Enhancer Required for Sox2 Expression in Mouse Embryonic Stem Cells. *PLoS ONE*, *9*(12), e114485. https://doi.org/10.1371/journal.pone.0114485

Lovén, J., Hoke, H. A., Lin, C. Y., Lau, A., Orlando, D. A., Vakoc, C. R., Bradner, J. E., Lee, T. I., & Young, R. A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*, *153*(2), 320–334. https://doi.org/10.1016/j.cell.2013.03.036

Maston, G. A., Evans, S. K., & Green, M. R. (2006). Transcriptional Regulatory Elements in the Human Genome. *Annual Review of Genomics and Human Genetics*, *7*(1), 29–59. https://doi.org/10.1146/annurev.genom.7.080505.115623

McCord, R. P., Kaplan, N., & Giorgetti, L. (2020). Chromosome Conformation Capture and Beyond: Toward an Integrative View of Chromosome Structure and Function. *Molecular Cell*, *77*(4), 688–708. https://doi.org/10.1016/j.molcel.2019.12.021

McSwiggen, D. T., Hansen, A. S., Teves, S. S., Marie-Nelly, H., Hao, Y., Heckert, A. B., Umemoto, K. K., Dugast-Darzacq, C., Tjian, R., & Darzacq, X. (2019). Evidence for DNA-mediated nuclear compartmentalization distinct from phase separation. *ELife*, *8*. https://doi.org/10.7554/eLife.47098

Medeiros, L. A., Dennis, L. M., Gill, M. E., Houbaviy, H., Markoulaki, S., Fu, D., White, A. C., Kirak, O., Sharp, P. A., Page, D. C., & Jaenisch, R. (2011). *Mir-290–295* deficiency in mice results in partially penetrant embryonic lethality and germ cell defects. *Proceedings of the National Academy of Sciences*, *108*(34), 14163–14168. https://doi.org/10.1073/pnas.1111241108

Mikhaylichenko, O., Bondarenko, V., Harnett, D., Schor, I. E., Males, M., Viales, R. R., & Furlong, E. E. M. (2018). The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes & Development*, *32*(1), 42–57. https://doi.org/10.1101/gad.308619.117

Narendra, V., Rocha, P. P., An, D., Raviram, R., Skok, J. A., Mazzoni, E. O., & Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science (New York, N.Y.)*, *347*(6225), 1017–1021. https://doi.org/10.1126/science.1262088

Nguyen, T. A., Jones, R. D., Snavely, A. R., Pfenning, A. R., Kirchner, R., Hemberg, M., & Gray, J. M. (2016). High-throughput functional comparison of promoter and enhancer activities. *Genome Research*, *26*(8), 1023–1033. https://doi.org/10.1101/gr.204834.116

Nizami, Z., Deryusheva, S., & Gall, J. G. (2010). The Cajal Body and Histone Locus Body. *Cold Spring Harbor Perspectives in Biology*, *2*(7), a000653–a000653. https://doi.org/10.1101/cshperspect.a000653

Parker, S. C. J., Stitzel, M. L., Taylor, D. L., Orozco, J. M., Erdos, M. R., Akiyama, J. A., van Bueren, K. L., Chines, P. S., Narisu, N., Black, B. L., Visel, A., Pennacchio, L. A., Collins, F. S., Becker, J., Benjamin, B., Blakesley, R., Bouffard, G., Brooks, S., Coleman, H., … Young, A. (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proceedings of the National Academy of Sciences*, *110*(44), 17921–17926. https://doi.org/10.1073/pnas.1317023110

Pombo, A., & Dillon, N. (2015). Three-dimensional genome architecture: players and mechanisms. *Nature Reviews Molecular Cell Biology*, *16*(4), 245–257. https://doi.org/10.1038/nrm3965

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Quinodoz, S. A., Bhat, P., Chovanec, P., Jachowicz, J. W., Ollikainen, N., Detmar, E., Soehalim, E., & Guttman, M. (2022). SPRITE: a genome-wide method for mapping higher-order 3D interactions in the nucleus using combinatorial split-and-pool barcoding. *Nature Protocols*, *17*(1), 36–75. https://doi.org/10.1038/s41596-021-00633-y

Quinodoz, S. A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J. M., Detmar, E., Lai, M. M., Shishkin, A. A., Bhat, P., Takei, Y., Trinh, V., Aznauryan, E., Russell, P., Cheng, C., Jovanovic, M., Chow, A., Cai, L., McDonel, P., Garber, M., & Guttman, M. (2018). Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell*, *174*(3), 744-757.e24. https://doi.org/10.1016/j.cell.2018.05.024

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., & Aiden, E. L. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*, *159*(7), 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021

Sabari, B. R., Dall'Agnese, A., Boija, A., Klein, I. A., Coffey, E. L., Shrinivas, K., Abraham, B. J., Hannett, N. M., Zamudio, A. V, Manteiga, J. C., Li, C. H., Guo, Y. E., Day, D. S., Schuijers, J., Vasile, E., Malik, S., Hnisz, D., Lee, T. I., Cisse, I. I., … Young, R. A. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. *Science (New York, N.Y.)*, *361*(6400). https://doi.org/10.1126/science.aar3958

Schuijers, J., Manteiga, J. C., Weintraub, A. S., Day, D. S., Zamudio, A. V., Hnisz, D., Lee, T. I., & Young, R. A. (2018). Transcriptional Dysregulation of MYC Reveals Common Enhancer-Docking Mechanism. *Cell Reports*, *23*(2), 349–360. https://doi.org/10.1016/j.celrep.2018.03.056

Serfling, E., Lübbe, A., Dorsch-Häsler, K., & Schaffner, W. (1985). Metal-dependent SV40 viruses containing inducible enhancers from the upstream region of metallothionein genes. *The EMBO Journal*, *4*(13B), 3851–3859. https://doi.org/10.1002/j.1460-2075.1985.tb04157.x

Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., & Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biology*, *16*(1), 259. https://doi.org/10.1186/s13059-015-0831-x

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., & Cavalli, G. (2012). Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell*, *148*(3), 458–472. https://doi.org/10.1016/j.cell.2012.01.010

Shen, Y., Yue, F., McCleary, D. F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V. V., & Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature*, *488*(7409), 116–120. https://doi.org/10.1038/nature11243

Skinnider, M. A., Squair, J. W., & Foster, L. J. (2019). Evaluating measures of association for single-cell transcriptomics. *Nature Methods*, *16*(5), 381–386. https://doi.org/10.1038/s41592-019-0372-4

Spitz, F., & Furlong, E. E. M. (2012). Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics*, *13*(9), 613–626. https://doi.org/10.1038/nrg3207

Stuart, T., Srivastava, A., Madad, S., Lareau, C. A., & Satija, R. (2021). Single-cell chromatin state analysis with Signac. *Nature Methods*, *18*(11), 1333–1341. https://doi.org/10.1038/s41592-021-01282-5

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, *126*(4), 663–676. https://doi.org/10.1016/j.cell.2006.07.024

The ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74. https://doi.org/10.1038/nature11247

Ulianov, S. V., Khrameeva, E. E., Gavrilov, A. A., Flyamer, I. M., Kos, P., Mikhaleva, E. A., Penin, A. A., Logacheva, M. D., Imakaev, M. V., Chertovich, A., Gelfand, M. S., Shevelyov, Y. Y., & Razin, S. V. (2016). Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Research*, *26*(1), 70–84. https://doi.org/10.1101/gr.196006.115

Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A., & Luscombe, N. M. (2009). A census of human transcription factors: function, expression and evolution. *Nature Reviews. Genetics*, *10*(4), 252–263. https://doi.org/10.1038/nrg2538

Visel, A., Rubin, E. M., & Pennacchio, L. A. (2009). Genomic views of distant-acting enhancers. *Nature*, *461*(7261), 199–205. https://doi.org/10.1038/nature08451

Weintraub, A. S., Li, C. H., Zamudio, A. V., Sigova, A. A., Hannett, N. M., Day, D. S., Abraham, B. J., Cohen, M. A., Nabet, B., Buckley, D. L., Guo, Y. E., Hnisz, D., Jaenisch, R., Bradner, J. E., Gray, N. S., & Young, R. A. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell*, *171*(7), 1573-1588.e28. https://doi.org/10.1016/j.cell.2017.11.008

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I., & Young, R. A. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell*, *153*(2), 307–319. https://doi.org/10.1016/j.cell.2013.03.035

Williamson, I., Hill, R. E., & Bickmore, W. A. (2011). Enhancers: From Developmental Genetics to the Genetics of Common Human Disease. *Developmental Cell*, *21*(1), 17–19. https://doi.org/10.1016/j.devcel.2011.06.008

Xie, Z., Bailey, A., Kuleshov, M. V., Clarke, D. J. B., Evangelista, J. E., Jenkins, S. L., Lachmann, A., Wojciechowicz, M. L., Kropiwnicki, E., Jagodnik, K. M., Jeon, M., & Ma'ayan, A. (2021). Gene Set Knowledge Discovery with Enrichr. *Current Protocols*, *1*(3). https://doi.org/10.1002/cpz1.90

Zhang, S. (2014). Sox2, a key factor in the regulation of pluripotency and neural differentiation. *World Journal of Stem Cells*, *6*(3), 305. https://doi.org/10.4252/wjsc.v6.i3.305

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137. https://doi.org/10.1186/gb-2008-9-9-r137

Zhou, H. Y., Katsman, Y., Dhaliwal, N. K., Davidson, S., Macpherson, N. N., Sakthidevi, M., Collura, F., & Mitchell, J. A. (2014a). A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. *Genes & Development*, *28*(24), 2699–2711. https://doi.org/10.1101/gad.248526.114

Zhou, H. Y., Katsman, Y., Dhaliwal, N. K., Davidson, S., Macpherson, N. N., Sakthidevi, M., Collura, F., & Mitchell, J. A. (2014b). A Sox2 distal enhancer cluster regulates embryonic

stem cell differentiation potential. *Genes & Development*, *28*(24), 2699–2711. https://doi.org/10.1101/gad.248526.114

Zhou, X., Edmonson, M. N., Wilkinson, M. R., Patel, A., Wu, G., Liu, Y., Li, Y., Zhang, Z., Rusch, M. C., Parker, M., Becksfort, J., Downing, J. R., & Zhang, J. (2016). Exploring genomic alteration in pediatric cancer using ProteinPaint. *Nature Genetics*, *48*(1), 4–6. https://doi.org/10.1038/ng.3466

FIGURE LEGENDS

Figure 1: BOUQUET uses genome topology to identify 3D cis-regulatory element communities.

A) Workflow of the cis-regulatory element (CRE) community-builder, Building Optimized Units of Quantified Enhancer Topologies (BOUQUET). High-confidence direct loops between CRE-pairs (green) are identified by H3K27ac HiChIP using Hi-LOW. Interaction of CTCF-bound insulator pairs forms insulated neighborhood (IN) loops (purple) thought to constrain CRE-CRE interactions. BOUQUET takes in CRE coordinates, loop calls, INs, and topologically associated domains (TADs) to connect CRE-pairs showing evidence of physical interaction. CRE-pairs are assembled into a CRE network. Using label-propagation, BOUQUET infers a given gene's CRE landscape by collecting a set of CREs that are densely connected (CRE subnetwork with high density of edges), called communities (highlighted in yellow). Unassigned CREs are added to the gene's community by co-occurrence within the same TAD. See Methods for additional details.

B) Number of CREs per 3D CRE community for communities with at least two CREs.

C) 3D CRE community gene expression level versus the number of CREs in the 3D CRE community (Spearman rho = 0.23, p-value = 2.2×10^{-127}).

D) Network diagrams of example 3D CRE communities. *Left:* a simple 3D CRE community, containing the focal gene *Zfp160* and another gene *Ppp2r1a*. *Right:* a complex 3D CRE community, containing the gene *Sox2*. Each 3D CRE community is highlighted in yellow in their respective network component diagram. While *Fxr1* and *Dnajc19* have high-confidence connections to CREs of *Sox2*'s 3D CRE community, they lack the degree of interconnection required to be considered as part of *Sox2*'s 3D CRE community.

E) A simple 3D CRE community. The promoters of *Zfp160* and *Ppp2r1a* are connected to each other and to a pair of downstream enhancers by high-confidence H3K27ac HiChIP loops. TADs and insulated neighborhoods are shown as black bars. The ends of these features are faded to show that they extend beyond the boundaries of this region. The transcriptionally inactive gene *Vmn1r233* was excluded from this CRE community. OSN SE = linear super-enhancers defined using OCT4/SOX2/NANOG as in Whyte et al., 2013.

F) A complex 3D CRE community. Several high-confidence H3K27ac HiChIP loops connect *Sox2* to the *Sox2* Control Region (SCR). The ends of TADs and insulated neighborhoods are faded to show that these features extend beyond the boundaries of this region. OSN SE = linear super-enhancers defined using OCT4/SOX2/NANOG as in Whyte et al., 2013.

Supplementary Figure 1: BOUQUET incorporates high-confidence HiChIP loops, insulated neighborhoods, and topologically associated domains to infer 3D CRE communities.

A) Generalized workflow of HiChIP/HiC Looping Organization Wrapper (HiLOW) and Building Optimized Units of QUantified Enhancer Topologies (BOUQUET). See Methods for details.

B) Fraction of enhancers and promoters involved in direct, high-confidence H3K27ac HiChIP loops; the fraction either involved in direct loops or falling within insulated neighborhoods (INs); and the fraction involved in direct loops, falling within INs, or falling within topologically associated domains (TADs). Pink indicates the percentage of enhancers/promoters covered by the given feature; gray indicates the percentage not covered.

C) Characteristics of the genomic regions (5kb bins) that interact with promoter regions by HiChIP loops. A single promoter-interacting region may have multiple characteristics.

D) Workflow of *de novo* IN calling and comparison with lifted-over, published INs using the same data. There are 22,589 interactions in common between *de novo* SMC1A ChIA-PET interactions and those lifted over from mm9 to mm10 (Dowen et al., 2014, Hnisz et al., 2016). INs are the subset of SMC1A-anchored loops also anchored by CTCF peaks on both sides that contain at least one gene. 13,777 insulated neighborhoods were observed in mESCs for mm10.

E) Visualization of INs identified with two methods in mm10 at a specific locus. Previously identified INs and significant interactions from SMC1A ChIA-PET were converted from mm9 to mm10 coordinates using liftover. INs and significant interactions were also identified *de novo* from the same raw data in mm10. Significant interactions and INs from both methods, as well as read coverage and anchors from *de novo* analysis are shown. INs are shown as arcs. High-confidence peaks and RPM-normalized read coverage of CTCF ChIP-Seq are also shown.

F) Fractions of enhancers, promoters, and the overall genome that fall within INs or topologically associated domains (TADs). Pink indicates the percentage covered by the given feature; gray indicates the percentage not covered.

G) Scatterplot of the contact frequency of CRE-pairs connected by direct loops (pink) or insulated neighborhoods (INs; blue) versus the genomic distance.

H) Percentage of high-confidence CRE connections in three categories. Pink denotes cases where the same IN is the smallest IN each CRE. Orange denotes cases where there is an IN that contains both CREs, but it is not the smallest for both CREs. Grey denotes cases where there is no IN that contains both CREs.

I) Fractions of CREs that can be assigned to 3D CRE communities with various topological evidence. CREs assignable by high-confidence HiChIP loops and insulated neighborhoods to “pre-communities” (pink), and CREs assignable to communities after incorporating TAD data (orange) are noted. Gray indicates the fraction of active CREs not assigned to a 3D CRE community.

J) Numbers of promoters and enhancers per 3D CRE community with at least two CREs.

K) Number of promoters (y-axis) versus the number of enhancers (x-axis) in each 3D CRE community. Spearman rho = 0.38, p-value = 2.1×10^{-158}.

Figure 2: Enhancer deletions selectively impact the expression of genes within the same community.

A) UCSC Genome Browser view of the CRISPR-Cas9 enhancer deletion site (DE, green). Constituent enhancer E4 of the linear super-enhancer associated with *D7Ertd143e* was deleted.This is a reanalysis using the same raw data and based on Fig. 2A of Hnisz et al., 2015. Single-cell ATAC-seq performed in mESCs was pooled across all cells for pseudobulk coverage.

B) UCSC Genome Browser view of the locus containing the *D7Ertd143e* 3DSE CRE community. Genes within the community are marked by the red box; neighboring genes from the same component but different communities are marked by the blue box. The CRISPR-Cas9 deletion site (DE) of a constituent enhancer of the super-enhancer is marked in green. Window size: 1.7 MB.

C) Network diagram of the *D7Ertd143e* component. Circles denote promoters, squares denote enhancers. Red lines indicate direct connections with the focal gene, *D7Ertd143e*. Grey lines indicate all other connections between CREs. The colored clouds denote distinct substructures within this component; the red cloud marks the community of *D7Ertd143e*, in which nodes are highly interconnected. The blue cloud marks linearly neighboring CREs that are less connected to *D7Ertd143e* and are only part of its larger component, not the *D7Ertd143e* community.

D) Change in expression of all genes (light gray), neighboring genes not within the *D7Ertd143e* 3DSE CRE community (dark gray), and genes within the *D7Ertd143e* 3DSE CRE community (red) following deletion of the *D7Ertd143e* constituent enhancer. These data are spike-in normalized.

E) TPM and spike-in-normalized expression of individual genes within the *D7Ertd143e* 3DSE CRE community. Gray dots denote wild type (WT) samples; red dots denote samples in which the constituent enhancer E4 of the *D7Ertd143e* super-enhancer has been deleted (DE).

F) TPM and spike-in-normalized expression of individual genes linearly neighboring the *D7Ertd143e* 3DSE CRE community. Gray dots denote wild type (WT) samples; red dots denote samples in which the constituent enhancer E4 of the *D7Ertd143e* 3DSE CRE community has been deleted (DE). Blue box denotes genes within the network component.

Supplementary Figure 2: Enhancer deletions in the *D7Ertd143e* 3DSE CRE community.

A) Change in expression of all genes (light gray), neighboring genes not within the *D7Ertd143e* 3DSE CRE community (dark gray), and genes within the *D7Ertd143e* 3DSE CRE community (red) following deletion of the *D7Ertd143e* constituent enhancer. These data are not spike-in normalized.

B) TPM and non-spike-in-normalized expression of individual genes within the *D7Ertd143e* 3DSE CRE community, related to Figure 2E. Gray dots denote wild type (WT) samples; red dots denote samples in which the constituent enhancer E4 of the *D7Ertd143e* super-enhancer has been deleted (DE).

C) TPM and non-spike-in-normalized expression of individual genes linearly neighboring the *D7Ertd143e* 3DSE CRE community, related to Figure 2F. Gray dots denote wild type (WT) samples; red dots denote samples in which the constituent enhancer E4 of the *D7Ertd143e* 3DSE CRE community has been deleted (DE). Blue box denotes genes within the network component.

Figure 3: *3DSE CRE communities recover and expand beyond linear SEs approaches.*

A) Comparison of the enhancer-only MED1 signal (y-axis) versus MED1 signal at enhancers and promoters (x-axis) in each 3D CRE community with at least two enhancers. Both x- and y- axis display log2 transformed values. Linear values of 50, 500 and 5000 are marked as grey dashed lines. Blue lines indicate cutoff thresholds for calling 3DSE communities using either enhancer-only or enhancers plus promoters MED1 signal. Spearman rho = 0.92.

B) Stitched enhancer ranking by MED1 signal according to ROSE, after Whyte et al., 2013. A subset of stitched enhancers with exceptional signal comprises linear super-enhancers.

C) 3D CRE community ranking by MED1 signal at enhancers. A subset of 3D CRE communities with exceptional signal comprises 3DSE E-CRE communities.

D) Comparison of the linear SE-identifying ROSE algorithm rank (stitched enhancer MED1 rank) and the 3D CRE community BOUQUET rank (enhancer-only MED1 rank) for the same assigned gene. The most signal-rich stitched enhancer was considered if multiple enhancers were assigned to the same gene by ROSE. 0,0 denotes highest rank in both BOUQUET and ROSE. Red dashed line denotes 3D super-enhancer cutoff threshold.

E) 3D CRE community gene expression versus enhancer signal for each 3D CRE community. Spearman rho = 0.23, p-value = 1.2×10^{-130}.

F) 3D CRE community ranking by MED1 signal at both enhancers and promoters. A subset of 3D CRE communities with exceptional signal comprises 3DSE EP-CRE communities.

G) Network diagram of histone genes. Circles denote promoters, squares denote enhancers. Four CRE communities (red, green, purple, and yellow) are identified. Altogether, this component comprised four 3DSE CRE regulatory communities.

H) UCSC Genome Browser view of histone genes. Distinct regulatory communities are outlined in red, green, yellow, and purple; encompassing box colors correspond to the colors of the nodes in the network diagram in panel (G).

I) Presence of binding sites for three example signaling transcription factors (TFs) STAT3, SMAD3, and TCF3 in the collected CREs within different types of community. Bars denote the presence of at least one binding site, as measured by the presence of a ChIP-seq peak for 0, 1, 2, or all 3 types of TFs. Here the binding sites within 3DSE CRE communities (red, 822 communities), typical enhancer communities (gray, 10,228 communities), and randomized sets of typical community CREs comparable in size to the 3DSE CRE communities (orange, 822 randomized sets), are compared.

Supplementary Figure 3: Linear and three-dimensional super-enhancers.

A) Workflow of the Rank-Ordering of Super-Enhancers (ROSE) algorithm for the identification of linear super-enhancers. Briefly, ROSE aggregates linearly proximal enhancers into stitched enhancers and ranks the stitched enhancers by their MED1 ChIP-Seq signal. Stitching occurs if pairs of constituent enhancers are within 12,500 bp of each other.

B) Overlap of the genes associated with linear super-enhancers (SEs) identified by ROSE, genes associated with 3DSE CRE communities ranked using enhancer-only MED1 ChIP-Seq signal (3DSE E-CRE communities), and genes associated with 3DSE CRE communities ranked using enhancer-and-promoter MED1 ChIP-Seq signal (3DSE EP-CRE communities).

C) UCSC Genome Browser view of genes *Rpap3* (blue) and *Pced1b* (red). The linear super-enhancer (red box, OSN SE track) is assigned to *Rpap3* (blue) based on proximity but is reassigned to *Pced1b* (red) under the 3DSE CRE community model based on genome topology data. Window size: 530 kb. OSN enhs = enhancers defined by OCT4/SOX2/NANOG overlap, as in Whyte et al., 2013.

D) UCSC Genome Browser view of the locus containing the *H2-M5* (red, gene track) 3DSE CRE community. In addition to the linear super-enhancer (SE) assigned to *H2-M5* (red, OSN SE track), BOUQUET assigns multiple typical enhancers to the *H2-M5* community (red, H3K27ac enhs track) and also reassigns the SE (blue, OSN SE track) originally assigned to *Olfr90* (blue, gene track) by ROSE to the *H2-M5* community as well. Window size: 1.4 MB. OSN enhs = enhancers defined by OCT4/SOX2/NANOG overlap, as in Whyte et al., 2013.

E) Analysis of the top 5 ranked gene ontology (GO) terms for genes associated with linear super-enhancers (linear SEs), genes associated with 3DSE CRE communities based on enhancer signal (3DSE E-CRE communities), and genes associated with 3DSE CRE community based on signal from enhancers and promoters (3DSE EP-CRE communities). The adjusted p-value for each term corresponds to the color denoted in the legend.

Figure 4: 3DSE CRE community members co-localize, are co-expressed, are co-accessible, and associate with co-factor puncta

A) Model of predicted outcomes from experimental design. Two regions-of-interest are depicted. In case 1, two gene DNAs and the mRNAs they encode are within a BRD4 co-factor punctum. These are

visualized using DNA-FISH, RNA-FISH, and BRD4 immunofluorescence (IF), respectively. In the model and below, *Esrrb* is represented in green, *Sptlc2* is represented in magenta, and BRD4 in cyan.

B) UCSC Genome Browser view of the locus containing the *Esrrb/Irf2bpl/Sptlc2* 3DSE CRE community. Locations for the immunofluorescent oligonucleotide probes used in RNA and DNA FISH experiments are marked with the corresponding colored bars: *Esrrb* (chr12:86,465,000-86,505,000; green), and *Sptlc2* (chr12:87,295,000-87,325,000; magenta).

C) Images of projections from mESC nuclei following sequential RNA/DNA FISH with BRD4 immunofluorescence. The regions of interest (ROIs), defined by the presence of one or both DNAs, are marked in each panel with an arrow. Blue = DAPI stain for chromatin; green = RNA and DNA probes for *Esrrb*; magenta = RNA and DNA probes for *Sptlc2*; cyan = anti-BRD4 antibody. Circled regions-of-interest = ROIs. Left scale bar = 5 micrometers. Right is a zoomed-in view of the same pair of ROIs. Top panels show results from DNA-FISH; bottom panels show results from RNA-FISH and IF. Annihilation of green/magenta/cyan shown as white in combined image.

D) Percentages of regions of interest (ROIs) overlapping nascent mRNA, DNA, and BRD4 fluorescence. *Left*: Percentage of ROIs where the presence (blue) or absence (gray) of BRD4 was detected under four conditions: co-expressed mRNA and co-localized genes detected; a single mRNA and a single gene detected; no mRNA expression and co-localized genes detected; and no mRNA expression and a single gene detected. Categories are mutually exclusive. *Right:* Percentage of ROIs where the presence (blue) or absence (gray) of BRD4 was detected with: the *Sptlc2* gene and the *Sptlc2* mRNA; the *Esrrb* gene and the *Esrrb* mRNA. Percentage of ROIs where the presence (light red) or absence (yellow) of the corresponding gene's mRNA was detected with: the *Sptlc2* gene and BRD4; the *Esrrb* gene and BRD4.

E) Distributions of the scATAC-seq-based odds ratio for co-accessibility of CRE pairs genome-wide, i.e., not within the same 3D CRE community (mean Odd Ratios (OR)=0.53, gray), CRE pairs within the same 3D CRE community (mean OR=0.59, blue), CRE pairs within the same 3DSE CRE community (mean OR=0.61,red), and CRE pairs within the same linear super-enhancer (mean OR=0.61,SE, light red). For comparison, 3D CRE community vs genome-wide ($p=3.6x10^{-97}$, bootstrapping procedure assuming normal distribution); 3DSE CRE community vs. genome-wide ($p=9.6x10^{-93}$, bootstrapping procedure assuming normal distribution); 3DSE vs 3D CRE community ($p=3.1x10^{-21}$, bootstrapping procedure assuming normal distribution).

F) Distributions of the scRNA-seq-based Spearman's correlation coefficient for co-expression of gene-pairs genome-wide (gray, mean r= 0.01), gene-pairs within the same 3D CRE community (blue, mean r=0.025), and gene-pairs within the same 3DSE CRE community (red, mean r=0.026). For comparison, 3D CRE community vs genome-wide ($p=2.9x10^{-95}$, Wilcoxon Rank Sum Test); 3DSE CRE community vs genome-wide ($p=3.5x10^{-226}$, Wilcoxon Rank Sum Test)

Supplementary Figure 4: 3DSE CRE community constituents co-localize, are co-expressed, and associate with transcriptional condensate proteins in distinct nuclear puncta.

A) Experimental workflow of sequential RNA/DNA fluorescence *in situ* hybridization (FISH) with BRD4 immunofluorescence (IF). Briefly, BRD4 IF and two-color RNA-FISH is performed; RNAs are degraded with RNAseA treatment, then DNA-FISH is performed. See Methods for additional details.

B) UMAP plot of scRNA-seq in 847 mESCs. Automatically determined clusters are in distinct colors.

C) Distribution of gene-pair expression correlations using scRNA-seq across eight correlation metrics. Bars from left to right: all gene-pairs (gray); gene-pairs within the same 3D CRE community (blue); gene-pairs within the same 3DSE CRE community (red).

D) UMAP plot of scATAC-seq in 3,171 mESCs. Automatically determined clusters are in distinct colors.

Figure 1

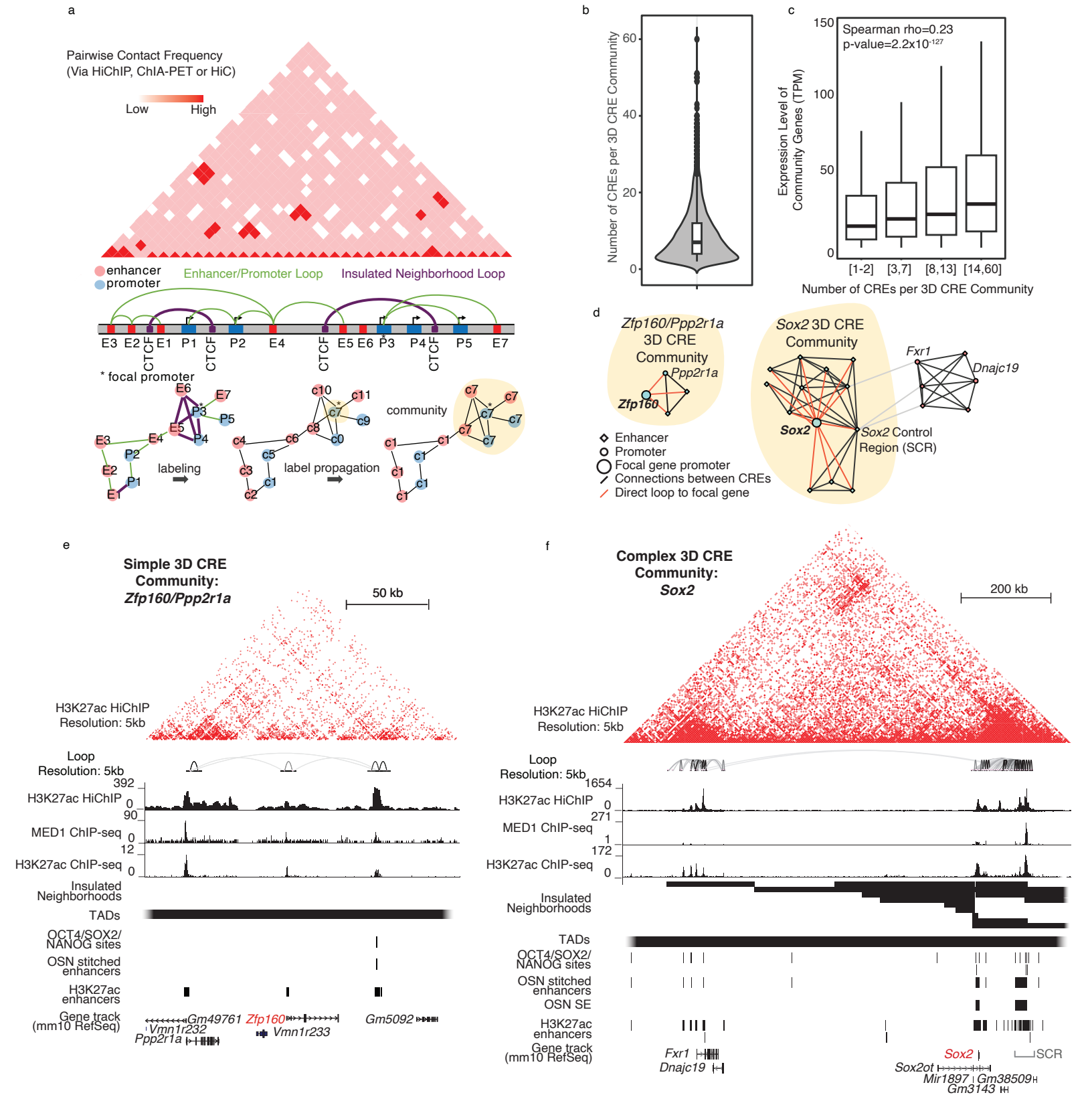

a
Pairwise Contact Frequency
(Via HiChIP, ChIA-PET or HiC)
Low
High
enhancer
promoter
Enhancer/Promoter Loop
Insulated Neighborhood Loop
E3 E2 CTCF E1 P1 CTCF P2 E4 CTCF E5 E6 P3 P4 CTCF P5 E7
* focal promoter
labeling
label propagation
community
b
Number of CREs per 3D CRE Community
c
Spearman rho=0.23
p-value=2.2x10-127
Expression Level of Community Genes (TPM)
[1-2] [3,7] [8,13] [14,60]
Number of CREs per 3D CRE Community
d
Zfp160/Ppp2r1a 3D CRE Community
Ppp2r1a
Zfp160
Sox2 3D CRE Community
Fxr1
Dnajc19
Sox2
Sox2 Control Region (SCR)
Enhancer
Promoter
Focal gene promoter
Connections between CREs
Direct loop to focal gene
e
Simple 3D CRE Community: Zfp160/Ppp2r1a
50 kb
H3K27ac HiChIP Resolution: 5kb
Loop Resolution: 5kb
H3K27ac HiChIP
MED1 ChIP-seq
H3K27ac ChIP-seq
Insulated Neighborhoods
TADs
OCT4/SOX2/ NANOG sites
OSN stitched enhancers
H3K27ac enhancers
Gene track (mm10 RefSeq)
Gm49761 Zfp160 Gm5092
Vmn1r232 Vmn1r233
Ppp2r1a
f
Complex 3D CRE Community: Sox2
200 kb
H3K27ac HiChIP Resolution: 5kb
Loop Resolution: 5kb
H3K27ac HiChIP
MED1 ChIP-seq
H3K27ac ChIP-seq
Insulated Neighborhoods
TADs
OCT4/SOX2/ NANOG sites
OSN stitched enhancers
OSN SE
H3K27ac enhancers
Gene track (mm10 RefSeq)
Fxr1
Dnajc19
Sox2
SCR
Sox2ot
Mir1897 Gm38509
Gm3143

Figure 2

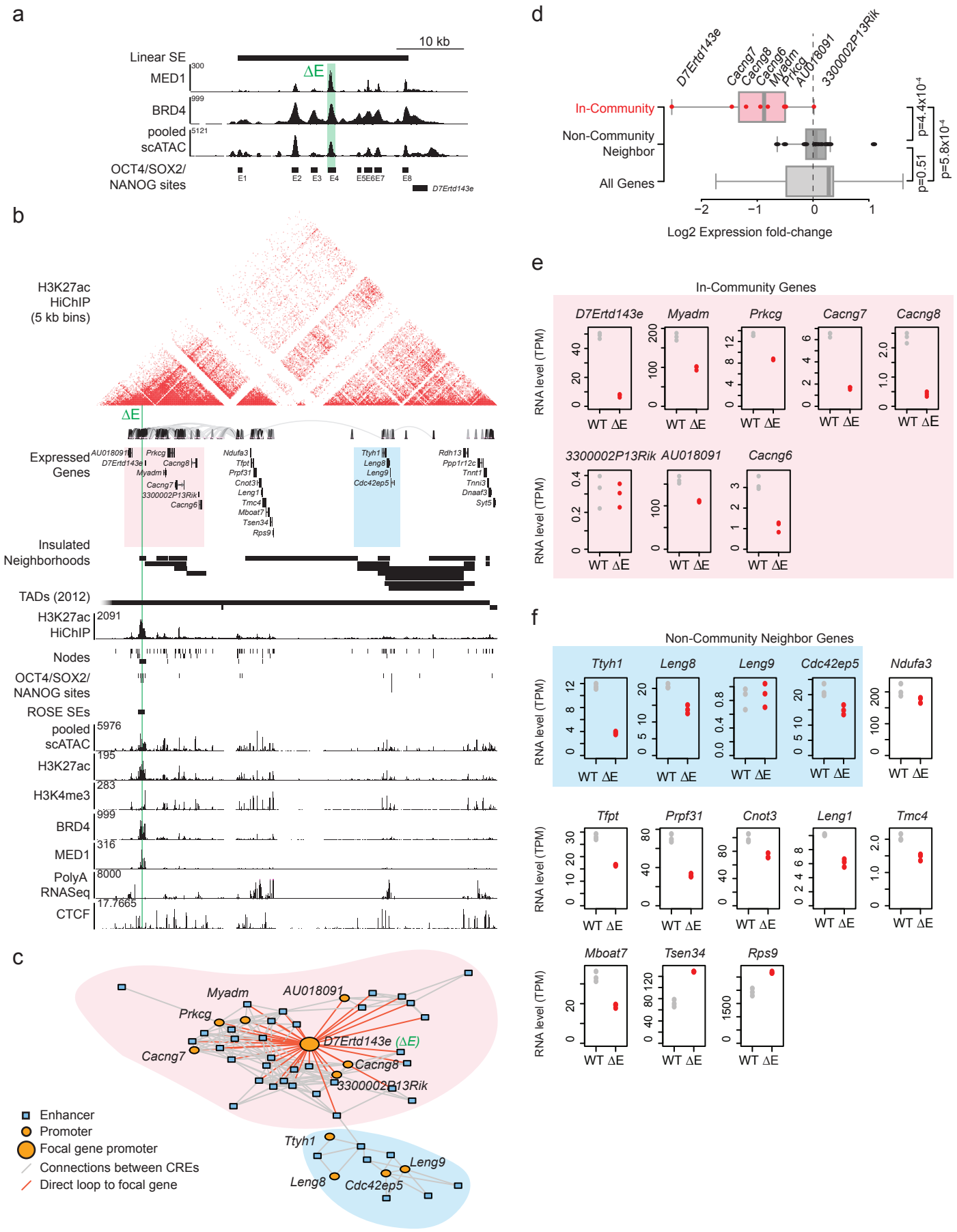

a
10 kb
Linear SE
MED1
BRD4
pooled scATAC
ΔE
OCT4/SOX2/ NANOG sites
E1
E2
E3
E4
E5E6E7
E8
D7Ertd143e
b
H3K27ac HiChIP (5 kb bins)
ΔE
Expressed Genes
Insulated Neighborhoods
TADs (2012)
H3K27ac HiChIP
Nodes
OCT4/SOX2/ NANOG sites
ROSE SEs
pooled scATAC
H3K27ac
H3K4me3
BRD4
MED1
PolyA RNASeq
CTCF
c
Myadm
AU018091
Prkcg
D7Ertd143e (ΔE)
Cacng7
Cacng8
3300002P13Rik
Ttyh1
Leng9
Leng8
Cdc42ep5
Enhancer
Promoter
Focal gene promoter
Connections between CREs
Direct loop to focal gene
d
D7Ertd143e
Cacng7
Cacng8
Cacng6
Myadm
Prkcg
AU018091
3300002P13Rik
In-Community
Non-Community Neighbor
All Genes
p=4.4x10-4
p=5.8x10-4
p=0.51
Log2 Expression fold-change
e
In-Community Genes
D7Ertd143e
Myadm
Prkcg
Cacng7
Cacng8
3300002P13Rik
AU018091
Cacng6
RNA level (TPM)
WT ΔE
f
Non-Community Neighbor Genes
Ttyh1
Leng8
Leng9
Cdc42ep5
Ndufa3
Tfpt
Prpf31
Cnot3
Leng1
Tmc4
Mboat7
Tsen34
Rps9

Figure 3

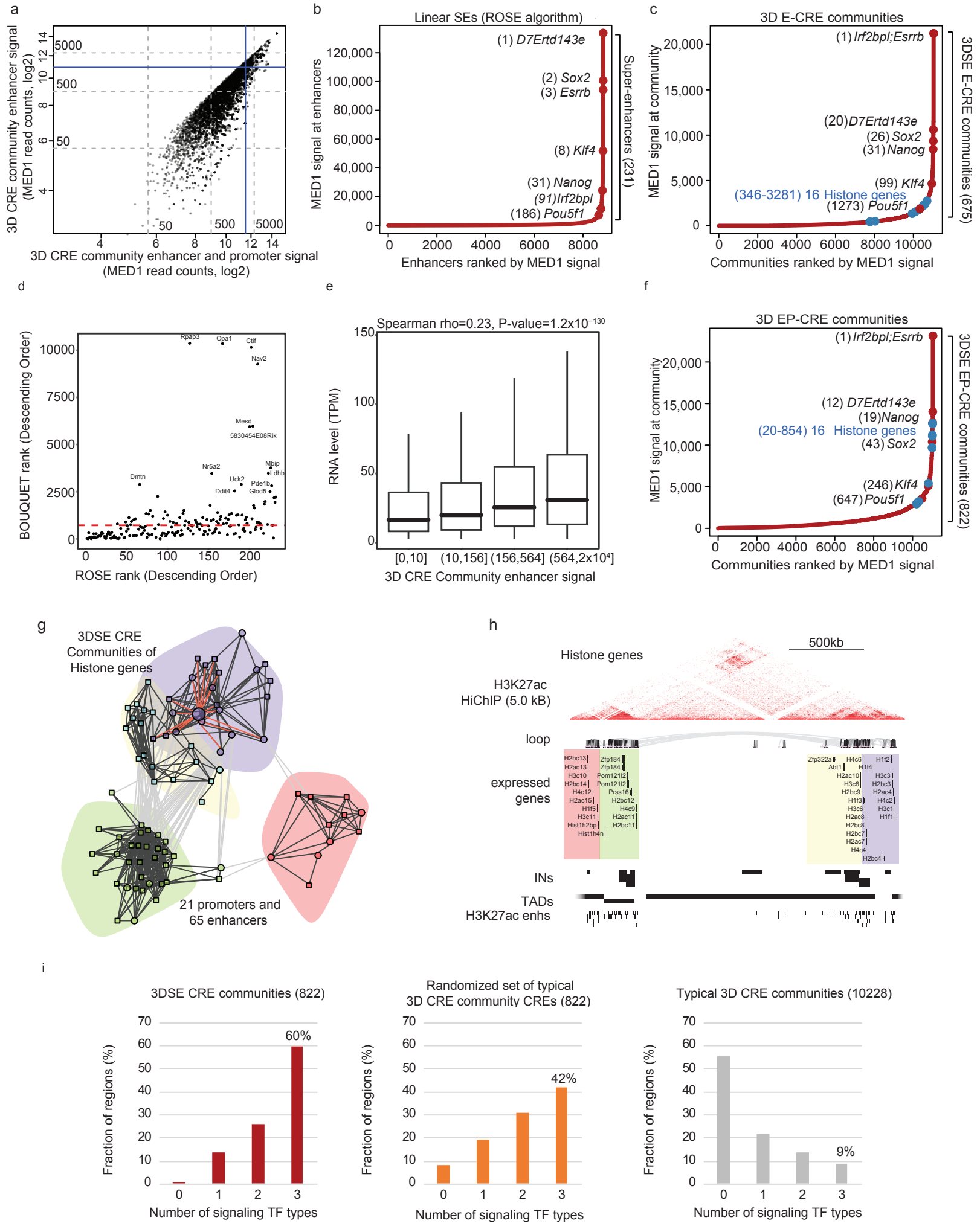
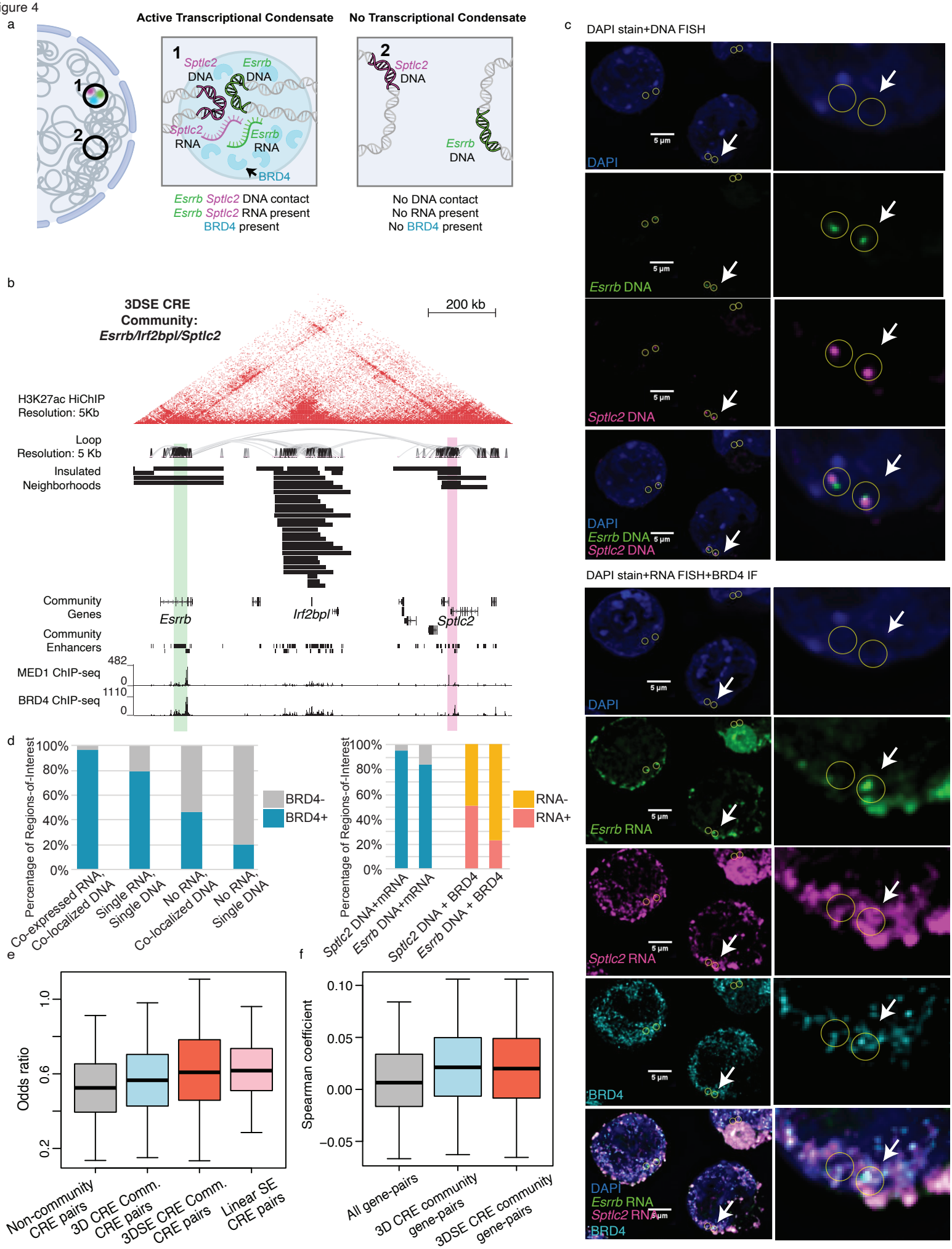
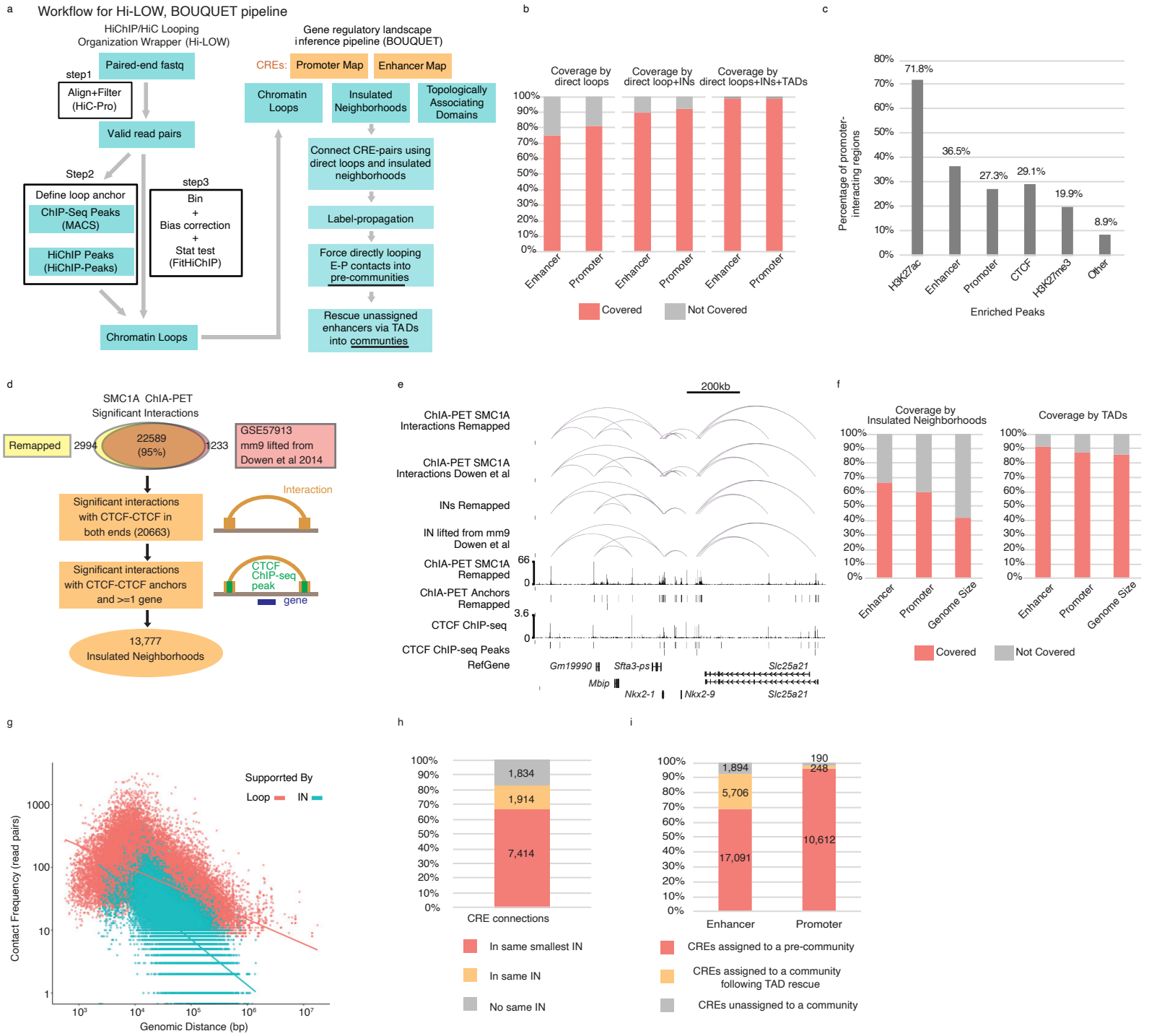

a
3D CRE community enhancer signal (MED1 read counts, log2)
3D CRE community enhancer and promoter signal (MED1 read counts, log2)
5000
500
50
b
Linear SEs (ROSE algorithm)
MED1 signal at enhancers
Enhancers ranked by MED1 signal
(1) D7Ertd143e
(2) Sox2
(3) Esrrb
(8) Klf4
(31) Nanog
(91)Irf2bpl
(186) Pou5f1
Super-enhancers (231)
c
3D E-CRE communities
MED1 signal at community
Communities ranked by MED1 signal
(1) Irf2bpl;Esrrb
(20)D7Ertd143e
(26) Sox2
(31)Nanog
(99) Klf4
(346-3281) 16 Histone genes
(1273) Pou5f1
3DSE E-CRE communities (675)
d
BOUQUET rank (Descending Order)
ROSE rank (Descending Order)
Rpap3
Opa1
Ctif
Nav2
Mesd
5830454E08Rik
Nr5a2
Mbip
Ldhb
Dmtn
Uck2
Pde1b
Ddit4
Glod5
e
Spearman rho=0.23, P-value=1.2x10^-130
RNA level (TPM)
3D CRE Community enhancer signal
[0,10]
(10,156]
(156,564]
(564,2x10^4]
f
3D EP-CRE communities
MED1 signal at community
Communities ranked by MED1 signal
(1) Irf2bpl;Esrrb
(12) D7Ertd143e
(19)Nanog
(20-854) 16 Histone genes
(43) Sox2
(246) Klf4
(647) Pou5f1
3DSE EP-CRE communities (822)
g
3DSE CRE Communities of Histone genes
21 promoters and 65 enhancers
h
Histone genes
500kb
H3K27ac HiChIP (5.0 kB)
loop
expressed genes
INs
TADs
H3K27ac enhs
i
3DSE CRE communities (822)
60%
Randomized set of typical 3D CRE community CREs (822)
42%
Typical 3D CRE communities (10228)
9%
Fraction of regions (%)
Number of signaling TF types

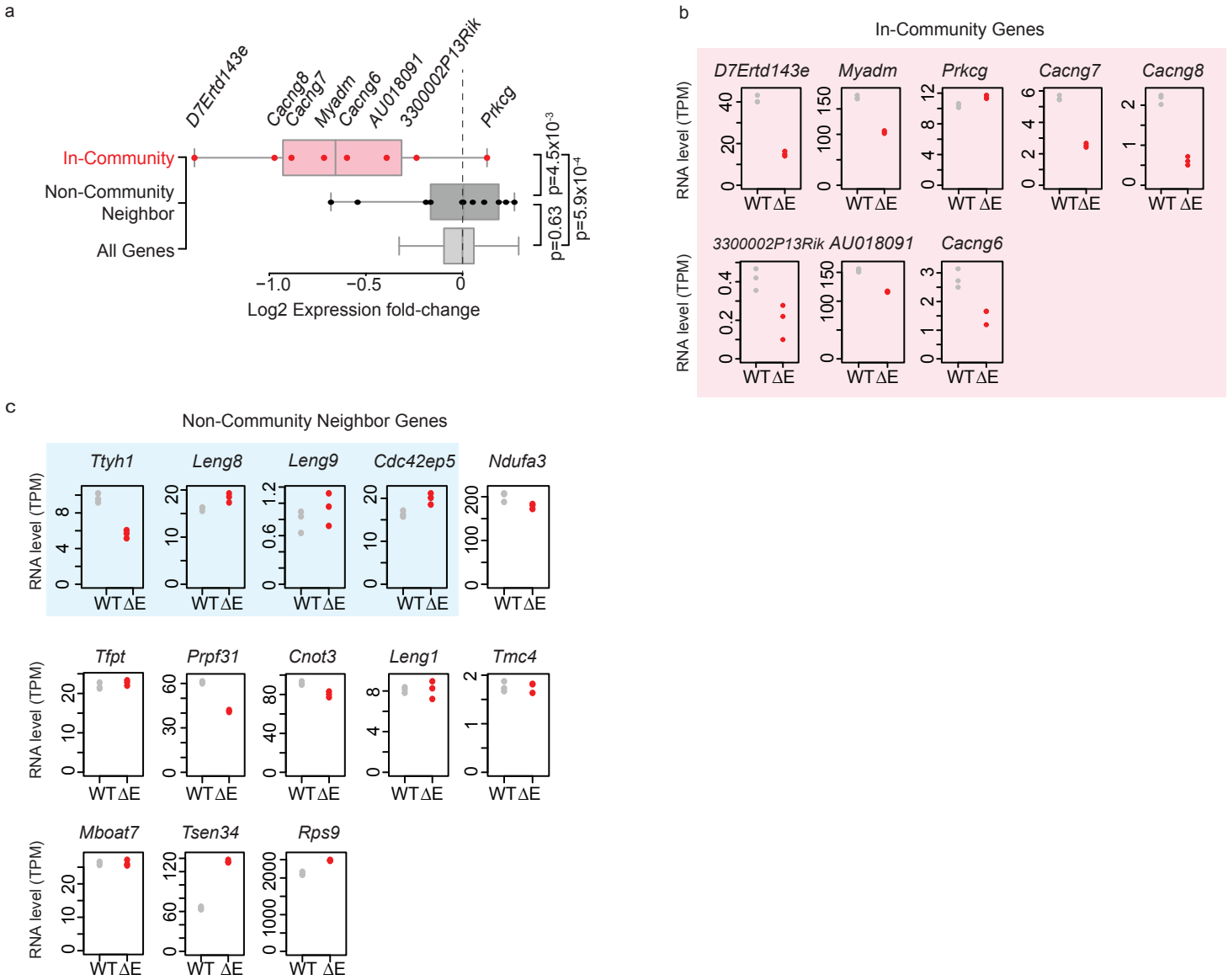
Figure 4


a
Active Transcriptional Condensate
No Transcriptional Condensate
1
2
Sptlc2 DNA
Esrrb DNA
Sptlc2 RNA
Esrrb RNA
BRD4
Esrrb Sptlc2 DNA contact
Esrrb Sptlc2 RNA present
BRD4 present
No DNA contact
No RNA present
No BRD4 present
b
3DSE CRE Community: Esrrb/Irf2bpl/Sptlc2
200 kb
H3K27ac HiChIP Resolution: 5Kb
Loop Resolution: 5 Kb
Insulated Neighborhoods
Community Genes
Esrrb
Irf2bpl
Sptlc2
Community Enhancers
MED1 ChIP-seq
BRD4 ChIP-seq
482
0
1110
0
c
DAPI stain+DNA FISH
DAPI
Esrrb DNA
Sptlc2 DNA
DAPI
Esrrb DNA
Sptlc2 DNA
5 µm
DAPI stain+RNA FISH+BRD4 IF
DAPI
Esrrb RNA
Sptlc2 RNA
BRD4
DAPI
Esrrb RNA
Sptlc2 RNA
BRD4
d
Percentage of Regions-of-Interest
100%
80%
60%
40%
20%
0%
BRD4-
BRD4+
Co-expressed RNA, Co-localized DNA
Single RNA, Single DNA
No RNA, Co-localized DNA
No RNA, Single DNA
RNA-
RNA+
Sptlc2 DNA+mRNA
Esrrb DNA+mRNA
Sptlc2 DNA + BRD4
Esrrb DNA + BRD4
e
Odds ratio
1.0
0.6
0.2
Non-community CRE pairs
3D CRE Comm. CRE pairs
3DSE CRE Comm. CRE pairs
Linear SE CRE pairs
f
Spearman coefficient
0.10
0.05
0.00
−0.05
All gene-pairs
3D CRE community gene-pairs
3DSE CRE community gene-pairs

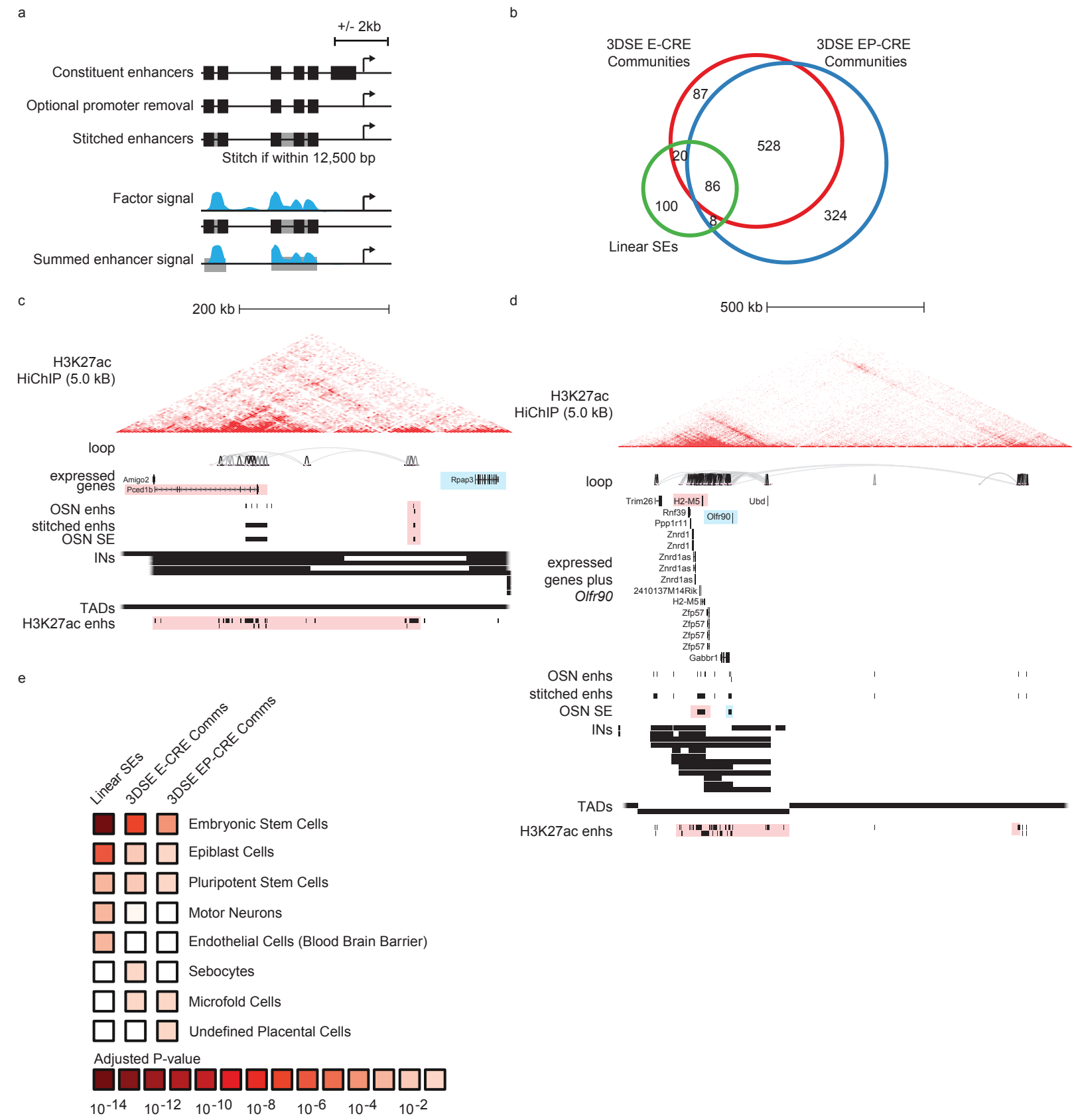
Supplemental Figure 1



Supplemental Figure 2

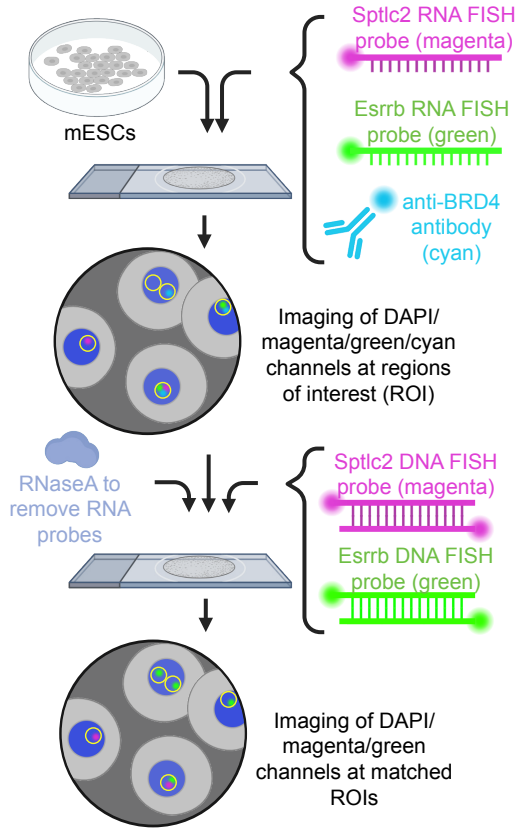


Supplemental Figure 3

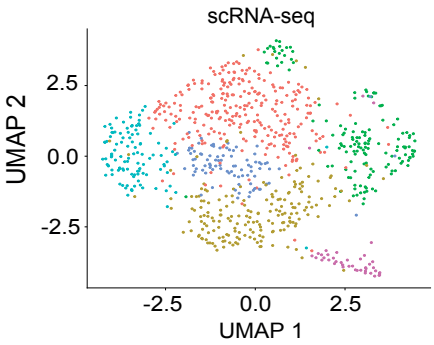


Supp Fig 4

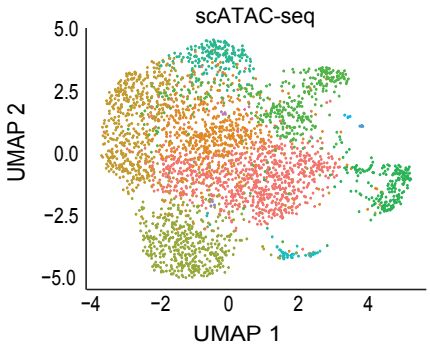
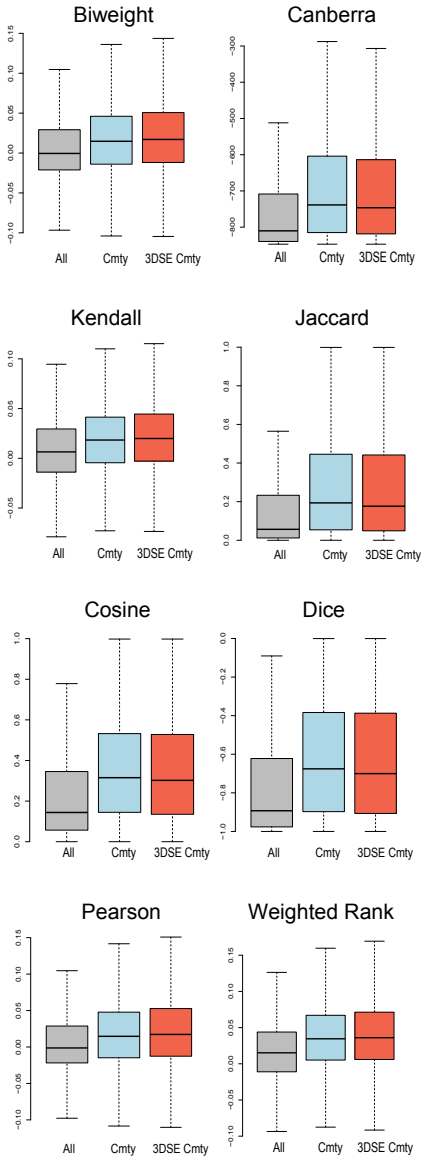
a



b



d



c