Deciphering Hierarchical Chromatin Structures and Preference of Genomic Positions in Single Mouse Embryonic Stem Cells

Yusen Ye[1*], Shihua Zhang[2,3,4], Lin Gao[1], Yuqing Zhu[5], Jin Zhang[6,7,8,9]

1School of Computer Science and Technology, Xidian University, Xi'an 710071, Shaanxi, China.
2NCMIS, CEMS, RCSDS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China.
3School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China.
4Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China.
5Center for Stem Cell and Translational Medicine, School of Life Sciences, Anhui University, Hefei, 230601, Anhui, China.
6Center for Stem Cell and Regenerative Medicine, Department of Basic Medical Sciences, and Bone Marrow Transplantation Center of the First Affiliated Hospital, Hangzhou, Zhejiang, China.
7Zhejiang Laboratory for Systems and Precision Medicine, Zhejiang University Medical Center, Hangzhou, Zhejiang, China.
8Institute of Hematology, Zhejiang University, Hangzhou, Zhejiang, China.
9Center of Gene/Cell Engineering and Genome Medicine, Hangzhou, Zhejiang, China.

*To whom correspondence should be addressed. Tel: +86-15802963380; Email: ysye@xidian.edu.cn

Abstract

Recent developments in the three-dimensional (3D) genome have revealed chromatin organization with hierarchical architectures at different genomic scales. Transcriptional regulators and their specific interactions implicate gene expression programs by complex chromatin structures. The exploration of single-cell 3D genome maps reveals that chromatin domains are indeed physical structures presenting in single cells and domain boundaries vary from cell to cell. However, the formation mechanism and functional implication of hierarchical chromatin structures in single cells is still blurry. To this end, we first develop a **hi**erarchical **c**hromatin **s**tructure identification algorithm (named as **HiCS**) from individual single-cell Hi-C maps, with superior performance in both accuracy and efficiency. The enrichment of hundreds of regulatory factors in domain boundaries suggests that in addition to the CTCF-cohesin complex, Polycomb, TrxG, pluripotent protein families and other multiple factors also contribute to shaping chromatin landscapes in single embryonic stem cells. We further decipher the preference of genomic positions-function relationship by grouping and annotating chromatin landscape clusters with different types of regulatory factors, and the relationship corresponds to precisely the largest five types of retrotransposons. Our results suggest that multiple types of regulatory factors interplaying with each other in specific genomic positions may navigate the preference of genomic positions, and then shape chromosome architecture and cell functions of single cells.

Introduction

Genome across a wide range of eukaryotic organisms is efficiently packaged and organized into hierarchical chromatin structures via ubiquitous architectural features, which is critical to gene regulation and dynamical changes in development and disease [1, 2]. These basic features consist of chromatin fibers, which fold into chromatin loops, such as enhancer-promoter interactions and architectural loops mediated by CCCTC-binding factor (CTCF) [3]. These fibers further fold into chromatin domains, referred to as topologically associating domains (TADs) or sub-TADs, which are associated with each other to generate chromosomal compartments. Each chromosome occupies a distinct volume or chromosome territory within the nucleus [3-5]. Genome architecture is an integral part of the chromatin landscape that transcription factors (TFs) must navigate to exert their regulatory roles [6]. Although most loci and chromosomes are characterized by a high degree of order and non-randomness, the precise functional roles and formation mechanism of these features remains obscure [7].

Some TFs, cofactors, and histone modifications that correlate with the chromatin structures have been identified to study features of chromatin organization [8-10]. For example, TAD boundaries are enriched with multiple factors including CTCF, cohesin, H3K4me3, H3K36me3, transcription start sites, housekeeping genes, etc, suggesting that CTCF binding, high levels of transcription activity, multiple histone modifications, and other regulatory factors may contribute to the formation of chromatin structures in mammals [7]. The Polycomb group (PcG) proteins are also associated with both local chromatin compaction and silencing as well as long-range chromatin interactions, which maintain a permissive regulatory for future activation [11].

Although TADs and their associated regulatory factors have been widely identified in multiple species and are highly conserved and stable across different cell types [5, 12], single-cell 3D genome analysis indicated that they display substantial cell-to-cell variation [13, 14]. Therefore, the bulk analysis only reflects properties of ensemble average structures from millions of cells, which may mask chromatin features appearing in a few cells or a single cell and limit our understanding of chromatin structures. A recent study reveals that domain structures often adopt globular conformation with strongly physical segregation of neighboring domains, and domain boundaries are preferentially located at CTCF- and cohesin-binding sites with a super-resolution chromatin tracing method [13]. More surprisingly, single-cell domain structures persist even after cohesin degradation [13]. These results suggest other TFs or epigenetic factors may contribute to the formation of chromatin structures. Currently, mouse embryonic stem cells (mESCs) have been served as a specific model cell system to elucidate the mechanisms of 3D genome organization and explore the relationship between chromatin structures and gene regulation. Hundreds of TFs and epigenetic modification profiles have been identified in mESCs [8-10]. The above observations push us to investigate the formation of hierarchical chromatin structures and their relationship with functional elements in single mESCs systematically.

Here, we first develop a **hi**erarchical **c**hromatin **s**tructure identification algorithm (named as **HiCS**) from single-cell Hi-C maps, which shows superior performance in both accuracy and efficiency. We reorganize atlas of ChIP-seq for mESCs, and reveal the patterns of hundreds of regulatory factors are either enriched or not in domain boundaries of single cells, suggesting that, in addition to CTCF-cohesin complex, Polycomb, TrxG, pluripotent protein families, different types

of histone modifications, and other multiple factors could promote the formation of chromatin structures in single mESCs. To explore how different types of regulatory factors are organized to drive chromatin structures, we group 13 large spatially organized chromatin landscape categories (consisting of 29 sub-categories) and annotate them with 7 different regulatory factor clusters (consisting of 27 sub-clusters). The clear patterns illustrate the emergence and functionality of chromatin structures, and provide a detailed view of the preference of genomic positions-function relationship. Furthermore, we discover that different retrotransposons interplaying with regulatory factors contribute to the maintaining of chromatin states and exactly match the above chromatin landscapes categories. Last but not the least, we find that genomic positions enriched by Alu/B2/B4 retrotransposons have higher preference scores for forming boundaries in G1 and ES phases in comparison with MS and LS/G2 phases, whereas genomic positions enriched by L1/ERVK retrotransposons display opposite tendency. In summary, we reveal that multiple types of regulatory factors interplaying with each other in specific genomic position could affect focal chromatin interactions, thereby changing interaction density or insulation strength of these regions. This further navigates the preference of genomic positions for forming boundaries, shape hierarchical chromosome architecture, and thus regulate gene expression and cell functions, even cell identity in single embryonic stem cells.

Results

Overview of HiCS. The key design of HiCS is to convert the problem of the identification of hierarchical chromatin structures or domains into finding peaks of insulation strength at different genome scales. The domain boundaries usually have higher insulation strength than their neighbors and a relatively large distance from any regions with higher strength (**Fig. 1a**). HiCS calculates two metrics for each bin including the insulation strength ρ and the minimum distance between the bin and any other bin with higher strength δ, and controls the number of peaks to obtain the hierarchical chromatin domains at different scales by α (**Fig. 1b and c**). Thus, HiCS is super-fast to identify a chromatin hierarchy that the structure of a higher level embraces the multiple smaller ones of a lower level (**Fig. 2b**).

Specifically, HiCS consists of four steps (**Methods**): (1) Implements a classic graph embedding method node2vec to impute the single cell Hi-C map (**Fig. 1a**); (2) Calculates the two metrics to find the peaks at a given scale α, and determines the domain boundaries for each chromosome of an individual cell (**Fig. 1b and c**); (3) Determines the hierarchical structures of a chromatin by adjusting the optimal structural identification parameter (**Fig. 1c**, **Fig. 2a, b**, and **Supplementary Fig. S1a**); (4) Applies a bi-clustering method to group chromatin positions and regulatory factors respectively and analyze the genomic structure-function relationship by combining hundreds of TFs and epigenetic factors with hierarchical chromatin structures (**Fig. 4**).

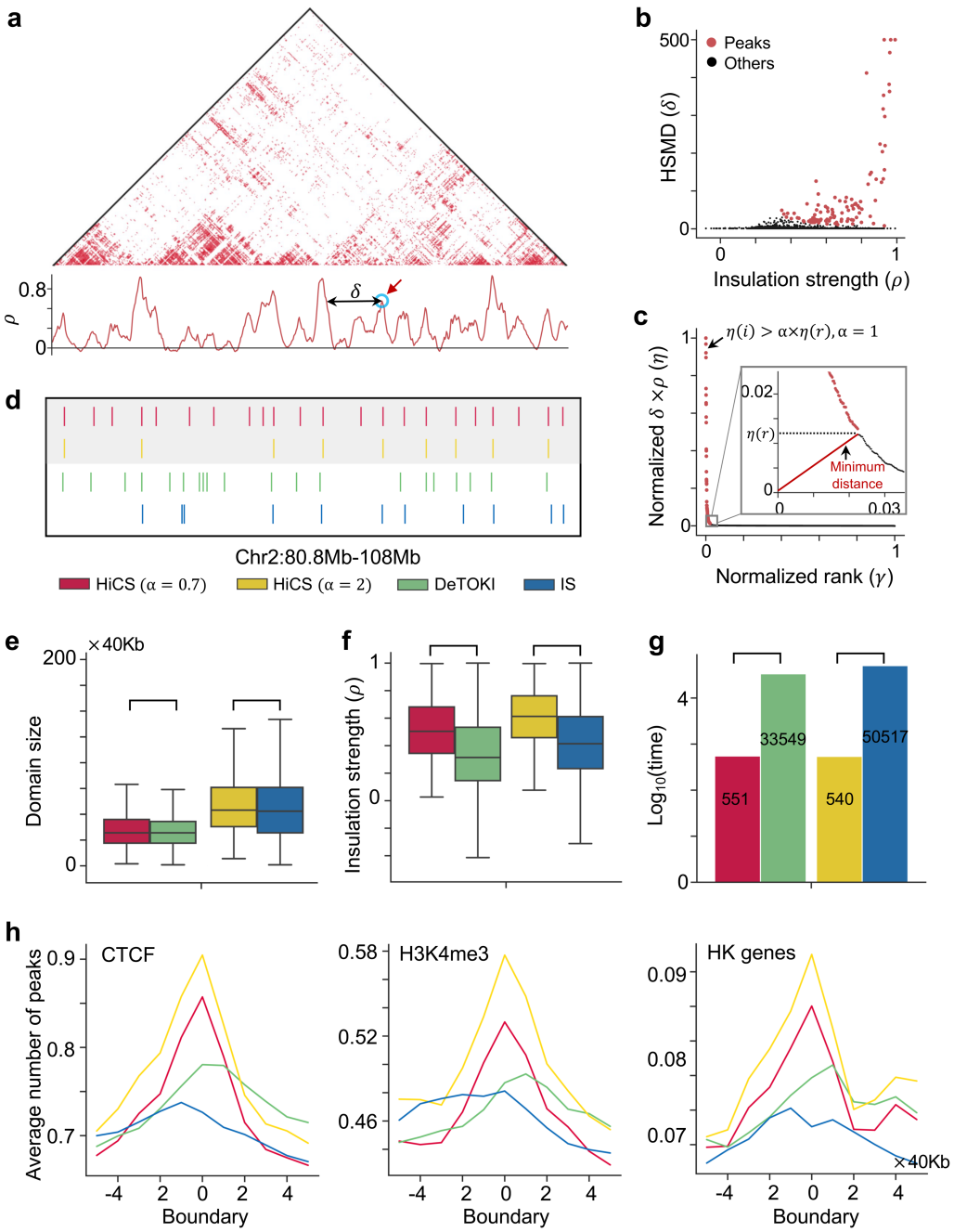


Figure 1. Illustration and efficiency of HiCS for determining the chromatin structures. a. An illustrative example of the preprocessed single-cell Hi-C contact map (top) and the insulation strength (ρ) of genomic positions (bottom). HSMD (δ) represents the minimum distance between the bin and any other bins with higher strengths. **b, c.** The decision graph (b) and the normalization value of $\eta = \rho \times \delta$ in a decreasing order (c) for the domain boundaries (colored in red) in the optimal structural identification parameter. The decision graph of optimal structural identification parameter (zoom box in c). **d.** A local example of domain boundaries at the region Chr2:80.8Mb-108Mb. **e, f.** Comparison of domain sizes and insulation strengths on Chromosome 2. **g.** Runtime (seconds) of different methods or parameters on Chromosome 2. **h.** The average number of CTCF peaks, H3K4me3 peaks, and HK genes at domain boundaries of single cells. The above results (d-h) are implemented by different methods across all single cells.

HiCS shows superior performance. We benchmark the performance of HiCS for domain detection against two methods: one is the commonly used method for bulk data, insulation score (IS) [15], and another is a single-cell TAD detection method, deTOKI [16]. We apply these methods to the preprocessed single-cell Hi-C data generated from mESCs [17]. To compare the performance of them fairly, we adjust the scaling parameter of HiCS to obtain comparable domain sizes with IS and deTOKI, respectively (**Fig. 1e**), with an example, we can see that HiCS can obtain more accurate chromatin structure boundaries at single-cell resolution (**Fig. 1a and d**). Actually, the insulation strengths of domains obtained from HiCS is significantly higher than those of IS and deTOKI respectively, suggesting its superiority to competing methods (**Fig. 1f**). The running time of HiCS is significantly less than both algorithms under the same hardware condition (**Fig. 1g**). Moreover, the domain boundaries detected by HiCS are more significantly enriched in multiple common factors, including CTCF, H3K4me3, Housekeeping (HK) genes (**Fig. 1h**), as well as RNA polymerase II (PolII), promoters, highly expressed genes, and average phastcon score (**Supplementary Fig. S1b**). Also, the boundaries detected by HiCS are more pronounced disappearance of the enhancers or the super enhancers (SEs), which unfavorably form boundaries in single cells (**Supplementary Fig. S1b**). Taken together, HiCS shows superior performance in both accuracy and efficiency.

The existence of hierarchical chromatin structures. We adjust the optimal structural identification parameters to generate multiple-scale chromatin structures at different genomic scales for 1315 single mESCs at 40kb resolution. We clearly observe four peaks of domain size and insulation strength distribution with different scaling parameters of 0.2, 1, 4, and 8, which we choose for the downstream analysis (**Fig. 2a, b** and **Supplementary Fig. S1a**). The domain scales of these four levels are approximately 200Kb~600Kb, 800Kb~1Mb, 2Mb~3Mb, and ~5Mb respectively. We show an example for hierarchical chromatin structures, which well match the local insulation strength of chromatin regions (**Fig. 2b**). Note that high-level boundaries with high δ may have lower insulation strength than low-level ones, and the boundaries in the same level may have different local insulation strengths (**Fig. 2b**).

The median number of boundaries identified by HiCS on chromosome 2 (Chr2) of single mESCs is about 161.5, 54, 18, and 30 for different genomic levels (**Supplementary Fig. S2a**). The median size of chromatin structures increases and the median boundaries' insulation strength enhances with the increase of genomic position level (**Supplementary Fig. S2b, c**). The boundaries show obvious cell-to-cell heterogeneity with a nonzero probability of being located at any genomic positions. 99.8% of genomic positions form boundaries in at least 1% of cells, while only 4.1% of genomic positions form boundaries in more than 14% of cells, and the highest number of genomic locations form boundaries in 5% of cells (**Fig. 3a**). We also observe that the probability forming boundaries enhances with the levels of genomic position increasing (**Supplementary Fig. S2d**). The above observations suggest that domain boundaries vary from cell to cell with nonzero probability at all genomic positions, and the preference of genomic positions for forming boundaries may shape the formation of hierarchical chromatin structures in single cells.

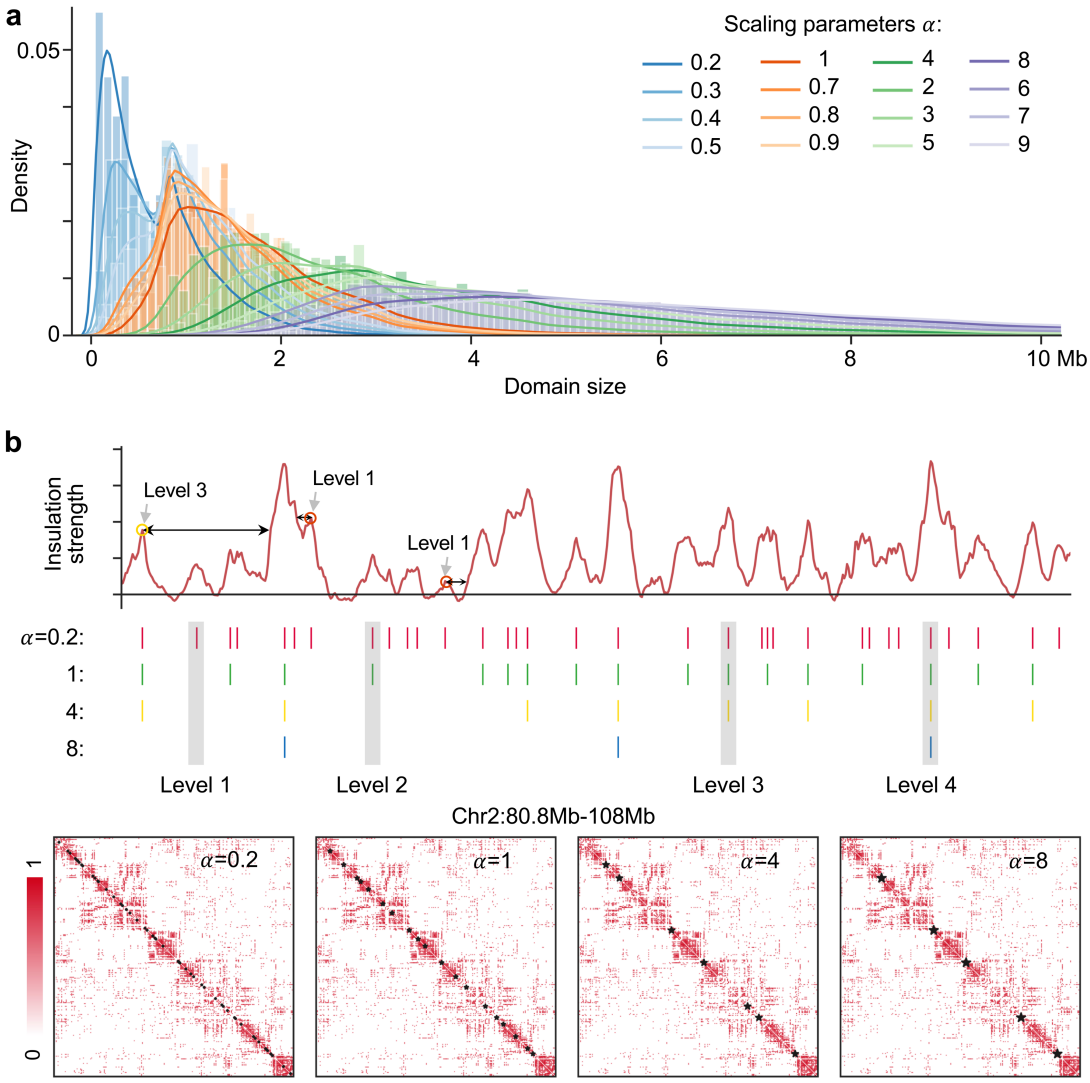


Figure 2. Identification of hierarchical chromatin structures. a. Density distribution of domain sizes at different multi-scale parameters. **b.** An example for hierarchical chromatin structures at the region Chr2:80.8Mb-108Mb. Circles mark three boundaries (two boundaries in level 1 have a magnitude difference in local insulation strength, and the boundary in level 3 has lower insulation strength than one of the boundaries in level 1).

Furthermore, we observe that the accessibility of genomic positions, the enrichment degree of HK genes and differentially expressed genes (DEGs), and the expression of genes gradually increases along with the increasing levels of genomic positions (**Fig. 3b, c and Supplementary Fig. S2e**). Conversely, enhancers and super enhancers (SEs) greatly become less enriched with the increase of levels (**Fig. 3b and c**). The emergence of highly transcriptional activity, especially DEGs, and the absence of enhancers, especially SEs, improve the probability of genomic positions forming boundaries of single cells, and thus shape hierarchical chromatin structures (**Fig. 3b and c**).

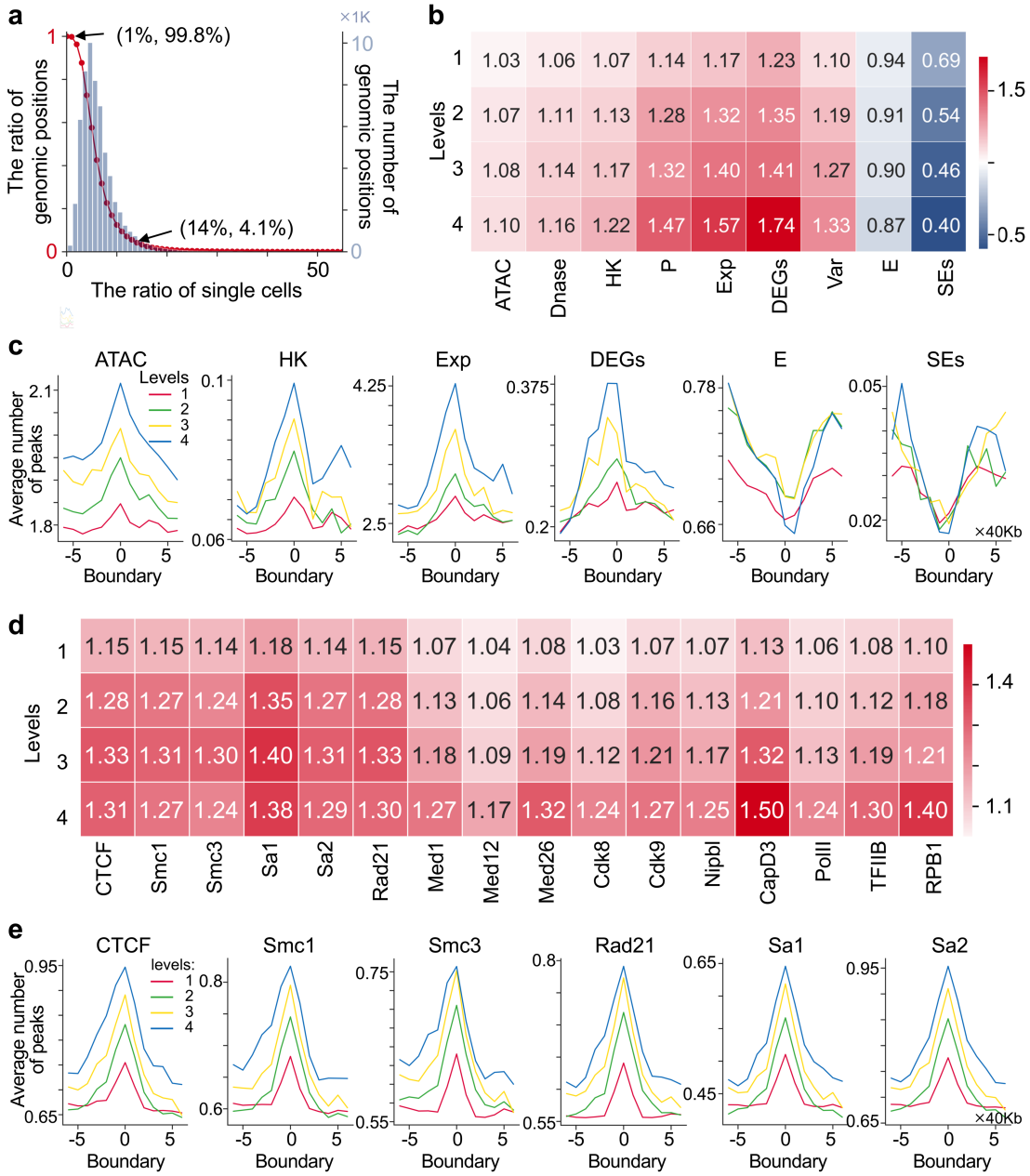


Figure 3. The existence of hierarchical chromatin structures. a. The ratio distribution of genomic positions forming boundaries at less than a given ratio of single cells in the left y-axis (such as 99.8% of genomic positions form boundaries in at least 1% of cells, respectively), and the number distribution of genomic positions forming boundaries in the right y-axis as the ratio of single cells increases (such as 1K genomic positions from boundaries in 5% of cells, respectively). **b-c.** The concentration scores (b) and the average number of ATAC-seq peaks and multiple regulatory elements (ATAC: ATAC-seq peaks, HK genes, Exp: gene expression value, DEGs: differentially expressed genes, E: enhancers, SEs) (c) around domain boundaries. **d.** The concentration scores for CTCF, cohesin, mediator- and Pol II-associated factors. **e.** The average number of peaks for CTCF and cohesin around domain boundaries. In b-e, the results were detected in the different genomic scales across all single cells. The concentration score is defined in **Supplementary Methods**.

Next, we check the enrichment of hundreds of TFs, chromatin regulators, and histone modifications at the boundaries of chromatin structure at the single-cell level. CTCF and cohesin have been proven to be key factors controlling the functional architecture of mammalian chromosomes forming TADs or sub-TADs by "loop extrusion" [18]. We observe that both CTCF and cohesin show similar enrichment patterns in boundaries of different genomic levels. The formation of boundaries of levels 2-4 requires almost the same concentration scores of CTCF and cohesin, while that of level 1 shows a significantly lower score than those of levels 2-4 (**Fig. 3d and e**), which potentially suggest some specific structure-function relationships.

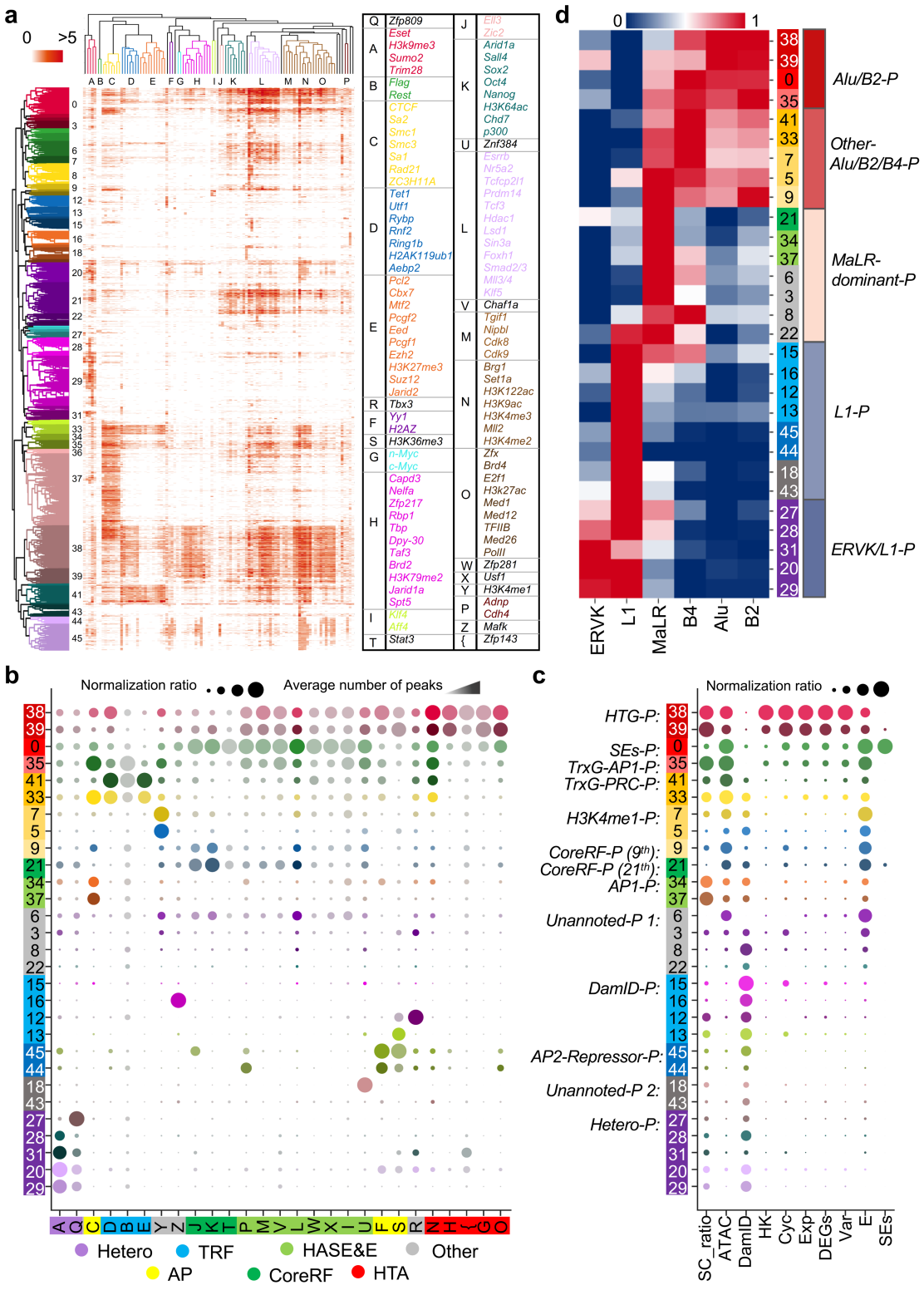
Although the CTCF-cohesin complex is critical for the formation of TADs in mammalian cells, a substantial number of boundaries remain unaffected after cohesin degradation in single cells, suggesting other modulators exist on domain boundaries [6, 19]. We indeed observe that mediators (*Med1*, *Med12*, *Med26*, *Cdk8*, *Cdk9*), *Nipbl*, PolII, TFIIB are all enriched in the boundaries of single cells, and as the level of boundaries increases, the degree of concentration scores gradually increases. Mediators are essential coactivators that are recruited to the regulatory regions of active genes and facilitate the ability of enhancer-bound TFs to recruit PolII to the promoters of target genes, and *Nipbl* has been proven to bind mediator to load cohesin and condensin at enhancers or promoters in ESCs [8, 9, 20]. The above results suggest that these factors may play important roles in shaping hierarchical chromatin structures of single cells.

The master factors of Polycomb repressive complex 1 (PRC1) and PRC2 just exhibit two different enrichment patterns at the domain boundaries of single cells. One type (*Aebp2*, *Rybp*, *Ring1b*, *Rnf2*) forms obvious single peaks at the domain boundaries, and the degree of enrichment gradually increases with the boundary level increasing (**Supplementary Fig. S3a**). Another type (*Ezh2*, *Pcl2*, *Suz12*, *Eed*) shows double peaks around the domain boundaries (**Supplementary Fig. S3b**). These two types of complexes have been proved to have distinct catalytic activities, but both are generally associated with transcriptional silencing [21]. We also observe that TrxG associated proteins (COMPASS: *Set1a*, *Mll2*, *Mll3/4,* and SWI/SNP: *Brg1*) are all enriched in domain boundaries, except for Mll3/4 (**Supplementary Fig. S3c**). In mammalian, *Set1a* reportedly contributes to most of the H3K4me3, and *Mll2* mediates H3K4me2 and H3K4me3 at developmental genes, while *Mll3/4* implements monomethylation of H3K4 at enhancers [22]. *Brg1* is an ATP-dependent chromatin remodeler, contributing to the maintenance of pluripotency and self-renewal in ESCs [8]. The above results suggest that the PRC and TrxG protein families could change focal chromatin interactions in different ways.

We find that the core TFs (*Oct4*, *Sox2*, and *Nanog*) controlling the pluripotent state do not prefer to appear in domain boundaries, which may be related to chromatin hubs occupied by super-enhancers/enhancers (**Supplementary Fig. S4a**) as reported in [23]. In addition, we also collect 14 additional TFs that may contribute to the pluripotent state of mESCs and investigate whether they are enriched in domain boundaries of single cells [8] (**Supplementary Fig. S4b and c**). The results indicate that 9 additional TFs (*Esrrb*, *Nr5a2*, *Klf4*, *Zfp281*, *Tcf3*, *Tcfcp2l1, Stat3*, *Prdm14*, and *Smad2/3*) that were previously shown to occupy both typical enhancers and SEs do not prefer to appear in domain boundaries, while 5 factors (*c-Myc*, *n-Myc*, *Zfx*, *Tbx3*, and *Yy*1) that were previously shown to occupy promoter-proximal sites are enriched in domain boundaries [8]. Among them, it is particularly interesting that *Smad2/3*, *Stat3*, and *Tcf3* signaling pathways were considered

242 as key modulators controlling mESCs pluripotent state transition by modifying chromatin states and
243 shaping chromatin structures [24-26].
244 In addition, we also observe additional 45 proteins are either enriched or not in domain
245 boundaries of single cells with different enrichment patterns (**Supplementary Figs. S5-7**), which
246 suggests their potential in shaping chromatin structures.
247 The above TFs and chromatin regulators have the most profound impact on cell states through
248 collaborative control of chromatin states and spatial structures. The enrichment of 14 histone-
249 modifying enzymes around domain boundaries of single cells, indicates that different histones show
250 a variety of enrichment patterns (**Supplementary Fig. S8**). For example, ZC3H11A shares
251 consistent enrichment patterns with CTCF-cohesin (**Fig. 3e**). H3K27me3 shares similar enrichment
252 patterns with PRC2, while H2AK119ub1 shares similar enrichment patterns with PCR1
253 (**Supplementary Figs. S3a and b**), which are consistent with their function in participating
254 chromatin modifications [21]. H3K4me1 marking of enhancers is not enriched in domain
255 boundaries. And H3K4me2 shares similar enrichment patterns with H3K4me3 around domain
256 boundaries, which all are associated with TrxG protein family (**Supplementary Fig. S3c**). It
257 suggests different histone modifications may cooperate with different TFs and chromatin regulators,
258 to modify chromatin states, shape the local chromatin interaction status and organize chromatin
259 spatial structures of single cells.
260 To sum up, we have observed hundreds of TFs, chromatin regulators, and histone
261 modifications are either enriched or not in domain boundaries of single cells, which show
262 differential enrichment patterns. The occupancy of these regulatory factors in specific genomic
263 positions will affect focal chromatin interactions, thereby changing the interaction density or
264 insulation strengths of these regions. These processes may navigate the preference of genomic
265 positions for forming boundaries, and then shape hierarchical chromosome architecture of single
266 cells.
267
268 **Regulatory factors navigate the preference of genomic positions.** To further elaborate on the
269 dynamic interplay between transcription factors, chromatin-associated proteins, epigenetic
270 modifications, and a spatially organized genome, we grouped 13 large spatially organized chromatin
271 landscapes categories (consisting of 29 sub-categories), which were annotated by 7 different
272 regulatory factor clusters (consisting of 27 sub-clusters). The result helps to explain the preference
273 of genomic positions for forming boundaries in single cells, and providing an increasingly complex
274 view of the genomic structure-function relationship (**Fig. 4 and Supplementary Fig. S9**).
275 ***The categories of regulatory factors.*** The result of hierarchical clustering indicates that these
276 categories mainly consist of the cluster of core regulatory factors (CoreRF), the clusters associated
277 with highly active activators for SEs and enhancers (HASE&E), the clusters of highly transcribed
278 activators (HTA), the clusters of transcriptional repressor factors (TRF), the cluster associated with
279 heterochromatin factors (Hetero), and the clusters of architectural proteins (AP). There are also a
280 few single factors (Other) (**Fig. 4a and Supplementary Fig. S9**). For a specific example with less
281 prior studies among these clusters, *Sumo2* is required to play critical roles in the canonical
282 *Zfp809/Trim28/Eset* complex via post-translational sumoylation of *Trim28*, which enhances the
283 recruitment of *Trim28* to the proviral DNA, resulting in the modification of proviral chromatin with

repressive histone H3K9me3 marks in turn [27]. These four factors were grouped with H3K9me3 together in the analysis below, which may organize the formation of heterochromatin (**Fig. 4a**). The annotation information and supporting materials of these all categories were summarized in **Supplementary Table S2**.



288

Figure 4. The systematic analysis on the preference of genomic positions. a. The bi-cluster of chromatin positions and regulatory factors. **b.** The column normalization ratio and the average number of peaks for regulatory factor-classes on the positions of chromatin landscape-classes. **c.** The column normalization ratio of different elements or factors, including SC_ratio (the ratio of single cells forming boundaries), ATAC, DamID (DamID-seq), HK genes, Cyc (mark genes for cell cycle), Exp (gene expression value), DEGs, Variable (variable scores for genes), E (Enhancers), and SEs, on the positions of chromatin landscape-classes. **d.** The heatmap shows the column normalization ratio of different retrotransposons based on different chromatin landscape-classes.

We also observe some interesting cooperative patterns among different clusters (**Supplementary Fig. S9b**). For example, PCR1 is strongly associated with HASE&E and HTA, especially with TrxG proteins, which is different from PCR2. Previous studies reported that both PRC and TrxG proteins were critical for orchestrating the stable activation and repression of gene expression, respectively, and maintaining a balanced state of gene expression [28]. AP2 showing an apparent discrepancy with AP1 is closely related to the HTG cluster, PRC1 class, and H3K36me3. The observation is consistent with the report in a recent study that *Yy1* is a structural regulator of enhancer-promoter loops, and contributes to transcription repressor associated with non-canonical PRC1, whereas CTCF preferentially occupies sites distal from these regulatory elements that tend to form larger loops and participate in insulation [21, 29]. In addition, the CoreRF cluster is strongly associated with the HASE&E cluster, but not the HTG cluster, while the HASE&E cluster is intimated to the HTG cluster. It suggests the HASE&E cluster may be a bridge between CoreRF and HTG clusters, which can be linked to various signaling pathways involved in the transcriptional network in mESCs.

Regulatory factors navigate the division of functional groups. We further analyze the preference for different classes of genomic positions with a ratio>1% in single cells (**Supplementary Fig. S10a**). We check the open genomic positions with high chromatin accessibility, and intuitively divide these chromatin regions into two categories by the ratio of single cells occurring boundaries (**Fig. 5a** and **Supplementary Fig. S10b**). One type preferentially forms boundaries of single cells, and the other is the opposite.

Of note, the largest type with weak open chromatin accessibility (13.7% of all positions, consisting of 12.6%, 1.1% from the 37th, 34th classes, respectively, defined as AP1-Positions) preferentially forming boundaries only are significantly occupied by AP1 factors (CTCF, cohesin, etc.) comparing with other types of factors (**Fig. 4b, c** and **Fig. 5a**). Previous studies have indicated that loop extrusion executed by CTCF and cohesin is a leading factor governing domain formation and facilitating chromatin folding. We find that the other subunits (*Sa1* and *Sa2*) of cohesin which occupy the same genomic locus and present similar enrichment patterns at boundaries of single cells, suggesting that both *Sa1* and *Sa2* may also participate in the maintenance of domain boundaries (**Fig. 3e**). ZC3H11A, a zinc finger protein, shows a uniform pattern as *Sa1* and *Sa2,* implying its novel role in shaping boundaries of chromatin domains (**Supplementary Fig. S8**). The chromatin positions that AP1 proteins occupy may be directly related to the macroscopic architecture of chromatin within the nucleus, and indirectly change the local chromatin context to exert regulatory functions.

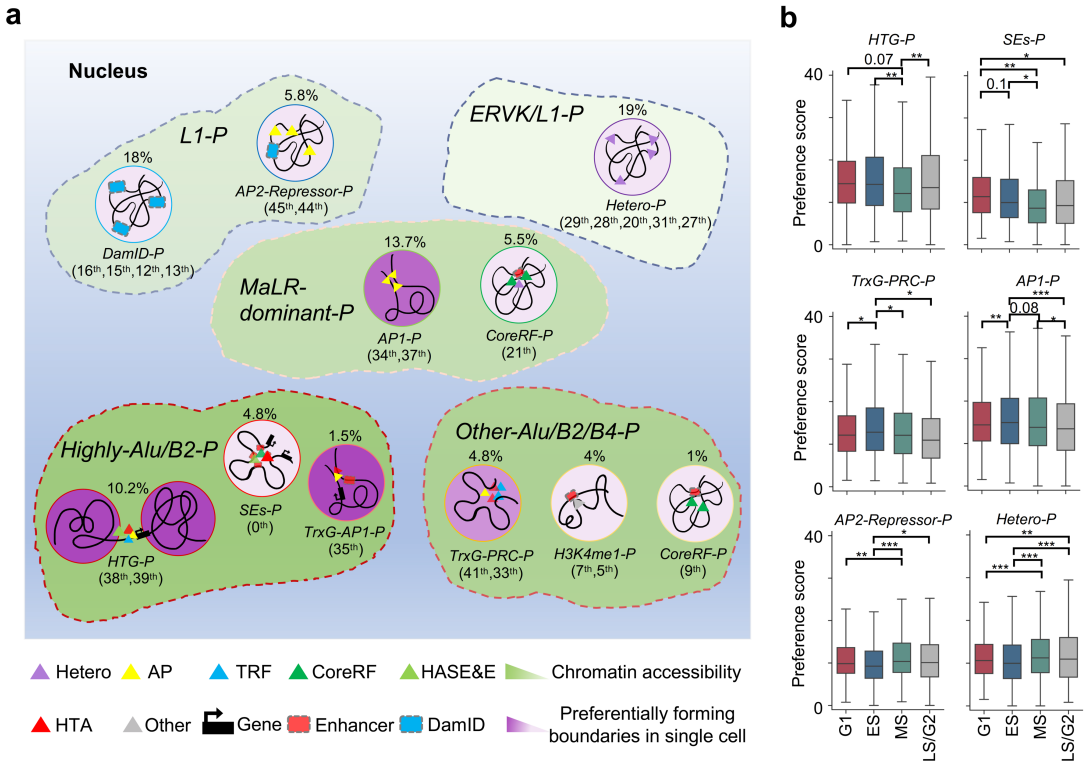


Figure 5. A schematic showing the preference of genomic positions. a. The graph presents the different clusters based on retrotransposons in the area marked by different colors of dotted line containing multiple functional groups associated with different regulatory factor-classes (the sub-categories serial number in above and the ratio occupying from all positions are marked for each functional group). The shades of purple on the background of circles indicate the preference of genomic positions forming boundaries in single cells for each functional group, and the shades of green on the background of these areas marked by dotted line indicate the degree of chromatin accessibility of genomic positions for each cluster based on retrotransposons. **b.** Preference scores of different chromatin landscapes categories across different cell states. Statistical significance is calculated by Welch's t-test (*P < 0.05, **P < 0.01, and ***P < 0.001). The preference score is defined in **Supplementary Methods**.

The second major category of open chromatin positions preferentially forming boundaries (7.5%, 2.7% from the 38th, 39th classes, defined as HTG-Positions) is associated with highly transcribed genes (**Fig. 4b, c** and **Fig. 5a**). These regions are mainly occupied by the HTA (particularly the TrxG class, playing important roles in orchestrating the stable activation of gene expression) and HASE&E clusters, but not the CoreRF cluster, which indicates that the CoreRF cluster hardly participates in the regulation of genes occupying in domain boundaries of single cells. Surprisingly, we also observe that other factors are weakly enriched in these chromatin positions, such as the H3K6me3, PRC1 (not PRC2), AP2 (not AP1) classes. The above results indicate that highly transcribed genes may promote the formation of chromatin structures, and AP2 and PRC1 may cooperate to maintain the architecture and epigenetic balance of domain boundaries [21].

352 The third major category of open chromatin positions preferentially forming boundaries (3.1%,
353 1.7% from the 41th, 33th classes, defined as TrxG-PRC-Positions, 1.5% from the 35th class, defined
354 as TrxG-AP1-Positions) is associated with PRC, TrxG, and AP1 proteins (**Fig. 4b, c** and **Fig. 5a**).
355 The 33th class of genomic regions is also associated with AP1 proteins with a higher probability of
356 forming domain boundaries of single cells compared with the 41th class, which suggests CTCF-
357 cohesin complexes may help these regions form a more stable structure. In addition, the 35th class
358 of genomic region with higher enrichment of TrxG proteins and lower enrichment of PRC proteins
359 compared with the 33th class, which results in higher open chromatin accessibility, enhancers
360 occupation, and probability forming domain boundaries in single cells. The above results indicate
361 that chromatin modifiers (TrxG and PRC proteins) can provide an additional layer of regulation by
362 changing chromatin structures, and balance the formation of domain boundaries by repressing and
363 activating chromatin states, respectively.
364 The major categories of the open chromatin positions unfavorably forming boundaries are
365 associated with highly active SEs or enhancers, which mainly contain three categories (**Fig. 4b, c**
366 and **Fig. 5a**). The first category (4.8% from the 0th class, defined as SEs-Positions) almost
367 encompasses more than 75% of SEs, which are strongly related with the HASE&E, CoreRF, TrxG,
368 and Med-PolII clusters. The second category (5.5%, 1% from the 21th, 9th classes, defined as
369 CoreRF-Positions) is mainly associated with enhancers occupied by CoreRF. The slight difference
370 between the 21th and 9th classes is that the 9th one shows higher chromatin accessibility, while the
371 21th one shows a stronger DamID enrichment, which may be related to their locations in the nucleus.
372 The third category (2.7%, 1.3% from the 7th, 5th classes, H3K4me1-Positions) are related to
373 enhancers enriched by H3K4me1. These regions taking SEs or enhancers associated with CoreRF
374 and H3K4me1 as focal regions generate dense chromatin structures mediated by different regulatory
375 factors in an orientation-independent manner, and unfavorably form boundaries in single cells with
376 weakened insulation strengths [30].
377 We also check the classes of genomic positions with low chromatin accessibility, all of which
378 unfavorably form boundaries in single cells. These chromatin positions are mainly divided into the
379 following categories (**Fig. 4b, c** and **Fig. 5a**). The first category (19% of all positions, consisting of
380 9.2%, 3.7%, 3.5%, 1.6%, and 1% from the 29th, 28th, 20th, 31th, and 27th classes, respectively, defined
381 as Hetero-Positions) are associated with the Hetero cluster. The second category (18% of all
382 positions, consisting of 2.8%, 2.3%, 1.9%, and 1.7% from the 16th, 15th, 12th, and 13th classes,
383 respectively, defined as DamID-Positions) maintains high enrichment of the DamID signal. The
384 above observations suggest that the presence of heterochromatin factors reduce the probability of
385 single-cell domain boundaries formation. Finally, we observe an interesting and specific category
386 (4.6%, 1.2% from the 45th, 44th classes, defined as AP2-Repressor-Positions), which are associated
387 with the AP2 (*Yy1* and H2AZ) and H3K36me3 classes (**Fig. 4a** and **Fig. 5a**). Both *Yy1* and H2AZ
388 are involved in maintaining poised chromatin and facilitate the organization of genome architecture
389 (**Supplementary Table S2**) [29, 31-33].
390 To summarize, we obtain the following key results by the above analysis: (1) The genomic
391 positions occupied by architectural proteins (CTCF and cohesin), highly transcribed genes, and
392 TrxG proteins preferentially form boundaries in the single cells analysis. (2) The genomic positions
393 taking SEs or enhancers associated with CoreRF or marked by H3K4me1, heterochromatin factors,

394 and repressing factors unfavorably form boundaries in single cells with weakened insulation
395 strengths. (3) Different from CTCF and cohesin, AP2 (*Yy1* and H2AZ) can contribute to the
396 formation or disappearance of boundaries by preferentially occupying interacting enhancers and
397 promoters, and ultimately regulate the balance between gene activation and inhibition. (4) PRC1
398 participates in the formation of domain boundaries supported by highly transcribed genes, while the
399 same phenomenon is not detected in PRC2.
400
401 **Retrotransposons are associated with these functional groups.** We observe that the division of
402 functional groups is strongly associated with the regulation of retrotransposons. The genomic
403 positions are clearly divided into five functional units based on categories of retrotransposons,
404 including Highly-Alu/B2-Positions, Other-Alu/B2/B4-Positions, MaLR-dominant-Positions, L1-
405 Positions, and ERVK/L1-Positions (**Fig. 4d** and **Fig. 5a**).
406 Firstly, all genomic regions (Hetero-Positions, DamID-Positions, AP2-Repressor-Positions)
407 with low accessibility and unfavorably forming boundaries are enriched by L1 elements, which tend
408 to occupy gene-poor, heterochromatic B compartments that interact with lamina-associated domains
409 in previous studies [34]. Among the defined regions, we are surprised to find that Hetero-Positions
410 are specially associated with ERVK elements, which may indicate that ERVK acts as specific roles
411 in regulating embryonic development as reported in [35, 36]. The above observation also implies
412 that the previously unannotated 18th and 43th genomic positions may be related to heterochromatin
413 organization.
414 We also observe that genomic regions (HTG-Positions, SEs-Positions, TrxG-AP1-Positions,
415 TrxG-PRC-Positions, H3K4me1-Positions, the 9th class of CoreRF-Positions) are enriched by
416 Alu/B2/B4, which may be related to euchromatin organization [34]. HTG-Positions, SEs-Positions
417 as well as TrxG-AP1-Positions with higher Alu/B2 enrichment than others suggest that the
418 enrichment of Alu/B2 may indicate the transcription level of genes, and promote the formation of
419 hierarchical chromatin structures by regulating gene transcription and SEs/Enhancer activation.
420 Besides the above genomic positions, what is interesting is that MaLR elements are enriched
421 in AP1-Positions, the 21th class of CoreRF-Positions, and other unannotated regions (the 6th, 3th, 8th,
422 and 22th classes). We first observe that heterochromatin factors (the Hetero cluster) are weakly
423 enriched in the 21th class compared to the 9th class (CoreRF-Positions), which may result in poised
424 chromatin states in the 21th positions (**Fig. 4b and c**). In addition, previous studies have shown that
425 domain boundaries mediated by AP1 proteins (e.g., CTCF and cohesin) may block the spread of
426 chromatin states [3]. It may suggest that the regions dominantly enriched by MaLR elements may
427 often undergo switches between euchromatin and heterochromatin.
428 In summary we find that: (1) L1-Positions with low accessibility and unfavorably forming
429 boundaries are associated with heterochromatin organization, and Alu/B2/B4-Positions are
430 associated with euchromatin chromatin, while MaLR-Positions may result in poised chromatin
431 states. (2) ERVK elements may contribute to the high level of heterochromatin formation, while
432 Alu/B2 may promote highly transcription of genes and highly activation of SEs/Enhancers. These
433 retrotransposons contribute to the maintaining of chromatin states, and interplay with other types of
434 regulatory factors, to navigate the preference of genomic positions and gene regulation in single
435 cells.

The preference of genomic positions regulates cellular states. To investigate the preference of genomic positions in the above functional groups along the cell cycle process, we check the dynamics of each functional group forming boundaries in single cells among four different cycle phases, including G1, early-S (ES), mid-S (MS), and late-S/G2 (LS/G2). We first reveal that genomic positions enriched by SEs show a significant preference for forming boundaries in the G1 phase (**Fig. 5b**). The observation suggests that highly activation of SEs in the phase may promote gene regulation and transcription for cell growth in size, and ensure biomaterials for DNA synthesis. In addition, the functional group (HTG-Positions) accompanied with highly transcribed genes exhibits a significant loss of boundaries in the MS phase, which may be because the rates of transcription and protein synthesis are low during DNA replication (**Fig. 5b**). We also observe that functional groups occupied by both CTCF-cohesin and TrxG-PRC complexes prefer to form boundaries in ES phases (**Fig. 5b**). The observation implicates in the clearest segmentation of chromatin structures at the beginning of DNA replication [37, 38]. Both complexes have been proven to modify local chromatin structure and regulate higher-order chromatin organization [6, 39]. And functional groups (Hetero-Positions and AP2-Repressor-Positions) associated with heterochromatin organization prefers to form boundaries in both MS and LS/G2 phases (**Fig. 5b**). Both phases may prepare for everything entering the mitosis phase with condensing chromatin states.

In general, genomic positions (HTG-Positions, SEs-Positions, TrxG-AP1-Positions, TrxG-PRC-Positions, H3K4me1-Positions, and the 9th class of CoreRF-Positions) enriched by Alu/B2/B4 retrotransposons have higher preference scores for forming boundaries of single cells in G1 and ES phases in comparison with MS and LS/G2 phases, whereas genomic positions (Hetero-Positions, AP2-Repressor-Positions, and DamID-Positions) enriched by L1/ERVK retrotransposons display opposite tendency, following by high preference scores for forming boundaries in both MS and LS/G2 phases (**Fig. 5b** and **Supplementary Fig. S11**). The above observations further expound that the dynamic interplay among different types of regulatory factors, retrotransposons, and chromatin structures could navigate gene regulation and cell functions, even cell identity in single embryonic stem cells.

Discussion

Several decades of research have shown that eukaryotic chromatin adopts a complex hierarchical architecture within the nucleus, which plays a key role in functional implications for almost all nuclear processes. Thus, the spatially organized chromatin architecture interplaying with multiple types of regulatory factors shape focal chromatin landscapes and then exert gene regulatory functions. Single-cell 3D genome analysis extends the limitation of bulk analysis to show substantial cell-to-cell variation and promote our understanding of chromatin structures in the individual cell. Recent discoveries on single-cell 3D genome have shed light on the relationship between CTCF-cohesin complexes and domain formation, but the more molecular details associated with regulatory factors remain to be investigated [13].

Here, we develop HiCS to detect hierarchical chromatin structures from single-cell Hi-C maps, and observe hundreds of regulatory factors, including TFs, chromatin regulators, and histone modifications, are either enriched or not in domain boundaries of single cells, which presents several

478 different enrichment patterns. The results suggest their potential cooperative associations in shaping
479 focal chromatin interactions, thereby changing interaction density or insulation strength of these
480 regions. We further group spatially organized chromatin landscapes categories and different
481 regulatory factor clusters, explaining the emergence and functionality of different chromatin
482 landscapes and providing a comprehensive view of the genomic structure-function relationship. We
483 also find that the retrotransposons exactly match spatially organized chromatin landscapes. The
484 above results indicate that these regulatory factors interplaying with each other exert gene regulatory
485 processes and control cell functions, even cell identity.
486 The chromatin structures within the nucleus operate in an obvious dynamic process driven by
487 both "loop extrusion" and attractive process induced by regulatory factors (associated with
488 compartmentalization). The process may condense or loose local chromatin landscapes in an
489 overlapping and concerted manner accompanied by adjusting insulation strength of chromatin
490 position and generating chromatin loops, and then shape gene expression programs during cell-fate
491 specification [37]. Further work is needed to leverage more specific chromatin structures,
492 particularly chromatin loops, of single cells with more abundant regulatory factors (TFs, chromatin
493 regulators, histone modifications, retrotransposons, RNA, even structural variations) to understand
494 structure-function relationships in complex tissues or diseases [40]. It will promote our
495 understanding of how multiple types of regulatory factors interact with chromatin topological
496 engines (such as loop extraction and compartmentalization) to regulate the gene-repression program,
497 determine cell functions and identity, and further explain tissue complexity and disease
498 development.
499

500 ## Methods

501 **Single-cell Hi-C and other genomic data processing**

502 ***Single-cell Hi-C data generated from mESCs.*** The single-cell Hi-C dataset used in this study
503 consists of 1992 diploid cells of mESCs grown in 2i media without feeder cells with stringent quality
504 control filter. This dataset involves a median number of 393506 restriction fragments, and 127233
505 distinct >1 kb contacting pairs on average per cell [17]. The top 1315 cells with >250000 contacts
506 per cell were selected for downstream analysis. Among them, 317, 341, 303, and 354 cells belong
507 to G1, early-S (ES), mid-S (MS), and late-S/G2 (LS/G2) phases labeled by fluorescence-activated
508 cell sorting (FACS) sort criterion, respectively.
509

510 ***An atlas of ChIP-seq for mESCs.*** We organized an atlas of ChIP-seq for hundreds of regulatory
511 factors of mESCs (**Supplementary Table S1**), including CTCF, cohesin (*Smc1, Smc3, Rad21, Sa1,*
512 *Sa2*), mediators (*Med1, Med2, Med26, Cdk8, Cdk9*), codensin (*Capd3, Nipbl*), PolII, TFIIB,
513 polycomb repressive complex (*Aebp2, Rybp, Ring1b, Rnf2, Suz12, Ezh2, Eed, Pcl2*), Trithorax
514 protein family (*Set1a, Mll2, Mll3/4, Brg1*), the core regulatory factors of mESC (*Oct4, Sox2, Nanog*),
515 the regulatory factors of ESC occupied on enhancers or SEs (*Esrrb, Nr5a2, Klf4, Stat3, Prdm14,*
516 *Zfp281, Tcf3, Tcfcp2l1, Smad2/3*), the regulatory factors of ESC occupied on promoter-proximal
517 sites or sites that border topological domains (*c-Myc, n-Myc, Zfx, Tbx3, Yy1*), and additional 45 TFs.
518 We collected 14 histone modification factors, including H3K4me1, H3K4me2, H3K4me3,
519 H3K27me3, H3K36me3, H3K79me2, H3k9me3, H3K9ac, H3K122ac, H3K64ac, H3K27ac, H2AZ,

ZC3H11A, and H2AK119ub1. In addition, we also collected two chromatin accessibility datasets (ATAC-seq and Dnase-seq) and the DamID-seq dataset.

For the ChIP-seq of regulatory factors, peaks were called using MACS2 software with q-value cut-off 1×10^{-5} [41]. The source information and supporting materials of these factors were summarized in **Supplementary Table S1 and S2**.

Regulatory elements and genes. Enhancers/SEs and gene expression datasets of mESCs were downloaded in GSE29278 [42]. Housekeeping genes were downloaded in Housekeeping and Reference Transcript Atlas (HRT Atlas v1.0, www.housekeeping.unicamp.br) [43]. PhastCons scores were downloaded from the UCSC Genome Browser via ftp://hgdownload.cse.ucsc.edu/goldenPath/mm9/phastCons30way/vertebrate [44]. Mouse cell-cycle annotated genes were obtained from the mouse genome informatics (MGI) (http://www.informatics.jax.org/), containing 891 genes relating to cell cycle process and regulation. We adopted Seurat to detect differential expression genes and the variable score of genes using *FindAllMarkers* and *FindVariableFeatures* functions based on single-cell RNA-seq data of ESCs, which consists of 182 cells labeled by FACS sort criterion, including 59, 58, and 65 cells belonging to "G1" phase, "S" phase, and "G2M" phase, respectively [45, 46].

Retrotransposons. Retrotransposons built from RepeatMasker annotations were downloaded from the UCSC Table Browser (http://genome.ucsc.edu/). We kept the top six categories of counts for the downstream analysis, including Alu, B2, B4, MaLR, L1, and ERVK.

HiCS

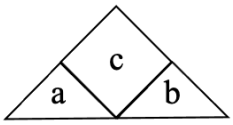
Impute contact probability for each chromosome of the individual cell. We first divided each chromosome into bins of specific size (40kb in this study), and counted the contact for each bin pair. Next, we modeled each chromosome as an unweighted network (each bin is one node, and each bin pair with non-zero contacts is added as one edge), and implemented a classic graph embedding method node2vec, which applies a biased random walks procedure, to compute the contact probability of edges by computing the cosine similarity of any two node embedding vectors, and obtained the imputed matrix A (**Fig. 1a**). We only kept the top 5% pairs for downstream analysis and the diagonal pairs were removed in our study.

Detect domain boundaries for each individual cell. Inspired by a fast density-based clustering method designed for grouping data points [47, 48], we take advantage of finding the cluster centers to detect domain boundaries for each chromosome of individual cells (**Fig. 1b and c**). Specifically, we define two indexes for each 40kb bin: (1) insulation strength $\rho(i)$ of the ith genomic position is defined as the ratio of $(I_{i,intra} - I_{i,inter})$ and $(I_{i,intra} + I_{i,inter})$ [49] using a 800kb sliding window size:

$$I_{i,intra} = I_a + I_b,$$

$$I_{i,inter} = I_c,$$

$$\rho(i) = \frac{I_{i,intra} - I_{i,inter}}{I_{i,intra} + I_{i,inter}},$$



where I_a, I_b, and I_c respectively represent the summation of interaction frequencies for region a, b, and c, and (2) minimum distance between the bin i and any other bin j with higher insulation strength is defined as $\delta(i)$:

$$\delta(i) = \min_{j:\rho(j)>\rho(i)} |i-j|, \quad (\delta_{max} < MAX).$$

We search for higher insulation strength of bin i in the range of $MAX = 500$ (20M genomic distance at 40kb resolution). Next, we define $\rho^{'}(i) = \rho^{'}(i)/\rho_{max}$ and $\delta^{'}(i) = \delta(i)/\delta_{max}$ such that both $\rho^{'}(i)$ and $\delta^{'}(i)$ are within range [0,1]. Then, we generate the rank γ of all bins for each chromosome by their $\eta(i) = \rho^{'}(i) \times \delta^{'}(i)$ in the descending and normalized the rank of each bin by $\gamma^{'}(i) = \gamma(i)/\gamma_{max}$. We define the optimal reflection point with $r = \underset{i=1:len}{\operatorname{argmin}}\, \eta(i)^2 + \gamma^{'}(i)^2$, where len is defined as the number of bins for a specific chromosome. The boundaries of the optimal structure are assigned by bins with $\eta(i) > \alpha \times \eta(r)$, $(\alpha = 1)$. The gap regions are defined by $I_{i,inter} = 0$ or have no contact with any other bins.

Determine the hierarchical chromatin structures. We ran the above procedures of detecting domain boundaries multiple times by define $\eta(i) > \alpha \times \eta(r)$ as a screening selection for different genomic levels, where the range of α are set as (0.1, 10) (**Fig. 2a**). In our study, we selected multiple scaling parameters α as $\{0.2,\ 1,\ 4,\ 8\}$ to obtain the hierarchical structures of chromatin, according to the distribution of the domain sizes and insulation strengths under different scaling parameters (**Fig. 2a and b**).

Decipher the categories of regulatory factors and spatially organized chromatin landscapes. We applied the hierarchical clustering method to group regulatory factors and chromatin positions, respectively (**Fig. 4a**). The classes of chromatin positions with the ratio>1% of all chromatin positions were selected for downstream analysis, which led to 29 classes of chromatin positions, along with 27 different transcriptional factor classes.

We further applied the hierarchical clustering method to merge the 27 classes into 7 large clusters, based on the Pearson correlation of the normalization ratio of the mean counts for peaks of regulatory factor classes (**Supplementary Fig. S9b**). These clusters or sub-clusters were manually annotated based on the annotation information (**Supplementary Table S2**). We then merged different chromatin position classes into 12 large spatially organized chromatin landscapes categories based on hierarchical clustering of correlation of regulatory factor classes, and manually annotated these categories based on the ratios of regulatory factor classes. We obtained and annotated 12 large spatially organized chromatin landscapes categories (consisting of 29 sub-categories), and 7 different regulatory factor clusters (consisting of 27 sub-clusters).

Data availability

All datasets analyzed in this study were published previously. The corresponding descriptions and preprocessing steps can be found in Supplementary Materials.

Software availability

The open-source HiCS python package and tutorial are available at GitHub (https://github.com/YusenYe/HiCS).

Acknowledgements

This work was supported by the National Natural Science Foundation of China [No. 62002275 to Y.Y., Nos. 62132015 & 61873198 to L.G., and Nos. 12126605&61621003 to S.Z.], the National Key Research and Development Program of China [No. 2019YFA0709501 to S.Z.], the Strategic Priority Research Program of the Chinese Academy of Sciences (CAS) [Nos. XDA16021400, XDPB17 to S.Z.].

Author Contributions

Y.Y. conceived the idea, implemented the algorithm and performed the analyses. Y.Y. interpreted the results. S.Z. and L.G. provided scientific insights on the applications. Y.Y. wrote the manuscript with feedback from all other authors. All of the authors read and approved the final manuscript.

Completing interests

The authors declare no completing interests.

References

1. Kempfer, R. and A. Pombo, *Methods for mapping 3D chromosome architecture.* Nature Reviews Genetics, 2020. **21**(4): p. 207-226.
2. Misteli, T., *The Self-Organizing Genome: Principles of Genome Architecture and Function.* Cell, 2020. **183**(1): p. 28-45.
3. Rao, Suhas S.P., et al., *A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping.* Cell, 2014. **159**(7): p. 1665-1680.
4. Lieberman-Aiden, E., et al., *Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome.* Science, 2009. **326**(5950): p. 289-293.
5. Dixon, J.R., et al., *Topological domains in mammalian genomes identified by analysis of chromatin interactions.* Nature, 2012. **485**(7398): p. 376-380.
6. Stadhouders, R., G.J. Filion, and T. Graf, *Transcription factors and 3D genome conformation in cell-fate decisions.* Nature, 2019. **569**(7756): p. 345-354.
7. Rowley, M.J. and V.G. Corces, *Organizational principles of 3D genome architecture.* Nature Reviews Genetics, 2018. **19**(12): p. 789-800.
8. Hnisz, D., et al., *Super-Enhancers in the Control of Cell Identity and Disease.* Cell, 2013. **155**(4): p. 934-947.
9. Young, Richard A., *Control of the Embryonic Stem Cell State.* Cell, 2011. **144**(6): p. 940-954.
10. Zepeda-Martinez, J.A., et al., *Parallel PRC2/cPRC1 and vPRC1 pathways silence lineage-*

specific genes and maintain self-renewal in mouse embryonic stem cells. Science Advances. **6**(14): p. eaax5692.

11. Cruz-Molina, S., et al., *PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation.* Cell Stem Cell, 2017. **20**(5): p. 689-705.e9.
12. Zuin, J., et al., *Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells.* Proceedings of the National Academy of Sciences, 2014. **111**(3): p. 996.
13. Bintu, B., et al., *Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells.* Science, 2018. **362**(6413): p. eaau1783.
14. Tan, L., et al., *Three-dimensional genome structures of single diploid human cells.* Science, 2018. **361**(6405): p. 924.
15. Crane, E., et al., *Condensin-driven remodelling of X chromosome topology during dosage compensation.* Nature, 2015. **523**(7559): p. 240-244.
16. Li, X., et al., *DeTOKI identifies and characterizes the dynamics of chromatin TAD-like domains in a single cell.* Genome Biology, 2021. **22**(1): p. 217.
17. Nagano, T., et al., *Cell-cycle dynamics of chromosomal organization at single-cell resolution.* Nature, 2017. **547**(7661): p. 61-67.
18. Schwarzer, W., et al., *Two independent modes of chromatin organization revealed by cohesin removal.* Nature, 2017. **551**(7678): p. 51-56.
19. Nora, E.P., et al., *Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization.* Cell, 2017. **169**(5): p. 930-944.e22.
20. Whyte, Warren A., et al., *Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes.* Cell, 2013. **153**(2): p. 307-319.
21. Piunti, A. and A. Shilatifard, *The roles of Polycomb repressive complexes in mammalian development and cancer.* Nature Reviews Molecular Cell Biology, 2021. **22**(5): p. 326-345.
22. Douillet, D., et al., *Uncoupling histone H3K4 trimethylation from developmental gene expression via an equilibrium of COMPASS, Polycomb and DNA methylation.* Nature Genetics, 2020. **52**(6): p. 615-625.
23. Buecker, C., et al., *Reorganization of Enhancer Patterns in Transition from Naive to Primed Pluripotency.* Cell Stem Cell, 2014. **14**(6): p. 838-853.
24. Cole, M.F., et al., *Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells.* Genes & development, 2008. **22**(6): p. 746-755.
25. Niwa, H., et al., *Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3.* Genes & development, 1998. **12**(13): p. 2048-2060.
26. Beattie, G.M., et al., *Activin A maintains pluripotency of human embryonic stem cells in the absence of feeder layers.* Stem cells, 2005. **23**(4): p. 489-495.
27. Yang, Bin X., et al., *Systematic Identification of Factors for Provirus Silencing in Embryonic Stem Cells.* Cell, 2015. **163**(1): p. 230-245.
28. Piunti, A. and A. Shilatifard, *Epigenetic balance of gene expression by Polycomb and COMPASS families.* Science, 2016. **352**(6290): p. aad9780.

682 29. Weintraub, A.S., et al., *YY1 Is a Structural Regulator of Enhancer-Promoter Loops.* Cell,
683 2017. **171**(7): p. 1573-1588.e28.
684 30. Furlong Eileen, E.M. and M. Levine, *Developmental enhancers and chromosome topology.*
685 Science, 2018. **361**(6409): p. 1341-1345.
686 31. Rege, M., et al., *Chromatin Dynamics and the RNA Exosome Function in Concert to*
687 *Regulate Transcriptional Homeostasis.* Cell Reports, 2015. **13**(8): p. 1610-1622.
688 32. Surface, Lauren E., et al., *H2A.Z.1 Monoubiquitylation Antagonizes BRD2 to Maintain*
689 *Poised Chromatin in ESCs.* Cell Reports, 2016. **14**(5): p. 1142-1155.
690 33. Thomas, M.J. and E. Seto, *Unlocking the mechanisms of transcription factor YY1: are*
691 *chromatin modifying enzymes the key?* Gene, 1999. **236**(2): p. 197-208.
692 34. Lu, J.Y., et al., *Homotypic clustering of L1 and B1/Alu repeats compartmentalizes the 3D*
693 *genome.* Cell Research, 2021. **31**(6): p. 613-630.
694 35. Hanna, C.W., et al., *Endogenous retroviral insertions drive non-canonical imprinting in*
695 *extra-embryonic tissues.* Genome Biology, 2019. **20**(1): p. 225.
696 36. Guo, F., et al., *The Transcriptome and DNA Methylome Landscapes of Human Primordial*
697 *Germ Cells.* Cell, 2015. **161**(6): p. 1437-1452.
698 37. Nuebler, J., et al., *Chromatin organization by an interplay of loop extrusion and*
699 *compartmental segregation.* Proceedings of the National Academy of Sciences, 2018.
700 **115**(29): p. E6697-E6706.
701 38. Ye, Y., L. Gao, and S. Zhang, *Circular Trajectory Reconstruction Uncovers Cell-Cycle*
702 *Progression and Regulatory Dynamics from Single-Cell Hi-C Maps.* Advanced Science,
703 2019. **6**(23): p. 1900986.
704 39. Schuettengruber, B., et al., *Genome Regulation by Polycomb and Trithorax: 70 Years and*
705 *Counting.* Cell, 2017. **171**(1): p. 34-57.
706 40. Jerkovic´, I. and G. Cavalli, *Understanding 3D genome organization by multidisciplinary*
707 *methods.* Nature Reviews Molecular Cell Biology, 2021. **22**(8): p. 511-528.
708 41. Zhang, Y., et al., *Model-based Analysis of ChIP-Seq (MACS).* Genome Biology, 2008. **9**(9):
709 p. R137.
710 42. Shen, Y., et al., *A map of the cis-regulatory sequences in the mouse genome.* Nature,
711 2012. **488**(7409): p. 116-120.
712 43. Hounkpe, B.W., et al., *HRT Atlas v1.0 database: redefining human and mouse*
713 *housekeeping genes and candidate reference transcripts by mining massive RNA-seq*
714 *datasets.* Nucleic Acids Research, 2021. **49**(D1): p. D947-D955.
715 44. Siepel, A., et al., *Evolutionarily conserved elements in vertebrate, insect, worm, and yeast*
716 *genomes.* Genome research, 2005. **15**(8): p. 1034-1050.
717 45. Buettner, F., et al., *Computational analysis of cell-to-cell heterogeneity in single-cell RNA-*
718 *sequencing data reveals hidden subpopulations of cells.* Nature Biotechnology, 2015.
719 **33**(2): p. 155-160.
720 46. Hao, Y., et al., *Integrated analysis of multimodal single-cell data.* Cell, 2021. **184**(13): p.
721 3573-3587.e29.
722 47. Rodriguez, A. and A. Laio, *Clustering by fast search and find of density peaks.* Science,
723 2014. **344**(6191): p. 1492-1496.

48. Ye, Y., L. Gao, and S. Zhang, *MSTD: an efficient method for detecting multi-scale topological domains from symmetric and asymmetric 3D genomic maps.* Nucleic Acids Research, 2019. **47**(11): p. e65-e65.
49. Luo, X., et al., *3D Genome of macaque fetal brain reveals evolutionary innovations during primate corticogenesis.* Cell, 2021. **184**(3): p. 723-740.e21.