DiffDomain enables identification of structurally reorganized topologically associating domains

Dunming Hua [1,2,#], Ming Gu[1,2,#], Xiao Zhang[1,2,#], Yanyi Du[1,2,#], Hangcheng Xie[1,2], Li Qi[3], Xiangjun Du[1,2], Zhidong Bai[4], Xiaopeng Zhu[5] and Dechao Tian[1,2,*]

[1]Department of Biostatistics and Systems Biology, School of Public Health (Shenzhen), Sun Yat-sen University, Shenzhen, Guangdong 510275, China
[2]Department of Biostatistics and Systems Biology, School of Public Health (Shenzhen), Shenzhen Campus of Sun Yat-sen University, Shenzhen, 518107, China
[3]Chongqing Municipal Center for Disease Control and Prevention, Chongqing 400042, China
[4]KLASMOE & School of Mathematics and Statistics, Northeast Normal University, Changchun, Jilin 130024, China
[5]Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA

[#]These four authors contributed equally
[*]Corresponding author: tiandch@mail.sysu.edu.cn

Abstract

Topologically associating domains (TADs) are critical structural units in three-dimensional genome organization of mammalian genome. Dynamic reorganizations of TADs between health and disease states are associated with transcription and other essential genome functions. However, computational methods that can identify reorganized TADs are still in the early stages of development. Here, we present DiffDomain, an algorithm leveraging high-dimensional random matrix theory to identify structurally reorganized TADs using chromatin contact maps. Method comparison using multiple real Hi-C datasets reveals that DiffDomain outperforms alternative methods for FPRs, TPRs, and identifying a new subtype of reorganized TADs. The robustness of DiffDomain and its biological applications are demonstrated by applying on Hi-C data from different cell types and disease states. Identified reorganized TADs are associated with structural variations and changes in CTCF binding sites and other epigenomic changes. By applying to a single-cell Hi-C data from mouse neuronal development, DiffDomain can identify reorganized TADs between cell types with reasonable reproducibility using pseudo-bulk Hi-C data from as few as 100 cells per condition. Moreover, DiffDomain reveals that TADs have clear differential cell-to-population variability and heterogeneous cell-to-cell variability. Therefore, DiffDomain is a statistically sound method for better comparative analysis of TADs using both Hi-C and single-cell Hi-C data.

Introduction

Recent development of mapping technologies, such as Hi-C [1], that probes the 3D genome organization reveals that a chromosome is divided into topologically associating domains (TADs) [2, 3]. TADs are genomic regions that chromatin loci are more frequently interacting with chromatin loci from the same TADs rather than from adjacent TADs. TADs are functional units for transcriptional regulation by constraining interactions between enhancers and promoters [4], for example. Although TADs are stable between cell types as revealed by earlier studies [2, 5], there are growing evidence for TAD reorganization in diseases [6–8], cell differentiation [5, 9, 10], somatic cellular reprogramming [11], between neuronal cell types [12], and between species [13, 14]. For example, extensive reorganizations of TADs are observed during somatic cell reprogramming, associating with dynamics of transcriptional regulation and changes in cellular identity [11]. TADs are also variable among individual cells as revealed by single-cell studies [15–20] and live-cell imaging [21]. Thus, it is important to identify reorganized TADs through comparative analysis to further understand the functional relevance of 3D genome organization, a major priority of current work in the field [22].

The majority of current methods call a reorganized TAD if at lesast one of its two boundaries changed between two conditions [11, 12, 23–28]. These methods enable easy integration with other analysis piplines and identified reorganized TADs have clear biological interpretation. However, they fail in identifying reorganized TADs without changing in boundaries, in additional to lacking statistical tests to differentiate ramdom pertubations and significant structrual reorganization of a TAD. Only a few nonparametric statistical methods are proposed to call TAD reorganization [29–32]. These methods define the structural similarity of a TAD by statistics from two Hi-C contact matrices, such as stratum-adjusted correlation coefficient used by DiffGR [30]. Distributions of the statistics on pairs of simulated Hi-C matrices are then used to compute empirarcal P values. However, these nonparametric statistical methods are conservative (see our own comparison later). TADs are relatively small, for example, median size of TADs is 185 kb [33]. Small size feature of TADs poses another computational challenge for identifying structurally rewired TADs using low-resolution Hi-C data. Importantly, identifying reorganized TADs using emerging single-cell Hi-C (scHi-C) data is largely under-explored. Other methods are developed for comparing Hi-C matrices at differnet scales and for different purposes: quantifying similarities of genome-wide Hi-C contact matrices [34, 35], identifying differential A/B compartments [36], and identifying differnetial chromatin interactions [37–39]. However, these methods are not tailored to compare Hi-C contact matrices at TAD-level, not optimal for identifying reorganized TADs (see our own comparison later). Therefore, new algorithms are needed to fill these gaps.

Here, we develop DiffDomain, a new parametric statistical method for identifying reorganized TADs. Its inputs are two Hi-C contact matrices from two biological conditions and a set of TADs identified in biological condition 1. This setting enables straightforward integration of DiffDomain with other analysis piplines of Hi-C data such as TAD calling and integrative analysis of multi-omics data. For each TAD, DiffDomain *directly* computes a difference matrix then normalize it properly, skipping the challenging normalization steps for individual Hi-C contact matrices. DiffDomain then borrows well-established theorectical results in ramdom matrix theory to compute a theorectical P value. We show that the assumptions of DiffDomain are reasonable. Method comparisons on real data reveal that DiffDomain has substantial advantages over alternative methods in false positive rates and accuracy in identifying truly reorganized TADs. Reorganized TADs identified by DiffDomain are biologically relevant in different human cell lines and disease states. Application to scHi-C data reveals that DiffDomain can identify reorganized TADs between cell types and TADs with differential variabilities among individual cells within the same cell type. Moreover, DiffDomain can quantify cell-to-cell variability of TADs betweeen individual cells. Together, these analyses demonstrate the power of DiffDomoain for better identification

of structually reorganized TADs using both bulk Hi-C and single-cell Hi-C data.

Results

Overview of DiffDomain

The workflow of DiffDomain is illustrated in Figure 1. Given two Hi-C contact matrices from two biological conditions and a set of TADs called in biological condition 1 (Fig. 1A). In this paper, TADs are called by Arrowhead [33] and Hi-C contact matrices are KR-normalized, unless otherwise stated. Our goal is to test if each TAD identified in biological condition 1 has significantly structural reorganization in biological condition 2. The core of DiffDomain is formulating the problem as a hypothesis testing problem, where the null hypothesis is that the TAD doesn't undergo significant structural reorganization in condition 2. Given two input Hi-C contact matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$, DiffDomain first log-transform them to adjust for the exponential decay of Hi-C contacts with increased 1D distances between chromosome bins. Their difference $\log(\boldsymbol{A}_1) - \log(\boldsymbol{A}_2)$ is calculated and denoted by $\boldsymbol{D}$ (Fig. 1B). $\boldsymbol{D}$ is further normalized by a 1D distance-stratified standardization procedure, similar to the procedures in HiC-DC+ [38] and SnapHiC [40]. Specifically, each d-off diagonal part of $\boldsymbol{D}$ is subtracted by its sample mean and divided by its sample standard deviation (Fig. 1C), $-N + 2 \leq d \leq N - 2$, reducing 1D distance-dependence of values in $\boldsymbol{D}$ and differences caused by variation in read depths between two biological conditions (see Supplementary Fig. S1 for two more detailed visualization). Intuitively, if a TAD is not significantly reorganized, normalized $\boldsymbol{D}$ would resemble a white noise random matrix, enabling us to borrow theoretical results in random matrix theory. Under the null hypothesis, DiffDomain assumes that $\boldsymbol{D}/\sqrt{N}$ is a generalized Wigner matrix (Fig. 1D), a well studied random matrix model. Its largest eigenvalue λ_N is proved to be fluctuating around 2. Armed with the fact, DiffDomain reformulates the reorganized TAD identification problem into the hypothesis testing problem:

$$H_0 : \lambda_N = 2 \text{ vs. } H_1 : \lambda_N > 2.$$

The key theoretical results empowering DiffDomain is that $\theta_N = N^{2/3}(\lambda_N - 2)$, a normalized λ_N, asymptotically follows a Tracy-Widom distribution with $\beta = 1$, denoted as TW_1. Thus, θ_N is chosen as the test statistic and one-sided P value is calculate as $P_{TW_1}(\theta_N \geq x)$. H_0 is rejected if P value is less than a predefined significant level α, which is 0.05 in this paper (Fig. 1E). The pseudocode is shown in Supplementary Methods A.1. For a set of TADs, P values are adjusted for multiple comparisons using Benjamini-Hochberg (BH) method as the default. Once DiffDomain identifies the subset of reorganized TADs, it further classifies them into six subtypes based on changes in their boundaries, beneficial for downstream biological analyses and interpretations (Fig. 1F, Supplementary Methods A.2). A few reorganized TADs in real Hi-C data are shown in Fig. 1G. Details are described in Methods.

Note that, although each biological condition may have multiple Hi-C replicates, DiffDomain takes the combined Hi-C contact matrix from the replicates as the input which is a common practice to generate a large number of Hi-C interactions [33]. Assumptions of DiffDomain are appropriate (see Methods section, Supplementary Results B.1, Supplementary Fig. S2). When $N = 10$, the Tracy-Widom distribution TW_1 is an adequate approximation of the exact distribution of θ_N [41]. Under common 10 kb resolution Hi-C data, $N = 10$ refers to TADs with 100 kb in length, much smaller than the median TAD length 185 kb [33]. Thus DiffDomain only computes the P value for TADs with at least 10 chromosome bins, a practical constraint. DiffDomain is robust to a varied number of sequencing reads, Hi-C resolution, and different TAD callers (Supplementary Results B.2, Supplementary Fig. S3, S4).

DiffDomain consistently outperforms alternative methods in multiple aspects

First, we assess the false positive rate (FPR) which is the ratio of the number of false positives to the number of true negatives. A smaller FPR means that the identified significantly reorganized TADs are more likely to be true. Due to the lack of gold-standard data, we resort to analyzing the proportions of significantly reorganized TADs between five different Hi-C replicates from the GM12878 cell line. These Hi-C replicates are generated by different experimental procedures and have a highly varied total number of Hi-C contacts (Supplementary Table S1). But the TADs are expected to have few structural changes between these Hi-C replicates. The proportion of identified reorganized TADs is treated as an estimate of FPR (Supplementary Methods A.3) and is expected to be small. The Hi-C resolution is chosen as 10 kb. Comparing GM12878 Hi-C replicates *combined* and *primary*, we find that DiffDomain, TADCompare [31], and HiCcompare [43] have FPRs that are smaller than the given significant level 0.05, suggesting good controls of FPR. In contrast, DiffGR [30], DiffTAD [29], and HiC-DC+ [38] have inflated FPRs (more than two-fold higher than 0.05), indicating poor controls of FPR (Fig 2A). Similar results are observed by repeating the above analysis to other GM12878 Hi-C replicates and 25 kb resolution Hi-C data (Supplementary Fig. S5).

A good control in FPR does not necessarily represents high power in detecting reorganized TADs between biological conditions. To investigate this, we compare TADs between two blood cell lines GM12878 and K562, DiffDomain identifies that 30.771% of GM12878 TADs are reorganized in K562. In contrast, TADCompare, HiCcompare, and HiC-DC+ only identify $\leq 8.256\%$ of GM12878 TADs that are reorganized in K562 (Fig 2B), suggesting that they are too conservative to identify reorganized TADs between biological conditions. Similar results are observed by repeating the above analysis to other human cell lines and 25 Kb resolution Hi-C data (Supplementary Table S2, Supplementary Fig. S6), demonstrating the robustness of the observations. Conservation of TADCompare is due to the fact that it is designed to scan every chromatin loci for potential reorganized TAD boundaries. But this analysis uses a given list of TADs, a common practice in Hi-C data analysis, which sharply narrows down the search space of TADCompare. Conservations of HiCcompare and HiC-DC+ are due to the fact that they are designed for detecting differential chromatin interactions, not specifically tailored for identifying reorganized TADs.

Compared with TADsplimer which specifically identifies *split* and *merge* TADs [25], DiffDomain identifies similar numbers of *split* and *merge* TADs between multiple pairs of human cell lines (Supplementary Fig. S7). Importantly, DiffDomain identifies that the majority (minimum 43.137%, median 81.357%, maximum 98.022%) of the identified reorganized TADs are the other four subtypes (Supplementary Fig. S8), which can not be detected by TADsplimer. For example, among the GM12878 TADs that are identified as reorganized in K562 by DiffDomain, *strength-chagne* is the leading subtype of reorganized TADs, consistent with the fact that both GM12878 and K562 are blood cell lines (Fig 2C). These results demonstrate that DiffDomain has substantial improvements over TADsplimer.

We next investigate the true positive rate (TPR). A higher TPR means that more truly reorganized TADs are correctly identified as reorganized TADs. Through an extensive literature search, we collect 65 TADs that are reorganized between 146 pairs of biological conditions in 15 published papers (Supplementary Table S3, Supplementary Methods A.4). We use these TADs as the gold standard data to compute the TPR (Supplementary Methods A.3) and also call these TADs as truly reorganized TADs. Four truly reorganized TADs, either correctly identified or missed by DiffDomain, are shown in Fig. 2D. HiCcompare and HiC-DC+, designed for identifying differential chromatin interactions, are not directly applicable to only testing reorganization of one single TAD and thus are excluded from the analysis. We find that the TPR of DiffDomain is 68.493%, which is 1.639, 2.60, and 16.665 times higher than that of alternative methods DiffGR, DiffTAD, and TADCompare, respectively (Fig. 2E). Compared with Diff-

Domain, DiffGR, DiffTAD, and TADCompare only uniquely identify 11, 10, and 1 truly reorganized TADs, respectively (Supplementary Fig. S9). Closer examination shows that DiffDomain has much smaller P values than other methods (Wilcoxon rank-sum test, $P \leq 1.62 \times 10^{-12}$, Fig. 2F), demonstrating that DiffDomain has stronger statistical evidence in favor of truly reorganized TADs. DiffDomain still misses 31.507% of possible pairwise comparisons of truly reorganized TADs. One reason is that some of the missed truly reorganized TADs have highly similar Hi-C contact matrices between biological conditions. For example, the missed truly reorganized TAD, chr11:1500000-2200000 (Fig. 2D), has the stratum adjusted correlation coefficient (SCC) [44] at 0.999. Generally, missed truly reorganized TADs have significantly ($P = 8.26 \times 10^{-6}$) higher SCC scores than those correctly identified reorganized TADs by DiffDomain (Fig. 2G). Since DiffGR uses SCC as the test statistic, this results also partially explains the low TPR (41.781%) of DiffGR and highlights that SCC alone is not optimal for identifying reorganized TADs. Another reason is that the resolutions of some Hi-C data are low since P values are moderately negatively associated with the maximum values of Hi-C contact matrices (Spearman's rank correlation coefficient $\rho = -0.534$).

Additionally, DiffDomain is efficient in memory usage and acceptable in computation time compared with alternative methods (Supplementary Results B.3, Supplementary Fig. S10).

In summary, compared with alternative methods, DiffDomain has multiple improvements including FPRs, proportions of identified reorganized TADs between different biological conditions, subtypes of reorganized TADs, and TPRs.

Reorganized TADs are associated with epigenomic changes

Armed with the advantages of DiffDomain over alternative methods, we explore the connections between TAD reorganization and epigenomic dynamics. We first showcase a GM12878 TAD that is significantly reorganized in K562 and classified as a *strength-change* TAD by DiffDomain (Fig. 3). The TAD covers a 445 kb region on chromosome 6. The TAD structural changes involve the vascular endothelial growth factor gene *VEGFA*, which is a major tumor angiogenic gene that is over-expressed in leukemia (See reviews [45, 46] for more details), consistent with the fact that K562 cells are chronic myelogenous leukemia cells. We find that the reorganized TAD has K562 specific functional annotations. The genomic region covered by the TAD is more accessible (1.71 times higher DNase peak coverage) in K562 than in GM12878 (Fig. 3B). The H3K27ac and H3K4me1 peak coverages of the TAD region in K562 are 3.24 and 3.28 times higher than the coverages in GM12878, respectively. In contrast, the H3K4me3 and H3K36me3 peak coverages of the TAD in K562 are only 1.31 and 1.25 times higher than the coverages in GM12878, respectively. Four regions in the TAD are annotated as super-enhancers [47] only in K562 (Fig. 3B). Note that, the TAD region is in A compartments in both cell types, suggesting that A/B compartments switch is not the reason for the gain in accessibility and histone modifications that are associated with gene activation. The normalized difference matrix $\boldsymbol{D}$ between Hi-C contact matrices of the TAD highlights that super-enhancer SE2 have increased Hi-C contacts with the *VEFGA* gene in K562 (Fig. 3C). To gain further insights into structural differences of the TAD in the two cell types, we compare the 3D structural representations of the TAD region. We run Chrom3D [48] 100 times to construct 100 possible 3D structures in each cell type for statistical comparisons. One possible 3D structure in each cell type illustrate the 3D structural differences of the TAD between GM12878 and K562 (Fig. 3D-E). Overall, the super-enhancer SE2, but not SE3, (Fig. 3B) are much spatially closer ($P < 2.22 \times 10^{-16}$) to VEGFA in K562 than in GM12878 (Supplementary Fig. S11). These results show that the reorganized GM12878 TAD in K562 has K562-specific chromatin organization and potential biological functions.

Generally, comparative analyses across multiple pairs of human normal and disease cell lines reveal that reorganized TADs that are of *strength-change*, *loss*, *zoom*, and *merge* have significant gain/loss

in the number of CTCF binding sites at TAD boundaries compared with other TADs, consisting with enrichment of CTCF binding sites in TAD boundaries (Supplementary Results B.4, Supplementary Fig. S12). Reorganized TADs can reveal cell identity (Supplementary Results B.5, Supplementary Fig. S6, S13). GM12878 (normal lymphoblastoid cell line) TADs that are reorganized in K562 (chronic myeloid leukemia cell lines) are enriched (P=0.01) in disease genes in chronic myelogenous leukemia. Across pairs of cell types, reorganized TADs tend to gain in chromatin accessibility, active transcription signals H3K27ac and K3K4me1 (Supplementary Results B.6, Supplementary Fig. S14, S15). Additionally, the reorganized TADs are biologically relevant to diffuse intrinsic pontine glioma (DIPG) (Supplementary Results B.7, Supplementary Fig. S16, S17, S18, Supplementary Table S4). Together, these results demonstrate the functional relevance of reorganized TADs in multiple human normal and disease cell lines.

Reorganized TADs are enriched with structural variations (SVs)

SVs can contribute to diseases by rewiring 3D genome organization. To further demonstrate the biological relevance of reorganized TADs, we systematically investigate the associations between SVs and reorganized TADs. High resolution SVs, including deletions, inversions, duplications, and translocations, from erythroleukemia (K562 cell line) and pediatric high-grade glioma (DIPG007 and DIPGXIII cell lines) are downloaded from Wang et al. [49]. The number of SVs and paired normal Hi-C data are summarized in Supplementary Table S5.

We find that reorganized TADs are enriched with SVs. First, reorganized TADs are enriched with the breakpoints of SVs. Two examples are visualized in Fig. 4A-B. K562 has 63 breakpoints from 45 SVs that lie in K562 TADs or GM12878 TADs. Among the 63 breakpoints, 27 (42.86%) lie in the GM12878 TADs that are reorganized in K562, which is significantly ($P = 8.96 \times 10^{-5}$) higher than the proportions from randomly sampled equal-numbered reorganized TADs (Fig. 4C). Similar enrichment results are also observed in pediatric high-grade glioma cell lines DIPG007 and DIPGXIII (Fig. 4C). The two breakpoints of a SV may not lie in reorganized TADs, but the genomic region defined by the two breakpoints of the SV can overlap with reorganized TADs (see two examples in Fig. 4D). Analyzing based on the genomic region covered by a SV, K562 has 33 SVs that overlap with K562 TADs or GM12878 TADs. Among those 33 SVs, 28 (84.85%) overlap with GM12878 TADs that are reorganized in K562, significantly ($P < 2.22 \times 10^{-16}$) higher than the proportions in random cases (Fig. 4F). Similar significant results are observed in DIPG007 and DIPGXIII cell lines (Fig. 4F), demonstrating that reorganized TADs are enriched with overlapping with the genomic regions covered by SVs. The enrichments are not sensitive to the Hi-C resolution because choosing Hi-C resolution at 10 kb instead of 25 kb used here gives similar significant results, except for the pair of normal human astrocytes (NHA) verse DIPGXIII (Supplementary Fig. S19). Together, these results reveal that reorganized TADs are enriched with SVs in terms of enrichments with the breakpoints of SVs and overlapping with the genomic regions covered by SVs, further demonstrating the biological relevance of reorganized TADs.

DiffDomain improves profiling of TAD reorganization related to SARS-CoV-2 infection

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causes over 640 million confirmed coronavirus disease 2019 (COVID-19) cases, including over 6.6 million deaths, worldwide as of December 2, 2022 (https://covid19.who.int), posing a huge burden to global public health. Wang et al. [50] is the first Hi-C study into the effects of SARS-CoV-2 infection on host 3D genome organization, finding a global pattern of TAD weakening after SARS-CoV-2 infection. However, the analysis uses aggregation domain analyses which cannot directly identify individual weakened TADs, in addition to missing other subtypes of TAD reorganization.

To further demonstrate the biological applications of DiffDomain, we reanalyze the data. We find that

20.58% (840 in 4082) mock-infected A549-ACE2 TADs are reorganized in SARS-CoV-2 infected A549-ACE2 cells. Among the reorganized TADs, *strength-change* TADs are the leading subtype (64.64%) (Fig. 5A) which is consistent with the global pattern of TAD weakening [50], verifying the reorganized TADs identified by DiffDomain. The following most frequent subtypes are *merge* TADs and *complex* TADs (18.45% and 10.95%, Fig. 5A), refining the characterization of TAD reorganization after SARS-CoV-2 infection. These reorganized TADs also enable refined profiling of transcriptional regulation in response to SARS-CoV-2 infection. Compared with the other TADs, the reorganized TADs have significantly higher numbers of up-regulated genes and down-regulated genes ($P \leq 1.27 \times 10^{-4}$, Fig. 5B, Supplementary Fig. S20A). Similar significant patterns are observed comparing *strength-change* TADs and *split* TADs with the other TADs ($P \leq 8.07 \times 10^{-3}$, Supplementary Fig. S20B), highlighting that the two subtypes have stronger connections with differentially expressed genes than other subtypes of reorganized TADs. In contrast, compared to the other TADs, the six subtypes of reorganized TADs have significantly higher numbers of both enhanced and weakened peaks of H3K27ac, SMC3, and RAD21 where H3K27ac is a marker for active enhancers and SMC3 and RAD21 are two critical cohesin subunits that regulate 3D genome organization (Supplementary Results B.8, Fig. 5C-E, Supplementary Fig. S20C-E). These results suggest that, after SARS-CoV-2 infection, the subtypes of reorganized TADs all have strong associations with epigenome reprogramme, and *strength-change* TADs and *split* TADs have strong associations with deregulation of gene expression, highlighting the importance of subtypes of reorganized TADs identified by DiffDomain.

DiffDomain characterizes cell-to-population and cell-to-cell variability of TADs using scHi-C data

Recent advances in scHi-C sequencing methods enable profiling of 3D genome organization in individual cells, revealing intrinsic cell-to-cell variability of TADs among individual cells. However, quantifying the variability is challenging due to the properties of scHi-C data such as high sparsity, low genome coverage, and heterogeneous [51, 52]. As a proof-of-concept, we apply DiffDomain to a moderate-sized scHi-C dataset from mouse neuronal development, which has 3646 cells with the median number of contacts per cell at 400,000 [53].

We first ask how many individual cells are sufficient to identify reorganized TADs between cell types with high reproducibility using pseudo-bulk Hi-C data (Supplementary Methods A.5.1). To do this, we design a sampling experiment to gradually increase the number of used individual cells (Supplementary Methods A.5.2). We find that DiffDomain can identity reorganized TADs between cell types with reasonable reproducibility using as few as one hundred sampled cells. For example, DiffDomain consistently identifies that a neuronal TAD, harboring neuronal marker genes *GM24071*, *LRFN2*, *MOCS1*, and *1700008K24RIK* [53], is reorganized in oligodendrocytes with numbers of cells starting at 100 (Fig. 6A). Consistent of DiffDomain on other example genomic regions are shown in Supplementary Fig. S21. On average, DiffDomain identifies that 19.25% neuronal TADs are reorganized in oligodendrocytes using only 250 sampled cells from each cell type, consistent (average Jaccard index at 0.49) with the identified reorganized TADs using all available cells in both cell types (Fig. 6B). Similar results are observed when identifying neonatal neuron 1 (the youngest structure type) TADs that are reorganized in cortical L2-5 pyramidal cells (adult type) (Fig. 6C-D). Jointly increasing the numbers of sampled cells in both cell types improves the performance of DiffDomain, as expected (Fig. 6B). In contrast, only increasing the number of sampled cells in one cell type has a limited boost in performance. For example, oligodendrocytes have only 257 cells but neurons has 1380 cells. Further increasing the number of sampled neurons from 250 to 500 has a slight performance improvement (Fig. 6B). The observation is further confirmed when comparing neuronal subtypes neonatal neuron 1 and neonatal cortical L2-5 pyramidal cells, in which the number of sampled cells in the latter subtype is no more than 150 (Fig. 6C) or 228

(Fig. 6D). Repeating the analysis by using bulk Hi-C data [54, 55] to create gold-standard reorganized TADs, we observed similar patterns in neuronal TADs that are identified in astrocytes. For example, sampling 150 cells in each cell type identifies 12.40% neuronal TADs that are reorganized in astrocytes on average. Among the reorganized TADs, 62.55% are also identified as reorganized TADs when bulk Hi-C data are used (Fig. 6E). Considering the medium number of contacts per cell at 400000, the merged Hi-C data from hundreds of cells are ultra-sparse pseudo-bulk Hi-C data. These results demonstrate that DiffDomain can work with ultra-sparse Hi-C data.

Next, we move to quantify the cell-to-population variability of TADs, that is, comparing TADs in individual cells to the population average. To do this, scHi-C data with 50 kb resolution from neonatal neuron 1 cells is imputed by scHiCluster [56] For each TAD, DiffDomain compares imputed Hi-C contact map of the TAD in each cell to the pseudo-bulk Hi-C contact map. Resulted *P* values reflect cell-to-population variability of TADs and thus are used by hierarchical clustering to divide TADs into three categories: high, medium, and low cell-to-population variational TADs (Fig. 6F). We find that TADs have clear differential cell-to-population variability. One example high cell-to-population variational TADs and their adjacent medium cell-to-population variational TADs in 9 cells are shown in Fig. 6H. Among the 2146 neonatal neuron 1 TADs, 8.90% (191) are high cell-to-population variational TADs, 7.50% (161) and 83.60% (1794) are medium and low cell-to-population variational TADs (Fig. 6G). They are distributed across chromosomes (Fig. 6I). Similar results are observed in other cell types (Supplementary Fig. S22). These results demonstrate that TADs have clear differential variability between individual cells and the population average, consistent with earlier observations [26, 57].

Next, we move to investigate the cell-to-cell variability of TADs. Requiring only one Hi-C contact matrix from each condition, DiffDomain can directly quantify cell-to-cell variability of TADs between individual cells using imputed scHi-C data. Note that, similar to other methods, pairwise comparison of TADs using scHi-C data from thousands of individual cells leads to exponential growth in runtime and thus is computational expensive [51]. As a proof-of-concept, we apply DiffDomain to scHi-C data from randomly selected 50 cortical L2-5 pyramidal cells and 50 adult astrocytes. We find that the cell-to-cell variability of TADs is heterogeneous. The heterogeneity is consistent among the pairwise comparisons but quite different from those from random scenarios in which equal-numbered reorganized TADs are randomly assigned in pairwise comparisons. The proportion of reorganized TADs is consistent among the 2,500 pairs of individual cells, ranging from 46.0% to 75.7% (Fig 6J). Across the 50 cortical L2-5 pyramidal cells, the proportions of TADs that are reorganized in a varied number of adult astrocytes are fairly consistent (Fig. 6K). Moreover, the proportions of TADs that are either low in cell-to-cell variability (reorganized in no more than 10 adult astrocytes) or high in cell-to-cell variability (reorganized in more than 40 adult astrocytes) are much higher than those from random scenarios (Fig. 6K, Supplementary Fig. S23), consistent with differential cell-to-population variability of TADs as reported in the previous paragraph. This observation is also concordance with randomized placement of TAD-like blocks in individual cells but with strong preference for TAD boundaries observed in bulk Hi-C data [18], further demonstrating the utilization of DiffDomain.

In summary, DiffDomain works on scHi-C data to identify reorganized TADs between cell types, identify TADs with differential cell-to-population variability, and characterize cell-to-cell variability of TADs.

Conclusion and Discussion

In this work, we present a statistical method, DiffDomain, for comparative analysis of TADs using a pair of Hi-C datasets. Extensive evaluation using real Hi-C datasets demonstrates clear advantages of DiffDomain over alternative methods for controlling false positive rates and identifying truly reorganized TADs with much higher accuracy. Applications of DiffDomain to Hi-C datasets from different cell lines and disease states demonstrate that reorganized TADs are enriched with structural variations and associated with CTCF binding sites changes and epigenomic changes, revealing their condition-specific biological relevance. By applying to a scHi-C data from mouse neuronal development, DiffDomain can identify reorganized TADs between cell types with considerable reproducibility using pseudo-bulk Hi-C data from as few as a hundred cells. Moreover, DiffDomain can reliably characterizes the cell-to-population and cell-to-cell variability of TADs using scHi-C data.

The major methodological contribution of DiffDomain is directly characterizing the differences between Hi-C contact matrices using high-dimensional random matrix theory. First, DiffDomain makes no explicit assumption on the input chromatin contact matrices, directly applicable to both bulk and single-cell Hi-C data. Second, DiffDomain computes the largest eigenvalue λ_N of a properly normalized difference contact matrix $\boldsymbol{D}$, enabling the quantification of the differences of a TAD using all chromatin interactions within the TAD. Third, leveraging asymptotic distribution of λ_N, DiffDomain computes theoretical P values which is much faster in computation than simulation methods used in alternative methods. The model assumptions are realistic (Methods). Last but not the least, the normalized difference contact matrix $\boldsymbol{D}$ can help pinpoint genomic regions with increased or decreased chromatin interactions within the reorganized TAD, enabling model interpretation and refined integrative analysis with other genomic and epigenomic data.

There is room for improvement. First, DiffDomain has the highest accuracy in detecting truly reorganized TADs, but it misses some truly reorganized TADs that only show subtle structural changes (see example in Fig. 2D, G). Developing more powerful model-based methods is future work. The manually created list of gold-standard reorganized TADs is deposited in the GitHub repository that hosts the source code, which would benefit the research community for better method development. Second, because of hierarchy of TADs and sub-TADs [2, 33], generalizations of DiffDomain to explicitly consider dependencies among TADs to further refine reorganized TADs identification and classification is future work. Third, it would be desirable to generalizing DiffDomain to compare other TAD-like domains [58–63]. scHi-C data is imputed by scHiCluster [56] for the characterization of cell-to-population and cell-to-cell variability of TADs. Benchmarking the effects of different imputation algorithms, including Higashi [26, 64] and scVI-3D [65], on quantifying cell-to-population and cell-to-cell variability of TADs is future work.

As a subset of TADs, the identified reorganized TADs could be a critical unit for refined integrative analyses of multi-omics data. As a proof-of-concept, we demonstrate this type of application on different human cell lines and diseases states including SARS-CoV-2 infected A549-ACE2 cells. Future work integrating multiple types of omics data and functional perturbation experiments [66] is necessary to elucidate the causal relationships between TAD reorganization and disease.

DiffDomain is a interpretable statistical method for enhanced comparative analysis of TADs and it works for both bulk and single-cell Hi-C data. The accelerated application of Hi-C and scHi-C mapping technologies would generate ever-growing numbers of bulk and single-cell Hi-C data from different health and disease states. DiffDomain and its future generalizations would be essential part of the Hi-C analysis toolkit for the emerging comparative analysis of TADs, which in turn would advance understanding of the genome's structure-function relationship in health and disease.

Methods

In this section, we first introduce the first part of DiffDomain: model-based method to identify reorganized TADs. We state the model assumptions and their verification using real Hi-C data before report the second part of DiffDomain: classification of reorganized TADs into six subtypes. Last is software availability.

Model-based method to identify reorganized TADs

In this paper, our aim is to identify a subset of TADs that are reorganized between two biological conditions, such as a pair of healthy and disease cell lines/tissues. Specifically, given a set of TADs identified in one biological condition, we aim to identify the subset of TADs that are reorganized in another biological condition. To achieve this goal, we develop DiffDomain that takes a set of TADs and their Hi-C contact matrices as the input. The core of DiffDomain is converting the comparison of Hi-C contact matrices into a hypothesis testing problem on their difference matrix, enabling us to borrow well-established theoretical results in high-dimensional random matrix theory.

Before explaining the hypothesis testing problem, we first introduce some mathematical notations and normalization operations. Let $\boldsymbol{A} = (A_{ij}) \in R_{\geq 0}^{N \times N}$ be a symmetric Hi-C contact matrix, where N rows represent the N consecutive and equal-length chromosome bins on the linear genome and $A_{ij} \in R_{\geq 0}$ represents non-negative Hi-C contact frequency between chromosome bins i and j, $1 \leq i, j \leq N$. Given two biological conditions 1 and 2, and a TAD with N chromosome bins in the biological condition 1, let $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ be the Hi-C contact matrices between the N bins in the TAD in the biological conditions 1 and 2, respectively. It is well-known that the Hi-C contact frequency A_{ij} exponentially decreases with a increased linear distance between bins i and j. We first log-transform the Hi-C contact matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ and compute their entry-wise differences, denoted by $\boldsymbol{D}$, as shown in Eq.(1).

$$\boldsymbol{D} = \log(\boldsymbol{A}_1) - \log(\boldsymbol{A}_2). \tag{1}$$

Values in Hi-C contact matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ could have large differences because of variation in read depths. For example, the GM12878 Hi-C experiment has 4.76 times more Hi-C contacts than the K562 Hi-C experiment (Table S2). Among the 889 GM12878 TADs on Chromosome 1, the averages in the 889 $\boldsymbol{D}$s range from 1.479 to 2.611, with a median at 2.229. To adjust for the differences due to variation in read depths, we normalize $\boldsymbol{D} = (D_{ij})_{i,j=1}^{N}$ by standardizing each of its k-off diagonal block by

$$(D_{ij} - \hat{\mu}_k)/\hat{\sigma}_k, \tag{2}$$

where $k = j - i$, $2 - N \leq k \leq N - 2$, $\hat{\mu}_k$ and $\hat{\sigma}_k$ are the sample mean and standard deviation of $(D_{mn})_{1 \leq m,n \leq N, n-m=k}$. Here, without abuse of notations, we continue to use $\boldsymbol{D}$ to denote the resulted normalized difference matrix. Note that the normalization is TAD-specific because two different TADs most likely have different $\hat{\mu}_k$ and $\hat{\sigma}_k$, $2 - N \leq k \leq N - 2$. Besides visualization in Fig. 1A-C, the effects of the above procedures for both bulk and single-cell Hi-C matrices from the same TAD are also visualized in Supplementary Fig. S1.

Intuitively, if a TAD does not undergo structural reorganization from biological condition 1 to biological condition 2, the differences between $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are caused by multiple factors including variation in read depths and random perturbations of 3D genome organization. Thus, we assume that $\boldsymbol{D}$ resembles a white noise matrix. Specifically, we assume that $\boldsymbol{D}/\sqrt{N}$ is a generalized Wigner matrix (detailed assumptions and justification are presented in the next subsection), one type of well-studied random matrix. The problem of identifying reorganized TADs is reformulated as the following hypothesis testing

problem:

$$H_0 : \boldsymbol{D}/\sqrt{N} \text{ resembles a generalized Wigner matrix,}$$
$$H_1 : \boldsymbol{D}/\sqrt{N} \text{ does not resemble a generalized Wigner matrix.}$$

The largest eigenvalue of $\boldsymbol{D}/\sqrt{N}$, denoted by λ_N, converges to 2 with increased N. This result helps us to reformulate the hypothesis testing problem as the following:

$$H_0 : \lambda_N = 2 \text{ vs. } H_1 : \lambda_N > 2. \tag{3}$$

Under H_0, $\theta_N = N^{2/3}(\lambda_N - 2)$, a normalized λ_N, converges in distribution to Tracy-Widom distribution with index $\beta = 1$, denoted as TW_1.

$$\theta_N \xrightarrow{d} TW_1. \tag{4}$$

In other words, under H_0, a TAD does not undergo structural reorganization in biological condition 2. Then the fluctuations of θ_N is governed by Tracy-Widom distribution TW_1. Thus, we choose θ_N as the test statistic and compute P value by

$$P \text{ value} = P_{TW_1}(\theta_N \geq x). \tag{5}$$

A smaller P value means that the TAD is more likely to be reorganized in condition 2.

For a set of TADs, P values are adjusted for multiple comparisons by a few methods with BH method as the default. The pseudocode of our DiffDomain algorithm is presented in Supplemental Methods A.1.

Model assumptions and their verifications

Given two KR-normalized Hi-C contact matrices, DiffDomain computes the normalized difference matrix $\boldsymbol{D}$, bypassing complicated further normalization of individual Hi-C contact matrices [67]. Thus, DiffDomain makes no explicit assumptions on the individual Hi-C contact matrices. DiffDomain only makes assumptions on the normalized difference matrix $\boldsymbol{D}$. First, under H_0, DiffDomain assumes that $\boldsymbol{D}/\sqrt{N}$ is a generalized Wigner matrix: a symmetric random matrix with independent mean zero upper diagonal entries. Symmetry is satisfied by $\boldsymbol{D}/\sqrt{N}$ because Hi-C contact matrices are symmetric. The independence assumption on the upper diagonal entries is violated by $\boldsymbol{D}/\sqrt{N}$ considering the well-known fact that Hi-C contact frequencies positively correlate with each other among nearby chromosome bins. However, the violation of the independence assumption does not substantially alter the properties of $\boldsymbol{D}/\sqrt{N}$ and Diffdomain (Supplementary Results B.1). Briefly, the empirical properties of $\boldsymbol{D}/\sqrt{N}$ and DiffDomain resemble the following theoretical properties: (i) empirical spectral distribution of a generalized Wigner matrix converging to the well-established semicircle law, (ii) $\lambda_N \to 2$ (Supplementary Results B.1, Supplementary Fig. S2A-D), (iii) unadjusted P values following a uniform distribution when H_0 is true and model assumptions are satisfied (Supplementary Results B.1, Supplementary Fig. S2E-F). The key result (4) requires one more assumption. It holds under the condition that the distributions of entries in generalized Wigner matrices have vanishing third moments as N tends to infinity [68]. After the standardization procedure (2), $D_{ij}/\sqrt{N}$ approximately follows a Gaussian distribution $N(0, 1/N)$ whose third moment is 0, satisfying the vanishing third moment assumption. Taken together, assumptions of DiffDomain are appropriate.

Reorganized TAD classification

Once a subset of reorganized TADs are identified, the classification of reorganized TADs is critical to interpret TAD reorganization and link them to dynamics of genome functions. Motivated by classifications in previous studies [10, 30, 31], DiffDomain classifies reorganized TADs into six subtypes: *strength-change*, *loss*, *split*, *merge*, *zoom*, and *complex* (depending on dynamics of TAD boundaries, Fig. 1F). More details are presented in Supplemental Methods A.2.

Software availability

The software is published under the GNU GPL v3.0 license. The source code of DiffDomain is available at https://github.com/Tian-Dechao/diffDomain.

Competing interest statement

The authors declare no competing interests.

Acknowledgement

This work was supported by the National Natural Science Foundation of China grant 12271536 (D.T.), National Key Research and Development Program of China grant 2021YFC2300102 (D.T.), GuangDong Basic and Applied Basic Research Foundation grant 2022A1515010043 (D.T.), Shenzhen Sustainable Research grant KCXFZ20211020172545006 (D.T.), National Natural Science Foundation of China grant 12171198 (Z.B.), and Jilin Provincial Foundation grant 20210101147JC (Z.B.). We thank Jiang Hu for helpful discussion on the theorectical properties of the proposed method, Jun Ding and Yang Zhang for help discussion that improved the manuscript, Jian Ma for helpful comments to improve the manuscript.

Author contributions

Conceptualization: X.Z. and D.T.; Methodology: M.G. and D.T.; Software: M.G. and D.T.; Investigation: D.H., M.G., X.Z, Y.D., H.X. and D.T.; Writing-Original Draft: D.T.; Writing-Review and Editing: L.Q., X.D., Z.B., X.Z. and D.T. Funding Acquisition: Z.B. and D.T.

References

[1] Erez Lieberman-Aiden, Nynke L Van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R Lajoie, Peter J Sabo, Michael O Dorschner, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326 (5950):289–293, 2009.

[2] Jesse R Dixon, Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S Liu, and Bing Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376–380, 2012.

[3] Elphège P Nora, Bryan R Lajoie, Edda G Schulz, Luca Giorgetti, Ikuhiro Okamoto, Nicolas Servant, Tristan Piolot, Nynke L Van Berkum, Johannes Meisig, John Sedat, et al. Spatial partitioning of the regulatory landscape of the x-inactivation centre. *Nature*, 485(7398):381–385, 2012.

[4] Li-Hsin Chang, Sourav Ghosh, and Daan Noordermeer. Tads and their borders: free movement or building a wall? *Journal of Molecular Biology*, 432(3):643–652, 2020.

[5] Jesse R Dixon, Inkyung Jung, Siddarth Selvaraj, Yin Shen, Jessica E Antosiewicz-Bourget, Ah Young Lee, Zhen Ye, Audrey Kim, Nisha Rajagopal, Wei Xie, et al. Chromatin architecture reorganization during stem cell differentiation. *Nature*, 518(7539):331, 2015.

[6] Denes Hnisz, Abraham S Weintraub, Daniel S Day, Anne-Laure Valton, Rasmus O Bak, Charles H Li, Johanna Goldmann, Bryan R Lajoie, Zi Peng Fan, Alla A Sigova, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*, 351(6280):1454–1458, 2016.

[7] Darío G Lupiáñez, Katerina Kraft, Verena Heinrich, Peter Krawitz, Francesco Brancati, Eva Klopocki, Denise Horn, Hülya Kayserili, John M Opitz, Renata Laxova, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, 161 (5):1012–1025, 2015.

[8] Jie Xu, Fan Song, Huijue Lyu, Mikoto Kobayashi, Baozhen Zhang, Ziyu Zhao, Ye Hou, Xiaotao Wang, Yu Luan, Bei Jia, et al. Subtype-specific 3D genome alteration in acute myeloid leukaemia. *Nature*, 611:387–398, 2022.

[9] Peter Hugo Lodewijk Krijger, Bruno Di Stefano, Elzo de Wit, Francesco Limone, Chris Van Oevelen, Wouter De Laat, and Thomas Graf. Cell-of-origin-specific 3d genome structure acquired during somatic cell reprogramming. *Cell stem cell*, 18(5):597–610, 2016.

[10] Yuwen Ke, Yanan Xu, Xuepeng Chen, Songjie Feng, Zhenbo Liu, Yaoyu Sun, Xuelong Yao, Fangzhen Li, Wei Zhu, Lei Gao, et al. 3d chromatin structures of mature gametes and structural reprogramming during mammalian embryogenesis. *Cell*, 170(2):367–381, 2017.

[11] Jia Wang, Haopeng Yu, Qian Ma, Pengguihang Zeng, Danya Wu, Yingping Hou, Xinyi Liu, Lumeng Jia, Jun Sun, Yilong Chen, et al. Phase separation of oct4 controls tad reorganization to promote cell fate transitions. *Cell Stem Cell*, 28(10):1868–1883, 2021.

[12] Warren Winick-Ng, Alexander Kukalev, Izabela Harabula, Luna Zea-Redondo, Dominik Szabó, Mandy Meijer, Leonid Serebreni, Yingnan Zhang, Simona Bianco, Andrea M Chiariello, et al. Cell-type specialization is encoded by specific chromatin topologies. *Nature*, 599:684–691, 2021.

[13] Ittai E Eres and Yoav Gilad. A TAD skeptic: is 3D genome topology conserved? *Trends in Genetics*, 37(3):216–223, 2021.

[14] Alessa R Ringel, Quentin Szabo, Andrea M Chiariello, Konrad Chudzik, Robert Schöpflin, Patricia Rothe, Alexandra L Mattei, Tobias Zehnder, Dermot Harnett, Verena Laupert, et al. Repression and 3D-restructuring resolves regulatory conflicts in evolutionarily rearranged genomes. *Cell*, 185 (20):3689–3704, 2022.

[15] Takashi Nagano, Yaniv Lubling, Tim J Stevens, Stefan Schoenfelder, Eitan Yaffe, Wendy Dean,

Ernest D Laue, Amos Tanay, and Peter Fraser. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, 502(7469):59–64, 2013.

[16] Vijay Ramani, Xinxian Deng, Ruolan Qiu, Kevin L Gunderson, Frank J Steemers, Christine M Disteche, William S Noble, Zhijun Duan, and Jay Shendure. Massively multiplex single-cell Hi-C. *Nature Methods*, 14(3):263–266, 2017.

[17] Ilya M Flyamer, Johanna Gassler, Maxim Imakaev, Hugo B Brandão, Sergey V Ulianov, Nezar Abdennur, Sergey V Razin, Leonid A Mirny, and Kikuë Tachibana-Konwalski. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature*, 544(7648): 110–114, 2017.

[18] Bogdan Bintu, Leslie J Mateo, Jun-Han Su, Nicholas A Sinnott-Armstrong, Mirae Parker, Seon Kinrot, Kei Yamaya, Alistair N Boettiger, and Xiaowei Zhuang. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, 362(6413), 2018.

[19] Elizabeth H Finn, Gianluca Pegoraro, Hugo B Brandao, Anne-Laure Valton, Marlies E Oomen, Job Dekker, Leonid Mirny, and Tom Misteli. Extensive heterogeneity and intrinsic variation in spatial genome organization. *Cell*, 176(6):1502–1515, 2019.

[20] Quentin Szabo, Axelle Donjon, Ivana Jerković, Giorgio L Papadopoulos, Thierry Cheutin, Boyan Bonev, Elphège P Nora, Benoit G Bruneau, Frédéric Bantignies, and Giacomo Cavalli. Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nature Genetics*, 52 (11):1151–1157, 2020.

[21] Michele Gabriele, Hugo B Brandão, Simon Grosse-Holz, Asmita Jha, Gina M Dailey, Claudia Cattoglio, Tsung-Han S Hsieh, Leonid Mirny, Christoph Zechner, and Anders Sejr Hansen. Dynamics of ctcf and cohesin mediated chromatin looping revealed by live-cell imaging. 2021. doi: 10.1101/2021.12.12.472242.

[22] Tom Misteli. The self-organizing genome: principles of genome architecture and function. *Cell*, 183(1):28–45, 2020.

[23] Natalie Sauerwald and Carl Kingsford. Quantifying the similarity of topological domains across normal and cancer human cell types. *Bioinformatics*, 34(13):i475–i483, 2018.

[24] Anthony M D'Ippolito, Ian C McDowell, Alejandro Barrera, Linda K Hong, Sarah M Leichter, Luke C Bartelt, Christopher M Vockley, William H Majoros, Alexias Safi, Lingyun Song, et al. Pre-established chromatin interactions mediate the genomic response to glucocorticoids. *Cell systems*, 7(2):146–160, 2018.

[25] Guangyu Wang, Qingshu Meng, Bo Xia, Shuo Zhang, Jie Lv, Dongyu Zhao, Yanqiang Li, Xin Wang, Lili Zhang, John P Cooke, et al. TADsplimer reveals splits and mergers of topologically associating domains for epigenetic regulation of transcription. *Genome biology*, 21(1):1–16, 2020.

[26] Ruochi Zhang, Tianming Zhou, and Jian Ma. Multiscale and integrative single-cell hi-c analysis with higashi. *Nature Biotechnology*, 2021.

[27] Raphaël Mourad. TADreg: a versatile regression framework for TAD identification, differential analysis and rearranged 3D genome prediction. *BMC Bioinformatics*, 23(1):1–14, 2022.

[28] Qing You Pang, Tuan Zea Tan, Vignesh Sundararajan, Yi-Chia Chiu, Edward Yu Wing Chee, Vin Yee Chung, Mahesh A. Choolani, and Ruby Yun-Ju Huang. 3D genome organization in the epithelial-mesenchymal transition spectrum. *Genome Biology*, 23(1):121, 2022.

[29] Rafal Zaborowski and Bartek Wilczynski. DiffTAD: Detecting differential contact frequency in topologically associating domains hi-c experiments between conditions. 2016. doi: 10.1101/093625.

[30] Huiling Liu and Wenxiu Ma. DiffGR: Detecting differentially interacting genomic regions from

Hi-c contact maps. *bioRxiv*, 2020.

[31] Kellen G Cresswell and Mikhail G Dozmorov. TADCompare: an R package for differential and temporal analysis of topologically associated domains. *Frontiers in genetics*, 11:158, 2020.

[32] Stephen Richer, Tian Yuan, Stefan Schoenfelder, Laurence Hurst, Adele Murrell, and Giuseppina Pisignano. Widespread allele-specific topological domains in the human genome are not confined to imprinted gene clusters. *bioRxiv*, 2022.

[33] Suhas SP Rao, Miriam H Huntley, Neva C Durand, Elena K Stamenova, Ivan D Bochkov, James T Robinson, Adrian L Sanborn, Ido Machol, Arina D Omer, Eric S Lander, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7): 1665–1680, 2014.

[34] Tao Yang, Feipeng Zhang, Galip Gürkan Yardımcı, Fan Song, Ross C Hardison, William Stafford Noble, Feng Yue, and Qunhua Li. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Research*, 27(11):1939–1949, 2017.

[35] Silvia Galan, Nick Machnik, Kai Kruse, Noelia Díaz, Marc A Marti-Renom, and Juan M Vaquerizas. Chess enables quantitative comparison of chromatin contact data and automatic feature extraction. *Nature Genetics*, 52(11):1247–1255, 2020.

[36] Abhijit Chakraborty, Jeffrey G Wang, and Ferhat Ay. dcHiC detects differential compartments across multiple Hi-C datasets. *Nature Communications*, 13(1):1–21, 2022.

[37] John C Stansfield, Kellen G Cresswell, and Mikhail G Dozmorov. multiHiCcompare: joint normalization and comparative analysis of complex Hi-C experiments. *Bioinformatics*, 35(17):2916–2923, 2019.

[38] Merve Sahin, Wilfred Wong, Yingqian Zhan, Kinsey Van Deynze, Richard Koche, and Christina S Leslie. HiC-DC+ enables systematic 3D interaction calls and differential analysis for Hi-C and HiChIP. *Nature Communications*, 12:3366, 2021.

[39] Abbas Roayaei Ardakany, Ferhat Ay, and Stefano Lonardi. Selfish: discovery of differential chromatin interactions via a self-similarity measure. *Bioinformatics*, 35(14):i145–i153, 2019.

[40] Miao Yu, Armen Abnousi, Yanxiao Zhang, Guoqiang Li, Lindsay Lee, Ziyin Chen, Rongxin Fang, Taylor M Lagler, Yuchen Yang, Jia Wen, et al. SnapHiC: A computational pipeline to identify chromatin loops from single-cell Hi-C data. *Nature Methods*, 18(9):1056–1059, 2021.

[41] Iain M Johnstone and Zongming Ma. Fast approach to the Tracy-Widom law at the edge of GOE and GUE. *the Annals of Applied Probability*, 22(5):1962, 2012.

[42] Juan Wang, Tina Yi-Ting Huang, Ye Hou, Elizabeth Bartom, Xinyan Lu, Ali Shilatifard, Feng Yue, and Amanda Saratsis. Epigenomic landscape and 3D genome structure in pediatric high-grade glioma. *Science Advances*, 7(23):eabg4126, 2021.

[43] John C Stansfield, Kellen G Cresswell, Vladimir I Vladimirov, and Mikhail G Dozmorov. HiCcompare: an R-package for joint normalization and comparison of HI-C datasets. *BMC Bioinformatics*, 19(1):1–10, 2018.

[44] Tao Yang, Feipeng Zhang, Galip Gürkan Yardımcı, Fan Song, Ross C Hardison, William Stafford Noble, Feng Yue, and Qunhua Li. Hicrep: assessing the reproducibility of hi-c data using a stratum-adjusted correlation coefficient. *Genome research*, 27(11):1939–1949, 2017.

[45] Anna-Karin Olsson, Anna Dimberg, Johan Kreuger, and Lena Claesson-Welsh. VEGF receptor signalling? In control of vascular function. *Nature reviews Molecular cell biology*, 7(5):359, 2006.

[46] Robert S Kerbel. Tumor angiogenesis. *New England Journal of Medicine*, 358(19):2039–2049, 2008.

[47] Denes Hnisz, Brian J Abraham, Tong Ihn Lee, Ashley Lau, Violaine Saint-André, Alla A Sigova,

Heather A Hoke, and Richard A Young. Super-enhancers in the control of cell identity and disease. *Cell*, 155(4):934–947, 2013.

[48] Jonas Paulsen, Tharvesh Moideen Liyakat Ali, and Philippe Collas. Computational 3d genome modeling using chrom3d. *Nature protocols*, 13(5):1137–1152, 2018.

[49] Xiaotao Wang, Yu Luan, and Feng Yue. EagleC: A deep-learning framework for detecting a full range of structural variations from bulk and single-cell contact maps. *Science Advances*, 8(24): eabn9215, 2022.

[50] Ruoyu Wang, Joo-Hyung Lee, Feng Xiong, Jieun Kim, Lana Al Hasani, Xiaoyi Yuan, Pooja Shivshankar, Joanna Krakowiak, Chuangye Qi, Yanyu Wang, Holger K. Eltzschig, and Wenbo Li. SARS-CoV-2 restructures the host chromatin architecture. 2021. doi: 10.1101/2021.07.20.453146.

[51] Tianming Zhou, Ruochi Zhang, and Jian Ma. The 3d genome structure of single cells. *Annual Review of Biomedical Data Science*, 4:21–41, 2021.

[52] Aleksandra A Galitsyna and Mikhail S Gelfand. Single-cell hi-c data analysis: safety in numbers. *Briefings in Bioinformatics*, 2021.

[53] Longzhi Tan, Wenping Ma, Honggui Wu, Yinghui Zheng, Dong Xing, Ritchie Chen, Xiang Li, Nicholas Daley, Karl Deisseroth, and X Sunney Xie. Changes in genome architecture and transcriptional dynamics progress independently of sensory experience during post-natal brain development. *Cell*, 184(3):741–758, 2021.

[54] Yan Jiang, Yong-Hwee Eddie Loh, Prashanth Rajarajan, Teruyoshi Hirayama, Will Liao, Bibi S Kassim, Behnam Javidfar, Brigham J Hartley, Lisa Kleofas, Royce B Park, et al. The methyltransferase setdb1 regulates a large neuron-specific topological chromatin domain. *Nature Genetics*, 49 (8):1239–1250, 2017.

[55] Sevil Sofueva, Eitan Yaffe, Wen-Ching Chan, Dimitra Georgopoulou, Matteo Vietri Rudan, Hegias Mira-Bontenbal, Steven M Pollard, Gary P Schroth, Amos Tanay, and Suzana Hadjur. Cohesin-mediated interactions organize chromosomal domain architecture. *The EMBO Journal*, 32(24): 3119–3129, 2013.

[56] Jingtian Zhou, Jianzhu Ma, Yusi Chen, Chuankai Cheng, Bokan Bao, Jian Peng, Terrence J Sejnowski, Jesse R Dixon, and Joseph R Ecker. Robust single-cell hi-c clustering by convolution- and random-walk–based imputation. *Proceedings of the National Academy of Sciences*, 116(28): 14011–14018, 2019.

[57] Tim J Stevens, David Lando, Srinjan Basu, Liam P Atkinson, Yang Cao, Steven F Lee, Martin Leeb, Kai J Wohlfahrt, Wayne Boucher, Aoife OShaughnessy-Kirwan, et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature*, 544(7648):59–64, 2017.

[58] Xiao Li, Guangjie Zeng, Angsheng Li, and Zhihua Zhang. DeTOKI identifies and characterizes the dynamics of chromatin TAD-like domains in a single cell. *Genome Biology*, 22(1):1–26, 2021.

[59] Angsheng Li, Guangjie Zeng, WANG Haoyu, Xiao Li, and Zhihua Zhang. DeDoc2 identifies and characterizes the hierarchy and dynamics of chromatin TAD-like domains in the single cells. *bioRxiv*, 2022.

[60] Tsung-Han S Hsieh, Claudia Cattoglio, Elena Slobodyanyuk, Anders S Hansen, Oliver J Rando, Robert Tjian, and Xavier Darzacq. Resolving the 3D landscape of transcription-linked mammalian chromatin folding. *Molecular Cell*, 78(3):539–553, 2020.

[61] Nils Krietenstein, Sameer Abraham, Sergey V Venev, Nezar Abdennur, Johan Gibcus, Tsung-Han S Hsieh, Krishna Mohan Parsi, Liyan Yang, René Maehr, Leonid A Mirny, et al. Ultrastructural details of mammalian chromosome architecture. *Molecular Cell*, 78(3):554–565, 2020.

[62] Douglas H Phanstiel, Kevin Van Bortle, Damek Spacek, Gaelen T Hess, Muhammad Saad Shamim,

Ido Machol, Michael I Love, Erez Lieberman Aiden, Michael C Bassik, and Michael P Snyder. Static and dynamic DNA loops form AP-1-bound activation hubs during macrophage development. *Molecular Cell*, 67(6):1037–1048, 2017.

[63] Jonathan A Beagan and Jennifer E Phillips-Cremins. On the existence and functionality of topologically associating domains. *Nature Genetics*, 52(1):8–16, 2020.

[64] Ruochi Zhang, Tianming Zhou, and Jian Ma. Ultrafast and interpretable single-cell 3D genome analysis with Fast-Higashi. *Cell Systems*, 13(10):798–807, 2022.

[65] Ye Zheng, Siqi Shen, and Sündüz Keleş. Normalization and de-noising of single-cell Hi-C data with BandNorm and scVI-3D. *Genome Biology*, 23(1):1–34, 2022.

[66] Haifeng Wang, Mengting Han, and Lei S Qi. Engineering 3D genome organization. *Nature Reviews Genetics*, 22(6):343–360, 2021.

[67] Anthony D Schmitt, Ming Hu, and Bing Ren. Genome-wide mapping and analysis of chromosome architecture. *Nature Reviews Molecular Cell Biology*, 17(12):743–755, 2016.

[68] Terence Tao and Van Vu. Random matrices: Universality of local eigenvalue statistics up to the edge. *Communications in Mathematical Physics*, 298(2):549–572, 2010.

[69] Kyle Xiong and Jian Ma. Revealing hi-c subcompartments by imputing inter-chromosomal chromatin interactions. *Nature communications*, 10(1):1–12, 2019.

[70] Hongbo Yang, Yu Luan, Tingting Liu, Hyung Joo Lee, Li Fang, Yanli Wang, Xiaotao Wang, Bo Zhang, Qiushi Jin, Khai Chung Ang, et al. A map of cis-regulatory elements and 3d genome structures in zebrafish. *Nature*, 588(7837):337–343, 2020.

[71] Hui Huang, Quan Zhu, Adam Jussila, Yuanyuan Han, Bogdan Bintu, Colin Kern, Mattia Conte, Yanxiao Zhang, Simona Bianco, Andrea M Chiariello, et al. Ctcf mediates dosage-and sequence-context-dependent transcriptional insulation by forming local chromatin domains. *Nature Genetics*, 53(7).

[72] Jared V Goodman, Tomoko Yamada, Yue Yang, Lingchun Kong, Dennis Y Wu, Guoyan Zhao, Harrison W Gabel, and Azad Bonni. The chromatin remodeling enzyme chd4 regulates genome architecture in the mouse brain. *Nature communications*, 11(1):3419, 2020.

[73] Da-Inn Lee and Sushmita Roy. Grinch: simultaneous smoothing and detection of topological units of genome organization from sparse chromatin contact count matrices with matrix factorization. *Genome biology*, 22(1):1–31, 2021.

[74] Boyan Bonev, Netta Mendelson Cohen, Quentin Szabo, Lauriane Fritsch, Giorgio L Papadopoulos, Yaniv Lubling, Xiaole Xu, Xiaodan Lv, Jean-Philippe Hugnot, Amos Tanay, et al. Multiscale 3d genome rewiring during mouse neural development. *Cell*, 171(3):557–572, 2017.

[75] Tiaojiang Xiao, Xin Li, and Gary Felsenfeld. The myc-associated zinc finger protein (maz) works together with ctcf to control cohesin positioning and genome organization. *Proceedings of the National Academy of Sciences*, 118(7), 2021.

[76] Di Zhang, Peng Huang, Malini Sharma, Cheryl A Keller, Belinda Giardine, Haoyue Zhang, Thomas G Gilgenast, Jennifer E Phillips-Cremins, Ross C Hardison, and Gerd A Blobel. Alteration of genome folding via contact domain boundary insertion. *Nature Genetics*, 52(10):1076–1087, 2020.

[77] Alessandro Bertero, Paul A Fields, Vijay Ramani, Giancarlo Bonora, Galip G Yardimci, Hans Reinecke, Lil Pabon, William S Noble, Jay Shendure, and Charles E Murry. Dynamics of genome reorganization during human cardiogenesis reveal an rbm20-dependent splicing factory. *Nature communications*, 10(1):1–19, 2019.

[78] Haipeng Xing, Yingru Wu, Michael Q Zhang, and Yong Chen. Deciphering hierarchical organiza-

tion of topologically associated domains through change-point testing. *BMC bioinformatics*, 22(1): 1–23, 2021.

[79] M Jordan Rowley, Michael H Nichols, Xiaowen Lyu, Masami Ando-Kuri, I Sarahi M Rivera, Karen Hermetz, Ping Wang, Yijun Ruan, and Victor G Corces. Evolutionarily conserved principles predict 3d chromatin organization. *Molecular Cell*, 67(5):837–852, 2017.

[80] Yanxiao Zhang, Ting Li, Sebastian Preissl, Maria Luisa Amaral, Jonathan D Grinstein, Elie N Farah, Eugin Destici, Yunjiang Qiu, Rong Hu, Ah Young Lee, et al. Transcriptionally active herv-h retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nature Genetics*, 51(9):1380–1388, 2019.

[81] Ruoyu Wang, Joo-Hyung Lee, Feng Xiong, Jieun Kim, Lana Al Hasani, Xiaoyi Yuan, Pooja Shivshankar, Joanna Krakowiak, Chuangye Qi, Yanyu Wang, Holger K. Eltzschig, and Wenbo Li. SARS-CoV-2 restructures the host chromatin architecture. 2021. doi: 10.1101/2021.07.20.453146.

[82] ENCODE Project Consortium et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414):57, 2012.

[83] Debyani Chakravarty, Jianjiong Gao, Sarah Phillips, Ritika Kundra, Hongxin Zhang, Jiaojiao Wang, Julia E Rudolph, Rona Yaeger, Tara Soumerai, Moriah H Nissan, et al. Oncokb: a precision oncology knowledge base. *JCO Precision Oncology*, 1:1–16, 2017.

[84] Warren A Whyte, David A Orlando, Denes Hnisz, Brian J Abraham, Charles Y Lin, Michael H Kagey, Peter B Rahl, Tong Ihn Lee, and Richard A Young. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, 153(2):307–319, 2013.

[85] Steffen Durinck, Paul T Spellman, Ewan Birney, and Wolfgang Huber. Mapping identifiers for the integration of genomic datasets with the r/bioconductor package biomart. *Nature Protocols*, 4(8): 1184–1191, 2009.

[86] Xiaopeng Zhu, Yang Zhang, Yuchuan Wang, Dechao Tian, Andrew S. Belmont, Jason R. Swedlow, and Jian Ma. Nucleome Browser: an integrative and multimodal data navigation platform for 4D Nucleome. *Nature Methods*, 19(8):911–913, 2022.

[87] Marie Zufferey, Daniele Tavernari, Elisa Oricchio, and Giovanni Ciriello. Comparison of computational methods for the identification of topologically associating domains. *Genome Biology*, 19 (1):1–18, 2018.

[88] Angsheng Li, Xianchen Yin, Bingxiang Xu, Danyang Wang, Jimin Han, Yi Wei, Yun Deng, Ying Xiong, and Zhihua Zhang. Decoding topologically associating domains with ultra-low resolution Hi-C data by graph structural entropy. *Nature Communications*, 9(1):1–12, 2018.

[89] Marilyn Safran, Irina Dalah, Justin Alexander, Naomi Rosen, Tsippi Iny Stein, Michael Shmoish, Noam Nativ, Iris Bahir, Tirza Doniger, Hagit Krug, et al. GeneCards Version 3: the human gene integrator. *Database*, 2010, 2010.

[90] Artem Barski, Suresh Cuddapah, Kairong Cui, Tae-Young Roh, Dustin E Schones, Zhibin Wang, Gang Wei, Iouri Chepelev, and Keji Zhao. High-resolution profiling of histone methylations in the human genome. *Cell*, 129(4):823–837, 2007.

[91] Nathaniel D Heintzman, Rhona K Stuart, Gary Hon, Yutao Fu, Christina W Ching, R David Hawkins, Leah O Barrera, Sara Van Calcar, Chunxu Qu, Keith A Ching, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics*, 39(3):311, 2007.

[92] Nathaniel D Heintzman, Gary C Hon, R David Hawkins, Pouya Kheradpour, Alexander Stark, Lindsey F Harp, Zhen Ye, Leonard K Lee, Rhona K Stuart, Christina W Ching, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, 459 (7243):108, 2009.

[93] Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):1–21, 2014.

[94] Shiqi Tu, Mushan Li, Haojie Chen, Fengxiang Tan, Jian Xu, David J Waxman, Yijing Zhang, and Zhen Shao. MAnorm2 for quantitatively comparing groups of ChIP-seq samples. *Genome Research*, 31(1):131–145, 2021.

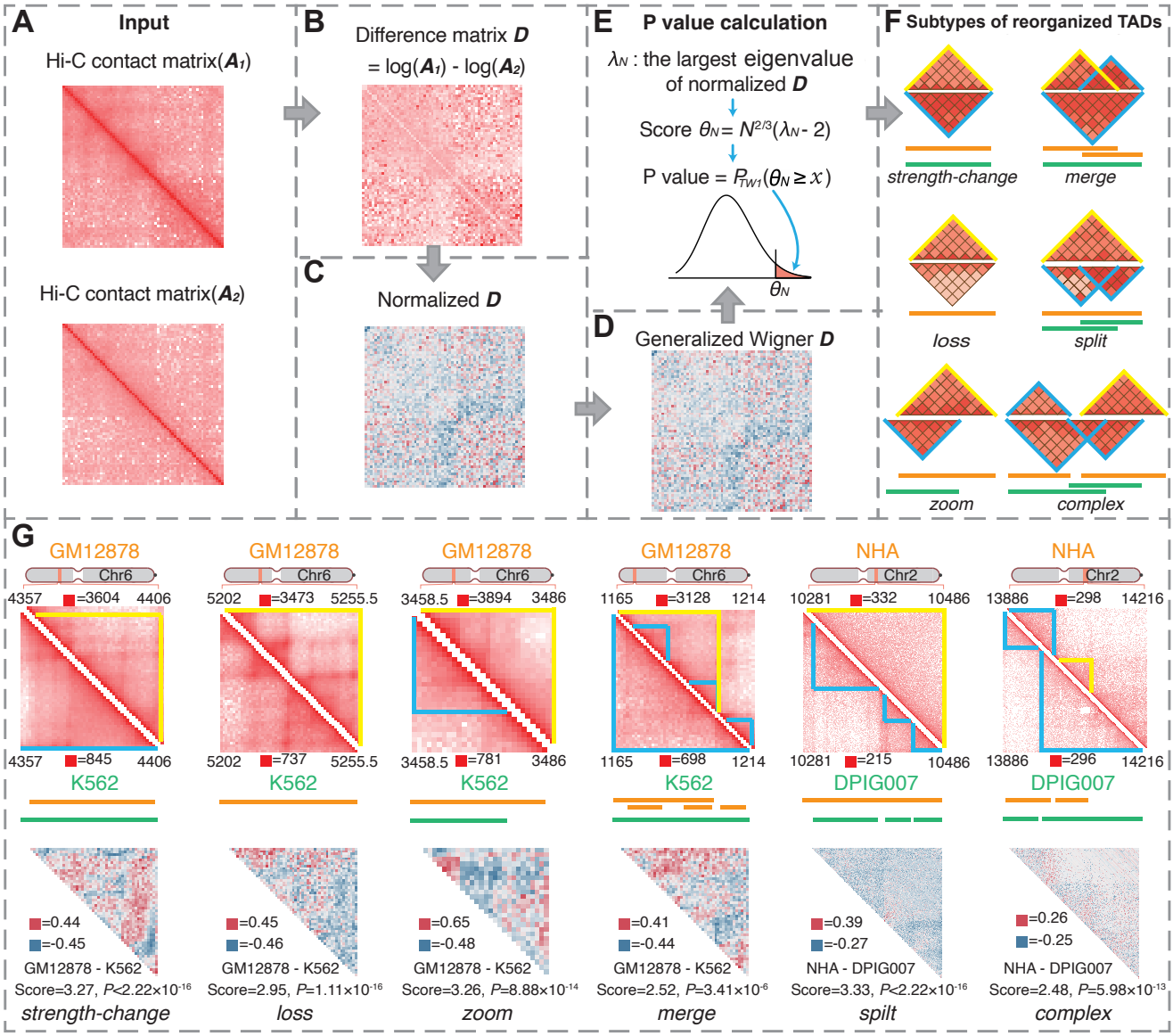

Figure 1: DiffDomain workflow and example outputs. (***A***) Input are a TAD in condition 1 and its two Hi-C contact matrices ($\boldsymbol{A}_1$ and $\boldsymbol{A}_2$) in two biological conditions 1 and 2. (***B***) Difference between log-transformed $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$, which is denoted as $\boldsymbol{D}$. (***C***) Normalization of $\boldsymbol{D}$ by a 1D distance-stratified standardization procedure. Its d-off diagonal part is normalized by d-off diagonal part-specific sample mean and sample standard deviation. (***D***) $\boldsymbol{D}$ is transformed by dividing $\sqrt{N}$. Under null hypothesis, it is assumed to be a generalized Wigner matrix. (***E***) P value is calculated based on the fact that θ_N, normalized largest eigenvalue of $\boldsymbol{D}$, follows Tracy-Widom distribution with $\beta = 1$ (denoted as TW_1 distribution). A TAD is identified as a reorganized TAD if P value ≤ 0.05. (***F***) Reorganized TADs are classified into six subtypes based on changes in TAD boundaries. (***G***) Example of the subtypes of reorganized TADs. Data are from two studies [33, 42]. *Top:* upper and lower triangular matrices represent Hi-C data in conditions 1 and 2, respectively; *Bottom:* normalized difference matrix $\boldsymbol{D}/\sqrt{N}$.

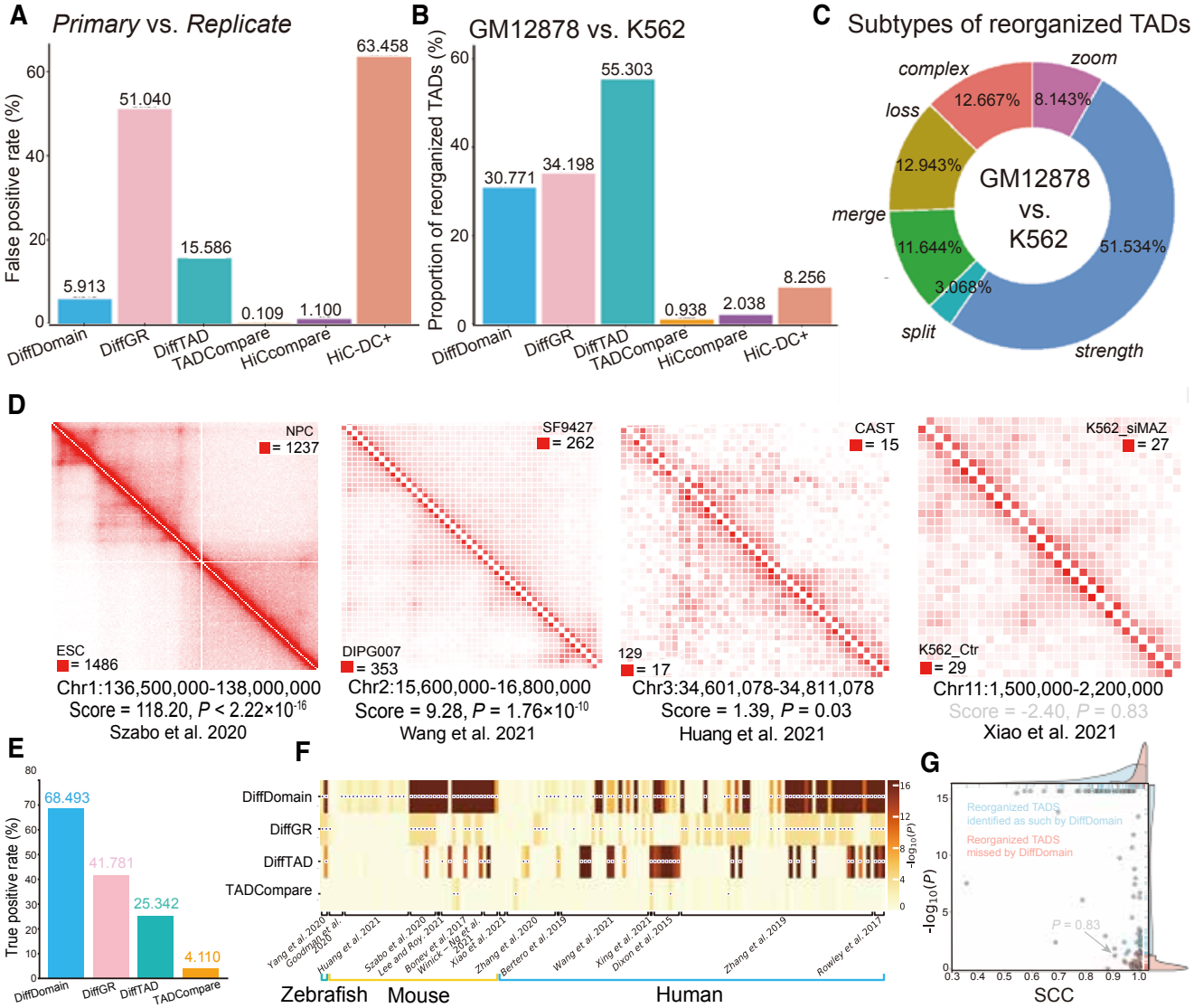

Figure 2: DiffDomain has better performance than alternative methods. (***A***) FPRs of DiffDomain and alternative methods in comparing two GM12878 Hi-C replicates (*combined* and *primary*). FPRs in comparing other Hi-C replicates of GM12878 are shown in Supplementary Fig. S5. FPR equals the ratio of the number of identified reorganized TADs to the number of TADs in GM12878. (***B***) Proportions of identified reorganized TADs by DiffDomain and alternative methods when comparing blood-related cell lines GM12878 and K562. TADs are GM12878 TADs. Results on other pairs of human cell lines are reported in Supplementary Fig. S6. (***C***) Percentages of the subtypes of reorganized TADs when comparing GM12878 and K562. Results on other pairs of human cell lines are reported in Supplementary Fig. S8. (***D***) Heatmaps showing Hi-C contact matrices of four truly reorganized TADs. Among them, three are correctly identified as such by DiffDomain (unadjusted $P \leq 0.05$, true positives) and one is missed by DiffDomain (unadjusted $P > 0.05$, false negative). Truly reorganized TADs are manually collected and treated as the gold standard positives (see Supplementary Methods A.4 for more details). (***E***) Barplot showing TPRs of DiffDomain and alternative methods. TPR equals the ratio of the number of reorganized TADs that are identified as such (true positives) to the number of reorganized TADs (positives). (***F***) Heatmap showing unadjusted P values for testing truly reorganized TADs. (***G***) Scatter points of unadjusted P values by DiffDomain (Y-axis) and SCCs [44] (X-axis) when testing the truly reorganized TADs.

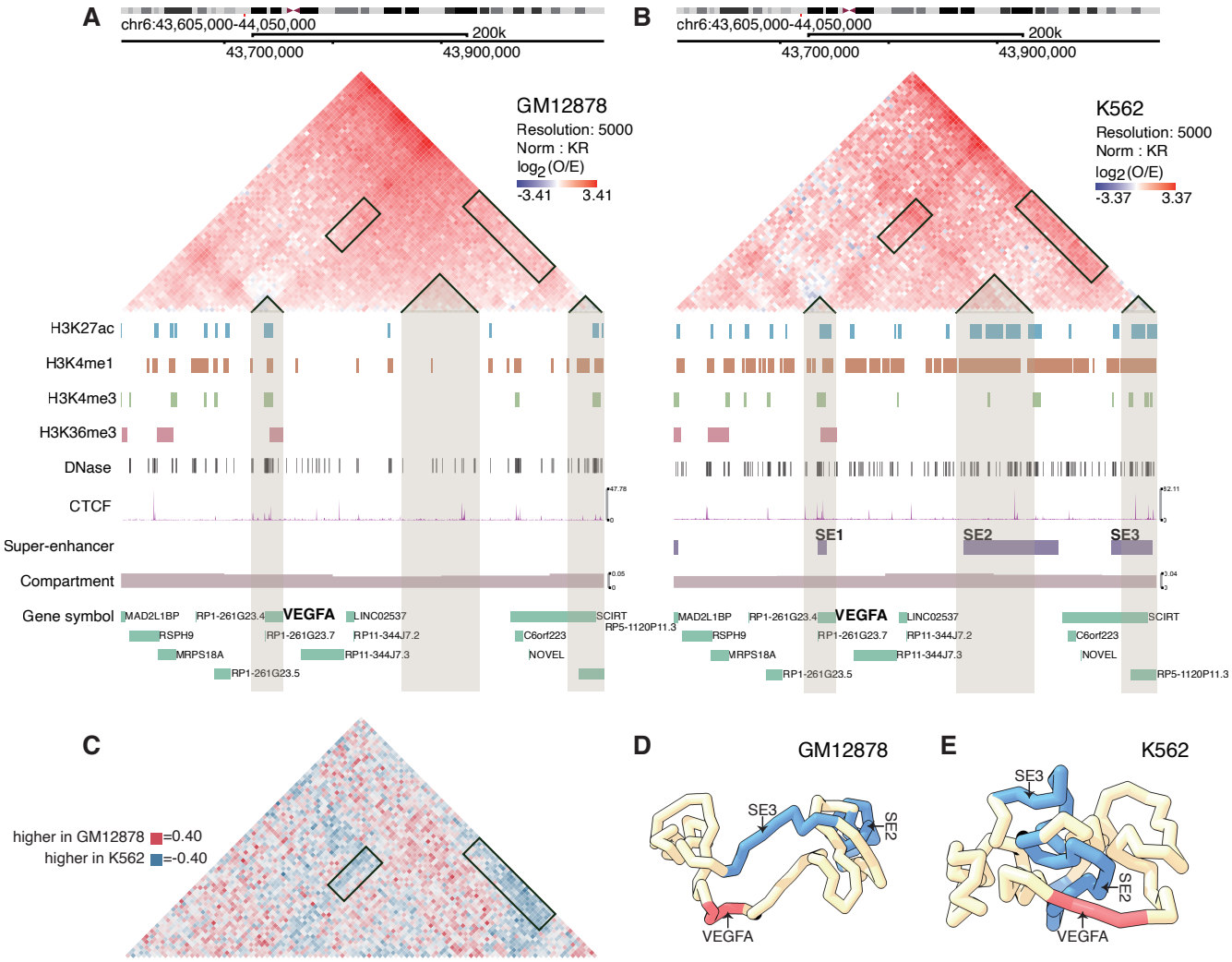

Figure 3: A reorganized TAD involves the angiogenic gene *VEGFA*. Hi-C contact matrices and 1D genomic tracks of the TAD region (Chr6:43605000-44050000) in GM12878 (***A***) and K562 (***B***) cell lines that shows differences in both Hi-C data and epigenomics profiles involving *VEGFA* gene. (***C***) The normalized difference matrix D betweeen the two cell lines highlights the differences in Hi-C contact maps. (***D-E***) Potential 3D models of the TAD in GM12878 and K562 that are estimated by Chrom3D, demonstrating the TAD structural differences between GM12878 and K562.

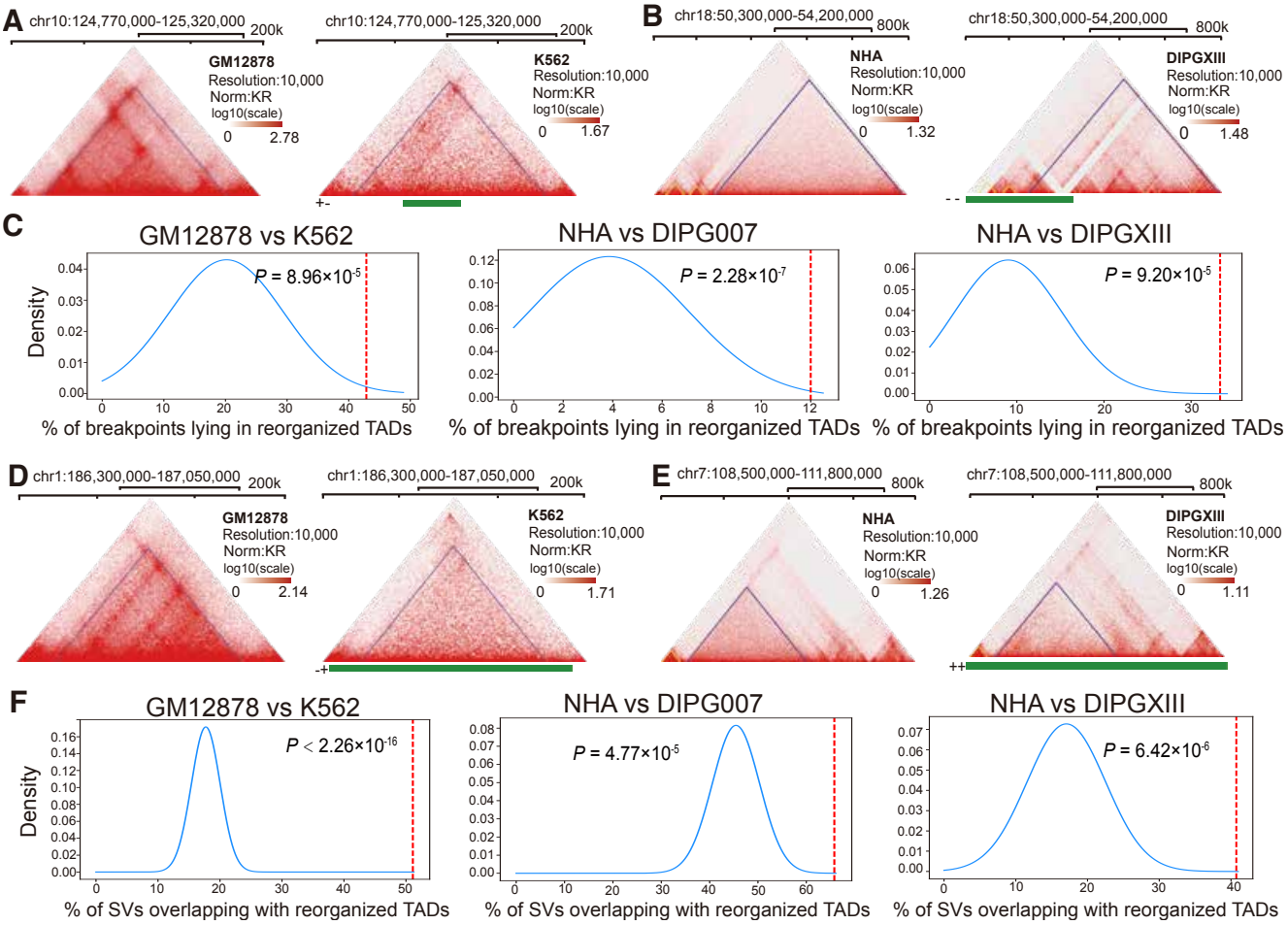

Figure 4: Enrichment of reorganized TADs with SVs. (***A-B***) Examples showing breakpoints of SVs lying in reorganized TADs. (***C***) Density curves comparing the proportions (vertical dashed lines) of breakpoints lying in reorganized TADs with the proportions (blue curves) from randomly cases. The reorganized TADs are GM12878 TADs reorganized in K562, NHA TADs reorganized in DIPG007, and NHA TADs reorganized in DIGPXIII, respectively. (***D-E***) Examples showing reorganized TADs overlapping with the genomic regions covered by SVs. (***F***) Density curves comparing the proportions (vertical dashed lines) of the genomic regions of SVs that overlap with reorganized TADs with the proportions (blue curves) from randomly cases. Here, the simulation design is randomly sampling equal-numbered reorganized TADs and the simulation is repeated 10,000 times. Abbreviations: $+-$, fusion of the 3′-end of a fragment to the 5′-end of another fragment; $++$, 3′-to-3′ fusion; $-+$, 5′-to-3′ fusion; $--$, 5′-to-5′ fusion [49]; NHA, normal human astrocytes; DIPG007 and DIPGXIII, pediatric high-grade glioma cell lines.

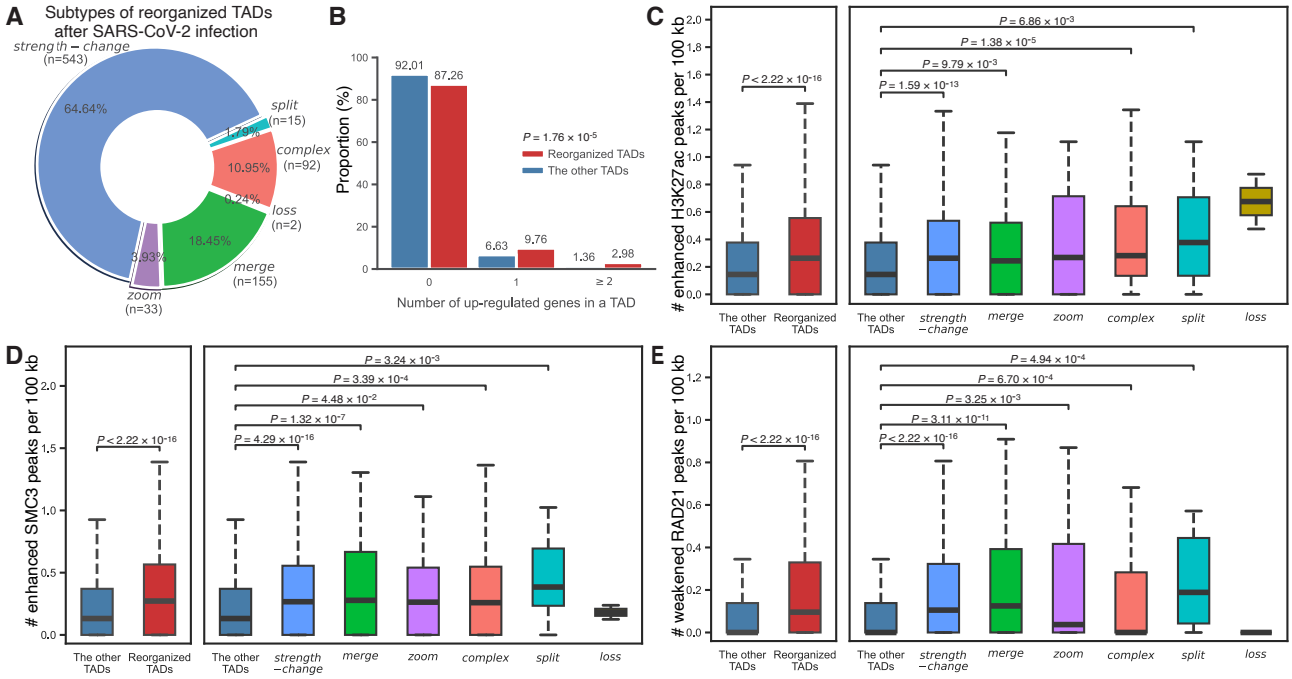

Figure 5: Mock-infected A549-ACE2 TADs that are reorganized in SARS-CoV-2 infected A549-ACE2. (***A***) Pie chart showing the percentages of subtypes of reorganized TADs in A549-ACE3 after SARS-CoV-2 infection. *Strength-change* TADs are the leading subtype. (***B***) Barplot comparing the number of up-regulated genes in the reorganized TADs and the other TADs. Boxplots comparing the numbers of enhanced H3K27ac peaks (***C***), enhanced SMC3 peaks (***D***), and weakened RAD21 peaks (***E***) per 100 kb. *Left*: comparing reorganized TADs with the other TADs; *right*: comparison stratified by the subtypes of reorganized TADs.

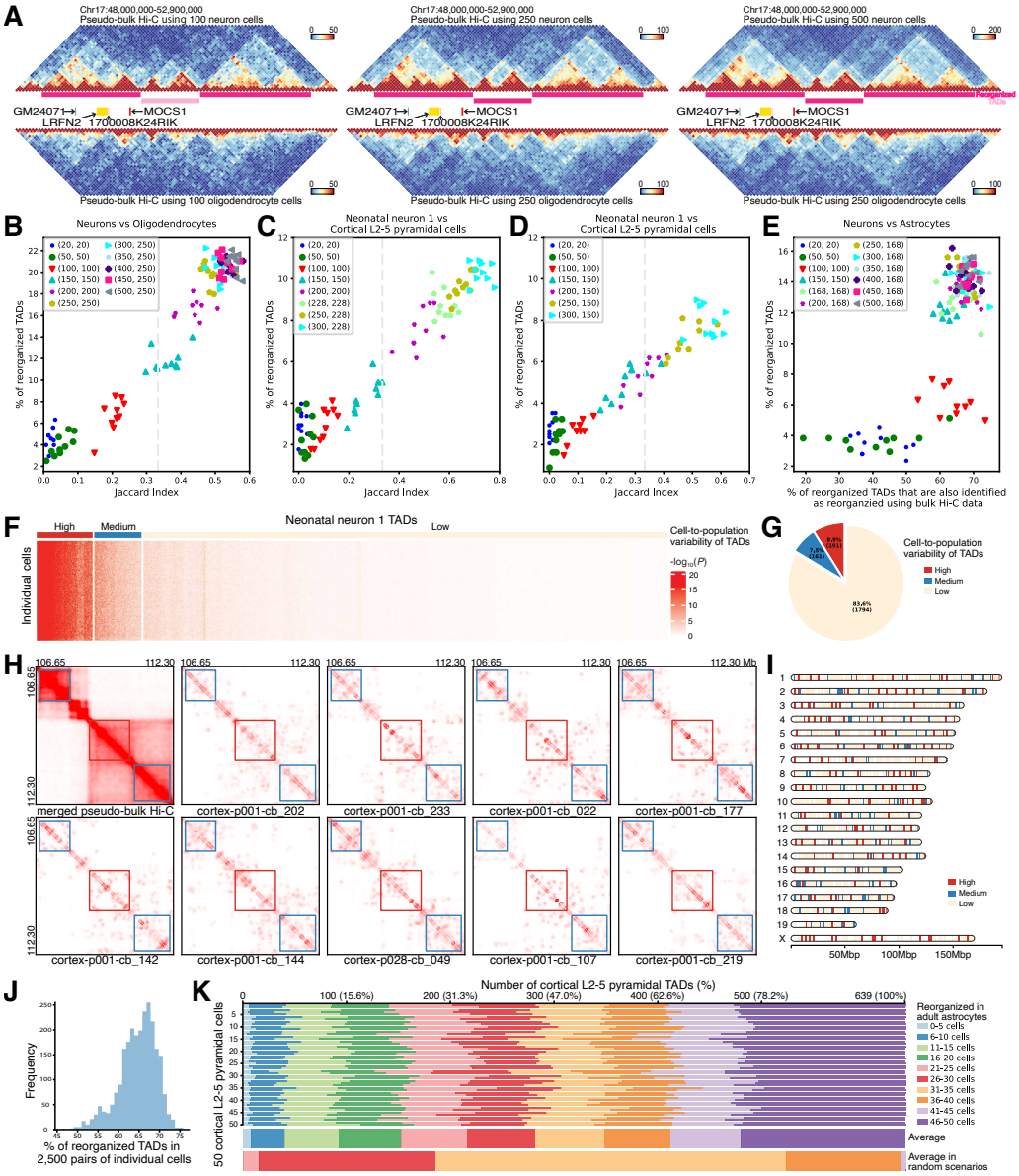

Figure 6: Application of DiffDomain on scHi-C data to identify reorganized TADs between cell types and characterize cell-to-population and cell-to-cell variability of TADs. (***A***) Visualization of the pseudo-bulk Hi-C contact maps and the identified neuronal TADs that are reorganized in oligodendrocytes (dark pink horizontal bars) using varied numbers of randomly sampled individual cells. Gene track shows four neuronal marker genes. (***B***) Scatter plot showing the proportions of neuronal TADs that are reorganized in oligodendrocytes using varied numbers (k_1, k_2) of randomly sampled individual cells from the two cell types. Their agreements with the set of reorganized TADs using the maximum numbers of cells are quantified by Jaccard index (X-axis). Vertical dashed line is $JI = 1/3$, representing that two equal-sized sets share half of reorganized TADs. (***C-D***) Scatter plots showing the proportions of neonatal neuron 1 TADs that are reorganized in cortical L2-5 pyramidal cells. Up to 228 and 150 cortical L2-5 pyramidal cells are randomly sampled, respectively. (***E***) Scatter plot showing the agreements between sets of reorganized TADs that are identified using (1) pseudo-bulk Hi-C data and (2) bulk Hi-C data. (***F***) Heatmap showing high, medium, and low cell-to-population variational TADs. P value is from comparing scHi-C contact map of a TAD in an individual cell to the pseudo-bulk Hi-C contact map that represents population average. Classification of TADs is done by hierarchical clustering. (***G***) Pie chart showing the percentages of the high, medium, and low cell-to-population variational TADs. (***H***) Heatmaps visualizing one high and two medium cell-to-population variational TADs. (***I***) Chromosome map showing the genomic locations of high, medium, and low cell-to-population variational TADs. (***J***) Histogram showing the percentages of reorganized TADs in 2,500 pairwise comparisons of 50 cortical L2-5 pyramidal cells and 50 adult astrocytes. (***K***) Stacked bar graph showing the number (the percentage) of the cortical L2-5 pyramidal TADs that are reorganized in a varied number (0 to 50) of adult astrocytes.