nature communications

Article

https://doi.org/10.1038/s41467-026-71406-y

CellLoop: Identifying single-cell 3D genome chromatin loops

Received: 30 June 2025

Accepted: 19 March 2026

Published online: 06 April 2026

Check for updates

Yusen Ye[1] ✉, Yiheng Wang[1], Shiji Liu[1], Yuxuan Hu[1], Liang Yu[1], Weibing Wang[1], Shihua Zhang[2,3,4], Hebing Chen[5] ✉ & Lin Gao[1] ✉

Single-cell 3D genome technologies provide unprecedented views of chromatin architecture, but the extreme sparsity and noise of contact maps limit robust detection of chromatin loops at the individual cell level. Here we present CellLoop, a computational framework for identifying chromatin loops from single-cell contact data by integrating intra-cellular and neighboring inter-cellular contacts through a density-based voting strategy. Applied to Dip-C data from the mouse brain, CellLoop achieves improved loop detection consistent with spatial distances and compartmentalization signals, revealing single cell-specific chromatin loops associated with transcriptional regulation and cell identity. In HiRES embryogenesis data, CellLoop enables finer cell subtype delineation by reducing confounding cell cycle effects. Integration with GAGE-seq and MERFISH data redefines spatial domain functions through chromatin loop dynamics. Together, CellLoop provides a scalable and accurate approach for characterizing chromatin loop variability at single-cell resolution and highlights the utility of 3D genome features in interpreting transcriptional and spatial heterogeneity.

Single-cell 3D genome technologies have rapidly advanced, offering unprecedented opportunities to examine the spatial organization of the genome and its role in gene regulation and cell identity[1–8]. Analytical frameworks have uncovered key structural features, including specific chromatin interactions[3,9–11], topological associating domains (TADs)[12,13], and compartments[14,15], enabling the establishment of critical links between 3D genome architecture and transcriptional control[16].

However, reliable identification of chromatin loops at single-cell resolution remains a major challenge due to the intrinsic sparsity and high dimensionality of contact maps. Existing methods have begun to address this limitation. For example, SnapHiC identifies chromatin loops across small pools of cells[17], and SimpleDiff quantifies interaction strength based on 3D spatial distances reconstructed from individual-cell genomes[3]. Yet SnapHiC relies on small cell populations and cannot capture individual cell-specific features, while SimpleDiff lacks the resolution to distinguish loops from general chromatin interactions. To date, no method has enabled de novo and robust identification of chromatin loops in individual single cells.

Here, we introduce CellLoop, an algorithm for single-cell loop detection based on a density-based center detection framework[18]. CellLoop integrates two complementary signals: (i) intra-cellular topology, which captures the spatial proximity of genomic loci within a single cell, and (ii) inter-cellular background strength, which reflects interaction probabilities across neighboring cells in a defined biological context. The algorithm evaluates each contact in parallel using two simple metrics, enabling efficient loop detection across thousands of single cells.

We benchmarked CellLoop using two orthogonal criteria: genome spatial distance (GSD) and compartment property consistency (CPC)

[1]School of Computer Science and Technology, Xidian University, Xi'an, Shaanxi, China. [2]NCMIS, CEMS, RCSDS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China. [3]School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing, China. [4]Key Laboratory of Systems Health Science of Zhejiang Province, School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Hangzhou, China. [5]Academy of Military Medical Sciences, Beijing, China. ✉e-mail: ysye@xidian.edu.cn; chenhb@bmi.ac.cn; lgao@mail.xidian.edu.cn

between loop anchors, demonstrating superior accuracy in single-cell level. Furthermore, we show that CellLoop can generate a loop frequency map (LFmap) to represent chromatin loop prevalence across cells, demonstrating improved accuracy and biological interpretability over existing approaches. By applying CellLoop to diverse single-cell 3D genomic datasets, including Dip-C data from the mouse brain, HiRES data from embryogenesis, and GAGE-seq and MERFISH data from the mouse cortex, we demonstrate that chromatin loops can serve as informative and biologically meaningful features. CellLoop uncovers cell-specific regulatory interactions, refines cell state annotations, and reveals spatial domain functions shaped by chromatin loops. Together, these findings highlight the utility of chromatin loops as interpretable and biologically meaningful features in single-cell genome architecture analysis.

Results

Overview of CellLoop algorithm

CellLoop identifies chromatin loops in individual cells using a density-based center detection algorithm[18], which integrates intracellular and neighboring inter-cellular contact maps through a re-voting strategy (Fig. 1a). By capturing both local topological associations and contextual support from neighboring cells, CellLoop enables robust identification of chromatin loops in sparse single-cell contact maps. Figure 1b illustrates the analytical workflow and potential applications of CellLoop, including the visualization and comparison of LFmap and CFmap, the identification and visualization of cell-type-specific chromatin loops, the characterization of chromatin loop features that define distinct cell states, and linking cell-type-specific loops with gene expression. Subsequent sections (Figs. 2–4) provide quantitative evaluations of these applications in detail.

CellLoop consists of two primary components (Fig. 1a): 1. Preference prediction module: The input consists of raw contact maps and an initial cell embedding (Methods), These are used to separately construct intra-cellular and inter-cellular networks. The intra-cellular network, where genomic bins are represented as nodes, and edge are weighted by contact frequency between genomic bins, reflects the strength of intrachromosomal topological associations within individual cells. The inter-cellular network, where cells are nodes and edge weights are based on the correlation coefficients of cell embedding vectors, captures the strength of chromatin interactions within a specific biological context. Importantly, this network is constructed using only the nearest neighboring cells to generate a single-cell-specific biological background signal. The combined data from varying support strengths and raw contact maps together inform the preference of chromatin interaction formation in a given cell. Notably, CellLoop incorporates neighborhood information only as a statistical voting reference to stabilize loop detection in sparse single-cell Hi-C data while preserving single-cell specificity. 2. Peaks calling module: Given the cell-specific interaction preference matrix, CellLoop formulates loop identification as a two-dimensional peak-calling problem. A density-based center detection algorithm is then applied to identify chromatin loops as local enrichment centers within the preference map of each cell.

Performance evaluation of CellLoop

We applied CellLoop to Dip-C data generated from the developing mouse cortex and hippocampus at a 20 Kb resolution[2]. CellLoop identified over 260,000 chromatin loops present in at least five single cells, with a median of 1769 chromatin loops per cell (Fig. 2a). Of these loops, more than 98% were detected in fewer than 64 cells, while 1,387 loops were observed in over 128 cells (Fig. 2b), suggesting a broad spectrum of cell specificity. The median number of loops per cell generally decreased with increasing genomic distance, consistent with the biological expectation that long-range loops are less frequent. However, two distance ranges exhibited notable enrichments: 408.5 loops per cell were detected between 40-80 Kb, and 357 loops per cell between 320-640 Kb (Fig. 2c). These enrichments may reflect distinct classes of chromatin loops, potentially corresponding to short-range transcriptional regulatory loops and mid-range structural or architectural loops, respectively. We further analyzed the gene composition of the identified loops and observed that, on average, 42.3% of single-cell chromatin loops contained at least one gene within their anchors (Fig. 2d), whereas only 35.9% of random loops did (P value ≈ 0), supporting the biological relevance of these structures (Supplementary Fig. 1).

To assess CellLoop's performance, we benchmarked it using two orthogonal metrics: GSD and CPC between the anchors of chromatin loops in individual cells (Methods). Biologically meaningful loops are expected to exhibit shorter GSD scores and higher CPC scores relative to control interactions. Consistent with this expectation, CellLoop-identified chromatin loops demonstrated significantly reduced GSD scores relative to controls (Fig. 2e). For example, in a representative cell, the mean GSD of control interactions was 1.6-fold higher than that of CellLoop-identified loops (Fig. 2f). Similarly, loops exhibited significantly elevated CPC scores (Fig. 2g), with a CPC value of 0.807 in the cell (Fig. 2h), indicating high consistency of loop anchors with compartmental identity.

Furthermore, we introduced an additional metric, the CpG density Consistency (CDC) score, to quantify the similarity of CpG density profiles between the two loop anchors. Chromatin loops detected by CellLoop exhibited significantly higher CDC scores than randomized control pairs, indicating that these loops tend to connect genomic regions with coordinated chromatin environments (Supplementary Fig. 2). To further evaluate the biological relevance of CellLoop-identified loops, we analyzed the ChIP-seq enrichment of five representative epigenetic factors, which revealed their preferential localization in regulatory regions, consistent with known biological patterns (Supplementary Fig. 3).

We also evaluated the performance of CellLoop across multiple resolutions. The results show that CellLoop maintains consistent detection accuracy, and the identified loops remain highly concordant across resolutions, demonstrating the robustness of the method (Supplementary Figs. 4 and 5).

In addition, ablation experiments demonstrated that both intra- and inter-cellular signals contribute to loop detection, with the inter-cellular component providing a stronger contribution. These results confirm that integrating both types of signals enhance the robustness of CellLoop (Supplementary Note 1 and Supplementary Fig. 6).

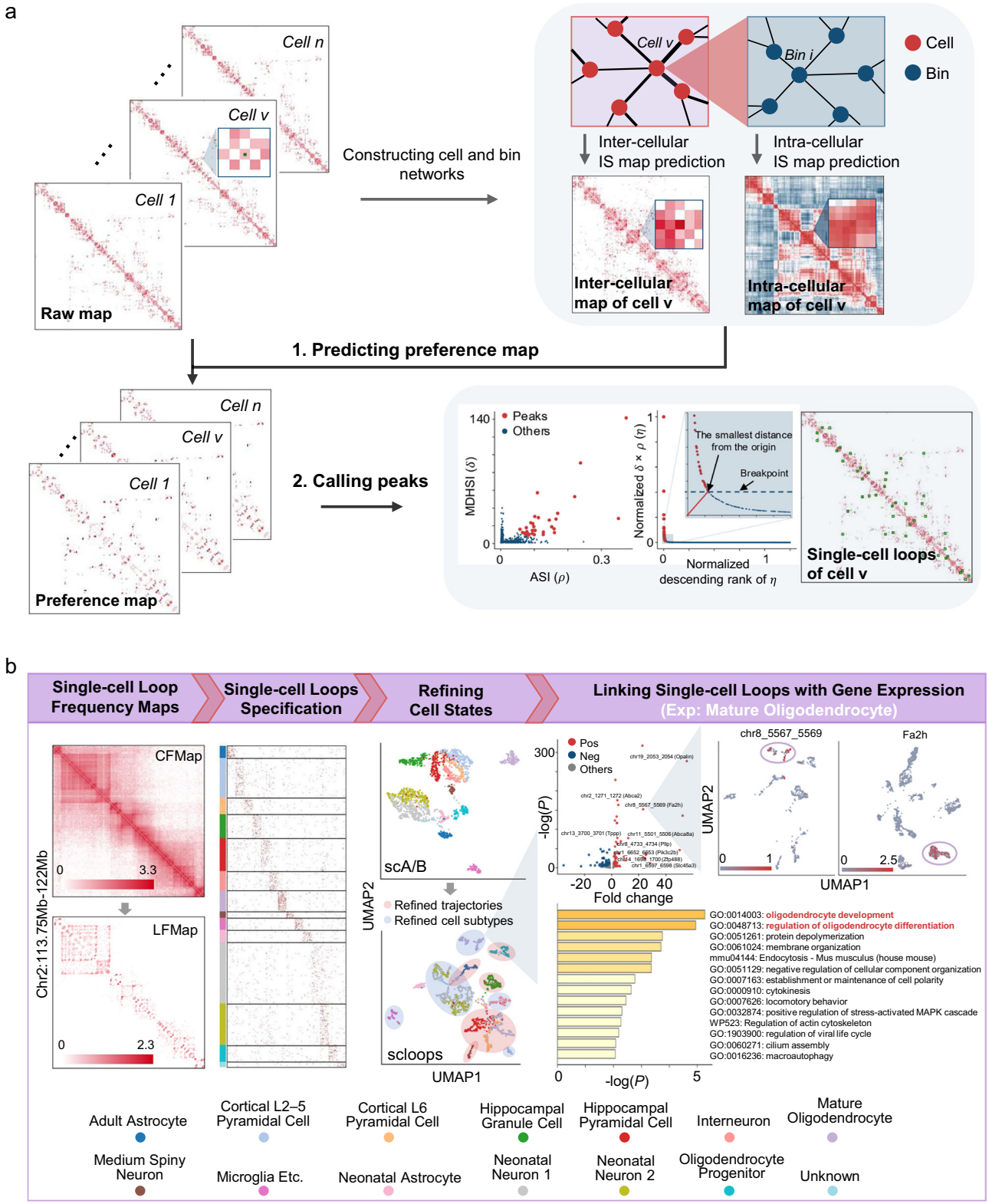
To further verify its generalizability, we applied CellLoop to the independent HiRES single-cell Hi-C dataset. Consistent with the findings from the Dip-C data, CellLoop identified reproducible and biologically meaningful chromatin loops in HiRES (Supplementary Fig. 7).

Finally, we assessed the computational efficiency of CellLoop. When parallelized across 10 CPU cores, CellLoop required an average of 0.99 sec per chromosome at 100 Kb resolution with a peak memory usage of 8.8 GB, and ~13 sec per chromosome at 10 Kb resolution with a peak memory usage of 109.2 GB (Supplementary Figs. 8 and 9, and Supplementary Note 1). The corresponding single-core performance metrics are provided in the Supplementary Information (Supplementary Figs. 8 and 9, and Supplementary Note 1). These results demonstrate the scalability of CellLoop for high-throughput analyses of large single-cell 3D genome datasets.

Performance evaluation of CellLoop via loop frequency map (LFmap)

To further leverage the single-cell loop identified by CellLoop, we constructed LFmap by aggregating chromatin loops across cells in the developing mouse cortex and hippocampus. Compared to

https://doi.org/10.1038/s41467-026-71406-y



conventional contact frequency map (CFmap), LFmap offer several distinct advantages in resolution, interpretability, and biological relevance: 1. Reduced genomic distance bias: LFmap substantially reduce the confounding effect of genomic distance, a major systematic bias in CFmap, as previously reported[19]. Across multiple chromosomes, the correlation coefficient between contact frequency and genomic distance was 0.77-0.84 in CFmap, whereas LFmap showed markedly reduced correlations (0.28-0.35) (Fig. 3a), with a representative case (Fig. 3b). Importantly, this reduction reflects the correction of background bias due to global contact decay, rather than the loss of biologically meaningful distance-dependent patterns. 2. Enhanced detection sensitivity: LFmap exhibit superior sensitivity in detecting chromatin loops. Using three most popular methods, including a commonly used loop detection method for bulk 3D genome map, HiCCUPs[20], a loop detection method for single-cell 3D maps, SnapHiC[17], and a supervised learning framework for chromatin loop

https://doi.org/10.1038/s41467-026-71406-y

Fig. 1 | Overview of the CellLoop Algorithm. a Schematic diagram of the CellLoop algorithm. CellLoop comprises two main components: Preference prediction module and Peaks Calling module. Module 1: CellLoop initiates the process by constructing cell and bin networks using raw contact maps and initial cell embeddings. It then generates interaction strength (IS) maps at both inter-cellular and intra-cellular levels. Based on these maps, CellLoop predicts a preference map that governs chromatin interaction formation within a specific single cell. Module 2: For each candidate chromatin interaction, CellLoop defines two metrics: the aggregated strength of chromatin interactions (ASI, ρ) and the minimum distance of higher-strength chromatin interactions (MDHSI, δ). CellLoop ranks the overall chromatin interaction score ($\eta = \rho \times \delta$) in descending order to infer breakpoints and differentiate chromatin loops from candidate interactions. **b** The significance of CellLoop in the Dip-C dataset is highlighted in the following aspects. The first column, the Chromatin Loop Frequency Map (LFmap), shows greater interpretive power compared to the Chromatin Contact Frequency Map (CFmap). The second column demonstrates how CellLoop detects cell-type-specific chromatin loop features. The third column illustrates that single-cell chromatin loop features (scloops) more effectively define cell states compared to other single-cell 3D features. The fourth column shows how CellLoop links cell-type-specific chromatin loops with gene expression in mature oligodendrocytes. Enrichment analysis was performed using a hypergeometric test, and *P* values (*P*) were adjusted for multiple comparisons using the Benjamini-Hochberg method to control the false discovery rate (FDR). Fold change was calculated as the ratio of the proportion of genes associated with a GO term in the target gene set to that in the background gene set. log represents the base-10 logarithm of the P value and fold change.

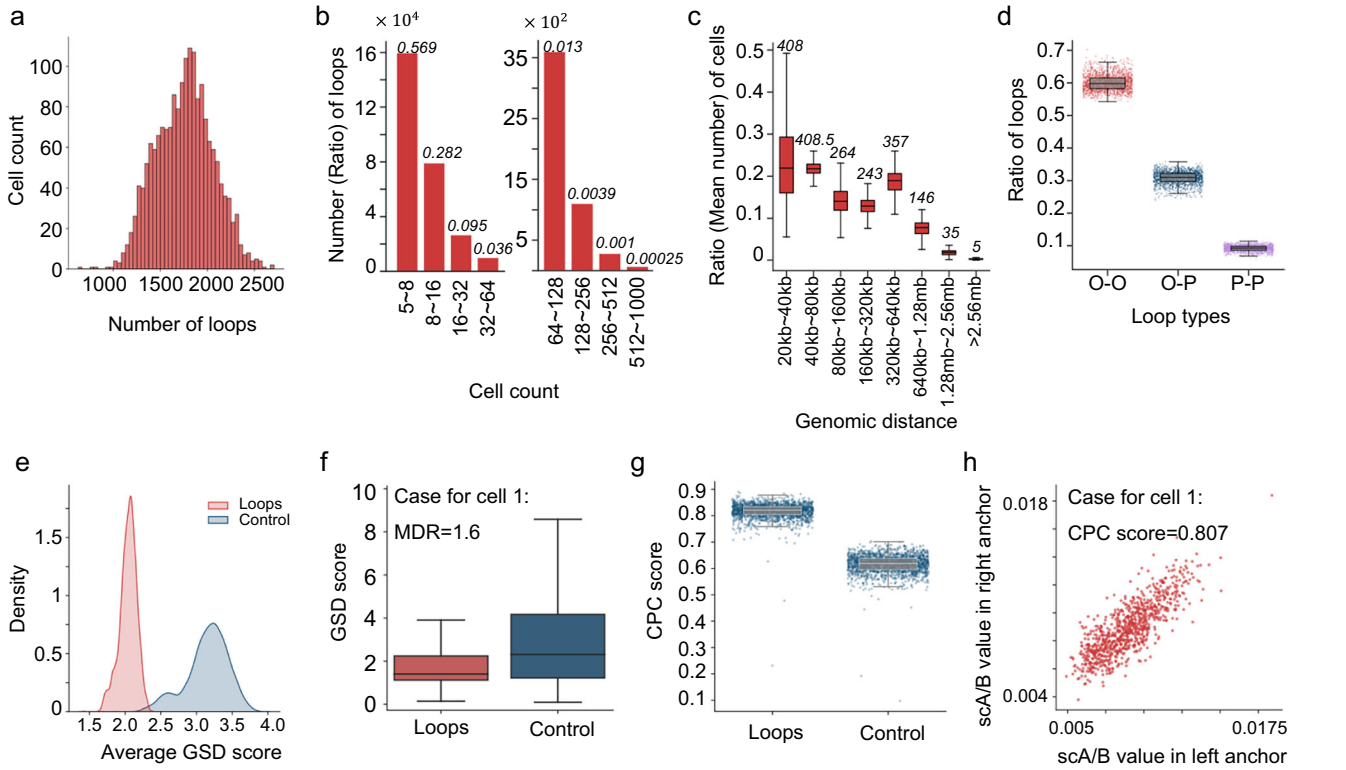
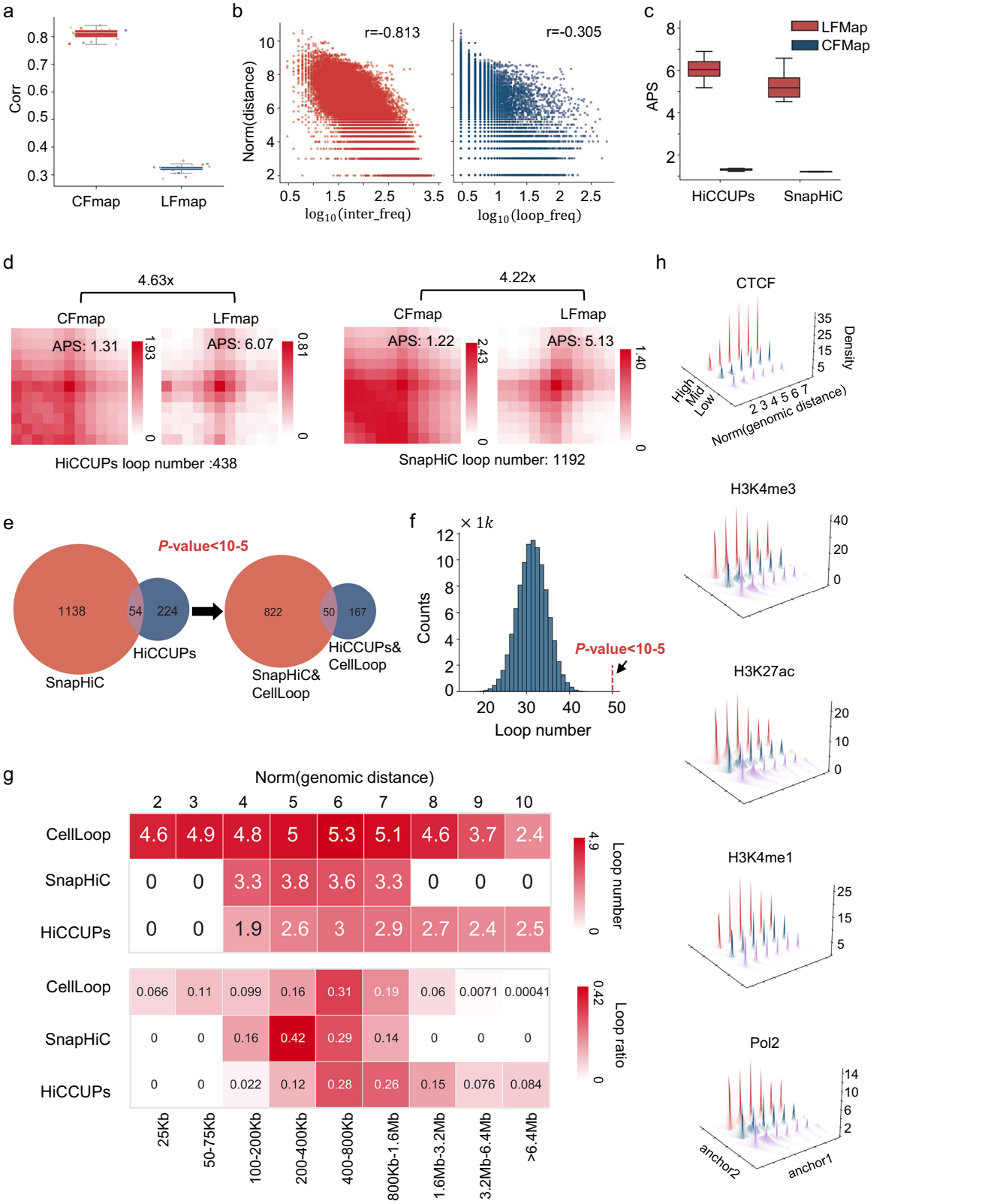


Fig. 2 | Performance evaluation of CellLoop algorithm in the Dip-C dataset at 20 Kb resolution. a Distribution of chromatin loop numbers across different single cells. **b** The number (and ratio) of chromatin loops detected in varying numbers of cells. For example, nearly 160,000 chromatin loops (accounting for 0.569) were detected simultaneously in 5–8 cells (Specifically, 5 ~ 8 indicates loops identified in more than 5 and up to 8 cells). **c** The mean number and ratio distribution of cells identifying each chromatin loop within different genomic distance intervals. Numbers above bars indicate average cell counts. **d** Proportion of different types of chromatin loops in single cells. P indicates that the chromatin loop contains an anchor corresponding to a promoter's transcription start site (TSS), whereas O denotes the opposite. Accordingly, chromatin loops can be categorized into three types: O-O, O-P, and P-P. **e** Density distribution of the average GSD scores for chromatin loops and their control interactions in single cells. **f** GSD scores of chromatin loops and their control interactions in cell 1. **g** Distribution of CPC scores for chromatin loops and their control interactions in single cells. **h** Distribution of scA/B values for the two anchors of chromatin loops and CPC scores for cell 1. In **c**, **d**, **f**, and **g**, the number (n) of samples is presented in Source Data files and box plots show the median (centre line), the interquartile range (box), and whiskers extend to 1.5×IQR.

detection, Peakachu[10], as references, we benchmarked loop signal enrichment. For chromosome 2, LFmap achieved aggregated peak scores (APS) over fourfold higher than CFmap for loops separately identified by all methods (Fig. 3c, d and Supplementary Fig. 10). Overall, LFmap recovered 52,000+ loops observed in at least three cells, covering 92.7% of the genome, whereas 1192 loops (11.4% of regions) by SnapHiC and 278 loops (5.3% of regions) by HiCCUPs (Supplementary Fig. 11a and b). 3. Comprehensive integration: LFmap enables robust integration across loop-calling methods[21]. Only 54 chromatin loops are commonly identified by both HiCCUPs and SnapHiC. Remarkably, LFmap encompasses 50 of them with high statistical enrichment (P value $< 10^{-5}$ in Fig. 3e and f and Supplementary Note 2). A similar trend was observed, with even higher statistical significance, for the common chromatin loops identified by SnapHiC and Peakachu (Supplementary Fig. 10). On average, LFmap covers 88.4% of loops detected by both methods, 69.5% of those identified solely by SnapHiC, and 75.0% by HiCCUPs alone across all chromosomes (P value approaching 0 from the hypergeometric test; Supplementary Fig. 11c and Supplementary Note 2). Notably, loops overlapping with LFmap exhibit significantly higher interaction frequencies than non-overlapping ones (Supplementary Figs. 10 and 11). 4. Extended genomic distance coverage: LFmap identifies chromatin loops across a broader range of genomic distances, particularly those less than 100 Kb (Fig. 3g, Supplementary Figs. 10 and 11e). High

https://doi.org/10.1038/s41467-026-71406-y



background noise often masks genuine short-range loops, leading them to be missed by previous methods. 5. Enrichment in regulatory elements: After excluding loops recognized by SnapHiC or HiCCUPs, the remaining LFmap loops are enriched for transcription and regulatory factors such as CTCF, H3K4me1, H3K4me3, H3K27ac, and Pol II (Fig. 3h and Methods). Loops enriched with CTCF predominantly span 400 Kb to 1.6 Mb, aligning with known CTCF/cohesion-linked loops[22], while those associated with enhancer or promoter markers are mainly within 100–200 Kb, consistent with transcription-linked loops[23].

In summary, LFmap enables biologically meaningful aggregation of single-cell chromatin loop data, outperforming previous methods in sensitivity, integration, and interpretability, and offering a powerful framework for exploring chromatin architecture.

Fig. 3 | Performance evaluation of CellLoop by LFmap in the Dip-C dataset. **a** Comparison of correlation coefficients between interaction frequency and genomic distance on the CFmap and LFmap across different chromosomes. **b** Left: Scatter plot of interaction frequency in the CFmap (inter_freq) versus genomic distance for interactions on chromosome 2. Right: Scatter plot of interaction frequency in the LFmap (loop_freq) versus genomic distance for interactions on chromosome 2 (see Supplementary Methods for Norm(distance)). **c** Distribution of the aggregate peak score (APS) for chromatin loops identified by HiCCUPs and SnapHiC on the CFmap and LFmap, respectively, across all chromosomes. **d** Comparison of APS for chromatin loops identified by HiCCUPs and SnapHiC on the CFmap and LFmap for chromosome 2. **e** Left: Overlap of chromatin loops identified by SnapHi-C and HiCCUPs. Right: Overlap of chromatin loops detected by SnapHi-C & CellLoop and HiCCUPs&CellLoop for chromosome 2 (HiCCUPs&-CellLoop represents chromatin loops identified by both methods simultaneously). **f** Expected distribution of chromatin loop numbers in the LFmap. These loops are detected simultaneously by SnapHi-C and HiCCUPs. Statistical significance (*P* value) was determined by comparing the observed overlap count to the overlap counts generated from 100,000 random repetitions. In e and f, A hypergeometric test was conducted, with a resulting *P* value of 0. **g** Distribution of chromatin loop numbers (and their ratios) at different genomic distances for three methods (see Supplementary Methods for Norm(genomic distance)). **h** Enrichment distribution of ChIP-seq peaks around chromatin loops present in the LFmap but not identified by either SnapHiC or HiCCUPs. Chromatin loops were divided into three groups according to their loop frequency values: Low (bottom 25%), Mid (middle 50%), and High (top 25%). In **a** and **c** the number (n) of samples is presented in Source Data files, and box plots show the median (centre line), the interquartile range (box), and whiskers extend to 1.5×IQR.

Linking cell type-specific chromatin loops with gene regulation

To explore the functional relevance of chromatin loops at the single-cell level, we used CellLoop-identified loops as features for dimensionality reduction. For comparison, we also derived features from single-cell A/B compartment values (scA/B), which were obtained following the established CpG density-based approach and widely applied in recent single-cell Hi-C studies[3,5]. While scA/B compartments remain the most commonly used structural features for cell type classification, they primarily capture large-scale nuclear partitioning. By contrast, chromatin loop-based features yielded clearer separation of cell types and finer discrimination of subpopulations in UMAP embeddings (Fig. 4a), highlighting the complementary value of loop-derived features as cell-type-specific structural markers that provide finer-scale regulatory information.

We developed a computational framework to classify CellLoop-detected chromatin loops into four categories (Supplementary Fig. 12 and Methods): 12,465 individual cell type-specific chromatin loops (Indi-Spe), 824 multiple cell type-specific chromatin loops (Multi-Spe), 1396 common chromatin loops (Com), and 29,481 other chromatin loops (Others) (Supplementary Fig. 13a). Representative examples of each category are shown in Supplementary Fig. 13b–d. It is worth noting that the "Others" category comprises a heterogeneous set of bona fide chromatin loops that were not further subclassified, rather than false positives (Supplementary Fig. 14). Notably, Indi-Spe and Multi-Spe loops tended to span shorter genomic distances (median: 60 Kb) compared to Com loops (median: 100 Kb; Supplementary Fig. 13e), and appeared in fewer cells, reflecting their subtype specificity (Supplementary Fig. 13a). Across 13 annotated cell types, we identified between 487 and 1546 Indi-Spe chromatin loops per type, as well as 36 Indi-Spe loops associated with unclassified cell states in the Dip-C dataset[2] (Fig. 4b and Supplementary Fig. 13f). Interestingly, non-neuronal cell types exhibited a greater significance than neuronal types (Supplementary Fig. 13f), highlighting structural heterogeneity beyond neuronal lineages.

To connect loops with gene regulation, we integrated chromatin loop annotations with cell type-specific transcriptomic profiles from matched Dip-C and MALBAC-DT[24] datasets (Supplementary Fig. 15a and b). For each cell type, we identified Indi-Spe-gene chromatin loops, Indi-Spe loops, containing at least one gene in either anchor. Among these, 108-363 loops per cell type were associated with genes exhibiting specific expression patterns in the same cell type (Supplementary Fig. 13g). These were further stratified into Indi-Spe-highgene and Indi-Spe-lowgene groups based on expression levels. The top 50 Indi-Spe-highgene loops per cell type are provided in Supplementary Fig. 12 (see Methods). Notably, both loop presence and gene expression exhibited marked cell type specificity (Fig. 4c). For example, a loop at Chr2:6475_6479 linked to *Sirpa* was enriched in microglia (Supplementary Fig. 15c and d), while a loop at Chr8:5567_5569 was specifically associated with the *Fa2h* gene in mature oligodendrocytes (Supplementary Fig. 15e and f).

Gene Ontology enrichment analysis of the Indi-Spe-highgene and Indi-Spe-lowgene categories further supported their biological relevance. Indi-Spe-highgene-associated genes were enriched in processes aligned with cell identity, such as "innate immune response" in microglia and "oligodendrocyte development" in mature oligodendrocytes. (Fig. 4d, e; Supplementary Figs. 15g, h, 16, and 17).

Together, these results demonstrate that CellLoop enables the identification of functionally relevant chromatin loops with cell-type specificity and provides a structural framework for linking 3D genome architecture to gene regulation.

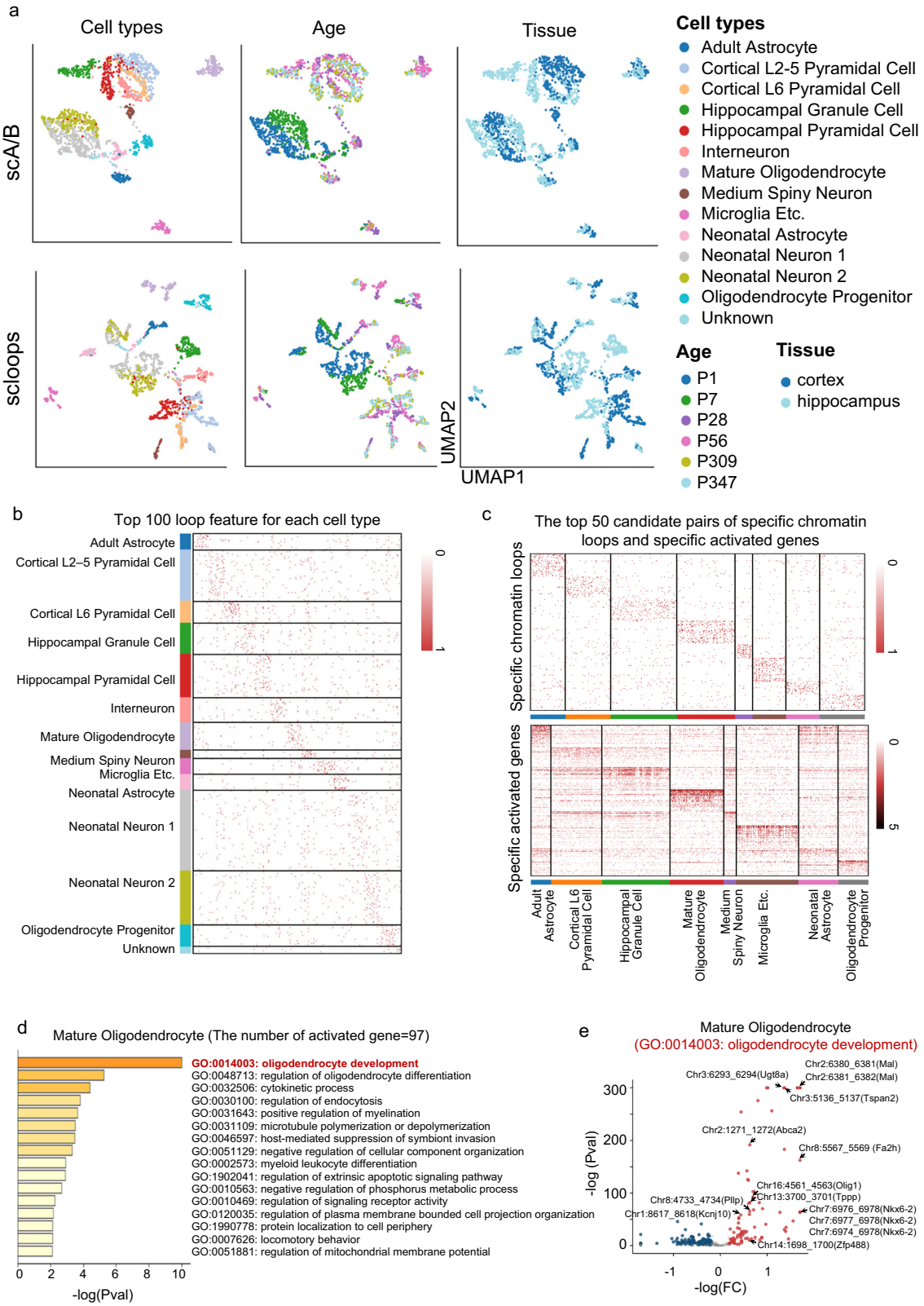
Single-cell chromatin loop signatures refine cell subtypes

To further assess CellLoop's capability in refining cell states, we applied it to dual-modality data from mouse embryonic development obtained using HiRES technology at 100 Kb resolution. In a previous study, these single cells were well-annotated based on RNA sequencing data using specifically expressed marker genes. Compared with Higashi, which infers chromatin architecture from single-modality 3D genome maps, CellLoop delineated cell types while attenuating confounding effects from cell cycle variation (Fig. 5a). Moreover, in comparison to single-cell gene expression features, chromatin loop features uncovered finer cell subclusters within annotated types, revealing subtle subtype-level distinctions (Supplementary Figs. 18a and 19a–c).

Within the early neuronal lineage, CellLoop identified two previously uncharacterized subpopulations, designated as subtypes 9 and 28, based on UMAP clustering derived from chromatin loop profiles (Fig. 5c). Differential loop analysis revealed 223 and 319 subtype-specific chromatin loops for subtypes 9 and 28, respectively (Supplementary Figs. 20–22). For instance, subtype 28 harbored a prominent chromatin loop (Chr2:1690_1700), whereas the same genomic region in subtype 9 exhibited a stripe-like architecture (Chr2:1690_1697/1698/1699/1700), consistent with progressive loop extrusion mediated by CTCF and cohesin[25] (Fig. 5d). Similarly, a specific loop (Chr18:763_768) was observed in subtype 9, whereas a neighboring loop (Chr18:763_769) was shared between both subtypes (Supplementary Fig. 23a), reflecting dynamic chromatin remodeling.

To investigate the regulatory potential of these loops, we identified 33 and 40 loop-gene pairs uniquely associated with subtypes 9 and 28, respectively (Supplementary Figs. 24–26). Gene ontology enrichment analysis revealed distinct functional themes: genes linked to subtype 9 loops were enriched in processes such as cell adhesion and axon guidance (Fig. 5c, top), while genes associated with subtype 28 were enriched in terms related to organ development, tissue morphogenesis, and regulation of developmental processes (Fig. 5c, bottom). These loop-gene pairs frequently connected gene promoters with distal regulatory elements, suggesting structural underpinnings for transcriptional

https://doi.org/10.1038/s41467-026-71406-y



regulation. One illustrative example is a loop in subtype 28 (Chr17:478_480), which potentially brings a super-enhancer into spatial proximity with the *Foxp4* promoter (Fig. 5e). Foxp4 is known to regulate embryonic development in multiple organs, including the lung, intestine, and nervous system[26]. In subtype 9, the chromatin loop Chr16:423_427 appears to link a distal enhancer to the *Gap43* gene (Supplementary Fig. 23b), a gene implicated in axonal growth, synaptic plasticity, and neuronal migration during development[27].

Collectively, these findings demonstrate that single-cell chromatin loop profiles can uncover previously unrecognized cell subtypes and reveal structural bases for transcriptional regulation. CellLoop thus provides a high-resolution framework for dissecting fine-grained cellular diversity during development.

Fig. 4 | Application of CellLoop to the Dip-C dataset for linking cell type-specific chromatin loops with gene functions at 100 Kb resolution. a UMAP visualization of cells using scA/B values and single-cell chromatin loop (scloops) features at 100 KB resolution. Cells are labeled by cell type, age, and tissue from left to right. **b** The top 100 most significant chromatin loops for each cell type, excluding the "Unknown" category. **c** The top 50 Indi-Spe-highgene pairs for each cell type. **d** Gene function enrichment analysis for gene sets from the Indi-Spe-highgene list of the Mature Oligodendrocyte cell type (see Supplementary Methods). **e** Volcano plot of Indi-Spe-highgene pairs for GO:0014003 term in Mature Oligodendrocyte cell type. In **d** and **e**, enrichment analysis was performed using a hypergeometric test, and *P* values (Pval) were adjusted for multiple comparisons using the Benjamini-Hochberg method to control the false discovery rate (FDR). Fold change (FC) was calculated as the ratio of the proportion of genes associated with a GO term in the target gene set to that in the background gene set. log represents the base-10 logarithm of the *P* value and fold change.

Single-cell chromatin loop signatures enhance spatial domain analysis

Recent advancements have enabled simultaneous mapping of single-cell 3D genomes and transcriptomes in the mouse cortex using GAGE-seq[6]. Additionally, MERFISH (multiplexed error-robust fluorescence in situ hybridization) has identified transcriptomes and spatial locations of single cells in the mouse primary motor cortex[28]. To evaluate the utility of CellLoop in delineating spatial domains (SDs) from the perspective of chromatin architecture, we applied it to the GAGE-seq dataset to identify single-cell chromatin loops at 100 kb resolution. Using the transcriptomic modality as an anchor, we inferred spatial coordinates for each cell via Tangram[29], thereby enabling the spatial localization of chromatin loop profiles (Methods).

We next employed STAGATE[30] to identify SDs using single-cell chromatin loop features (Fig. 6a). The resulting SDs exhibited strong concordance with major annotated cell types (Supplementary Fig. 27a), suggesting that spatial chromatin loop patterns are tightly coupled with cell-type identity. We focused on the fifth SD, which exhibited distinct transcriptional and structural characteristics, and stratified associated loop-linked genes into high- and low-expression cohorts. Within this SD, 2120 chromatin loops corresponded to significantly altered gene expression (one-sided *t*-test: $P = 1.22 \times 10^{-8}$ for high expression and $P = 2.5 \times 10^{-7}$ for low expression, Fig. 6b). A parallel analysis of the second SD further corroborated this association (one-sided *t*-test: $P = 1.88 \times 10^{-24}$ for high expression and $P = 6.85 \times 10^{-6}$ for low expression, Supplementary Fig. 27b).

Comparative analysis of SD5 and SD2 identified 2006 SD5-specific chromatin loops (one-sided *t*-test, $P = 1.16 \times 10^{-89}$; Fig. 6c and Supplementary Fig. 27c). Among these, 42 loop-gene pairs exhibited significant correlations between loop presence and expression specificity (one-sided *t*-test for loops: $P = 4.39 \times 10^{-6}$; for genes: $P = 2.82 \times 10^{-31}$; Fig. 6d), collectively termed chromatin loop-driven genes (CL-genes). Compared to those in SD2, CL-genes within SD5 displayed broader and more diverse functional enrichment (Fig. 6e, Supplementary Fig. 27d, Supplementary Data 1 and 2), underscoring the spatial specificity of chromatin architecture in shaping gene regulatory landscapes.

To further dissect functional divergence, we identified a control set of 29 loop-independent genes within the fifth SD that were spatially specific but not regulated by chromatin loops (termed normal-genes). Gene Ontology enrichment revealed that CL-genes were predominantly associated with neurodevelopmental processes such as axonogenesis, cell migration, and proliferation. In contrast, normal-genes were enriched for synaptic functions, including glutamatergic signaling, neurotransmitter transport, and synaptic vesicle exocytosis (Fig. 6e, f; and Supplementary Data 1 and 3). These findings suggest that CL-genes possess distinct regulatory and functional signatures relative to spatially specific but loop-independent genes.

At the level of individual loci, we observed that the *Mef2c* gene, previously implicated in autism spectrum disorder (ASD) pathogenesis, was regulated by a distal enhancer located at 83.1–83.2 Mb, marked by enriched enhancer signals and associated with elevated *Mef2c* expression (Fig. 6g). Notably, base-editing of *Mef2c* regulatory regions has been shown to alleviate ASD-like phenotypes in mouse models[31]. Additional examples include *Hbegf*, a gene facilitating tumor cell infiltration across the blood-brain barrier[32], and *EphA4*, implicated in amyotrophic lateral sclerosis and other neurodegenerative diseases[24,33], both of which appear to be regulated by long-range chromatin loops connecting promoters to distal enhancers (Supplementary Fig. 27e–h).

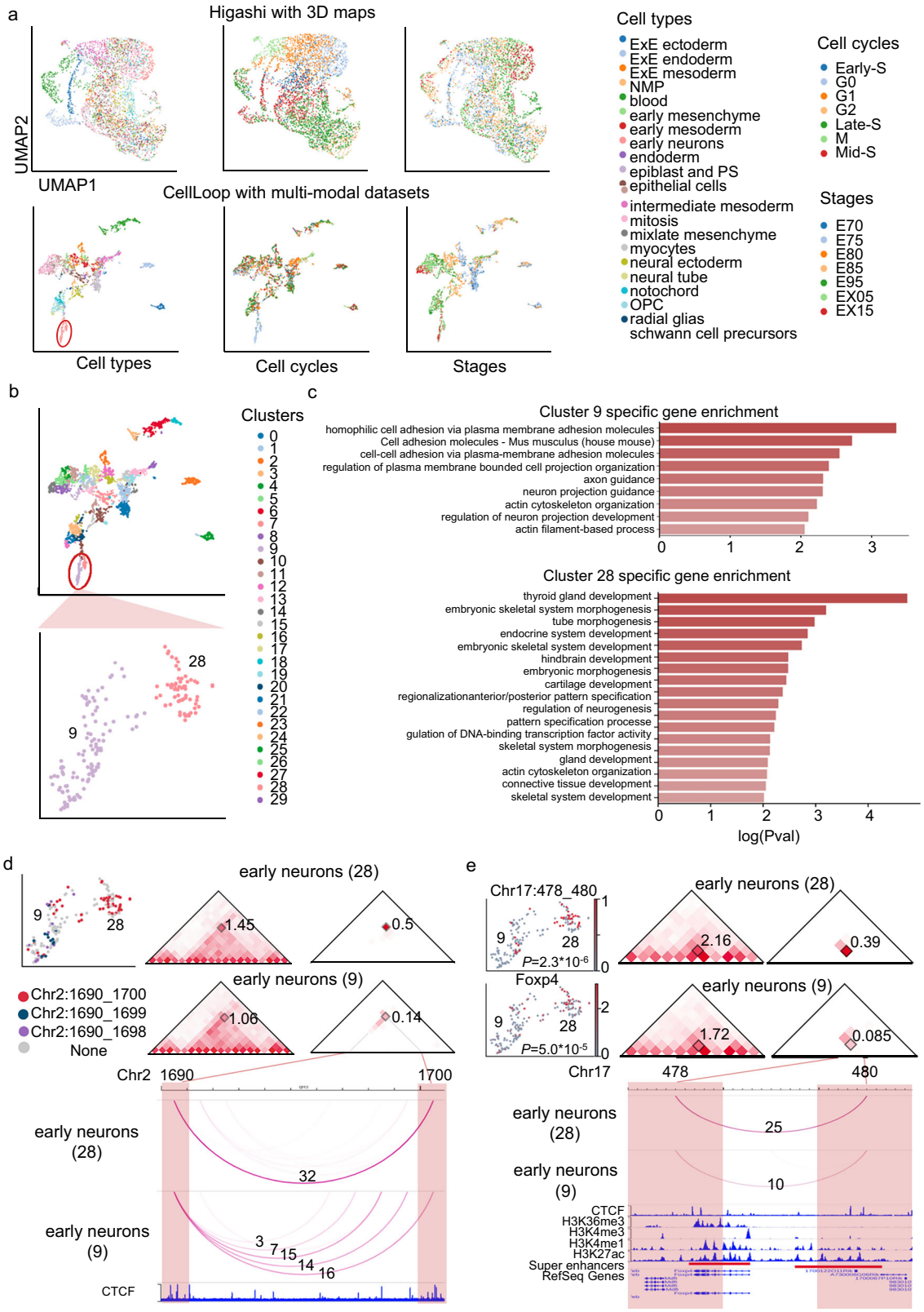
Collectively, our results demonstrate that CellLoop enables the mapping of spatially resolved long-range regulatory interactions, offering mechanistic insights into how 3D genome architecture contributes to spatial gene regulation and functional zonation in the brain.

Discussion

CellLoop identifies chromatin loops at single-cell resolution by formulating the problem as a density-based peak detection task, accounting for the multi-level background strength inherent in chromatin interaction data[18]. Leveraging two simple yet complementary metrics, CellLoop enables efficient and scalable detection of chromatin loops across tens of thousands of individual cells. The resulting LFmap enhances biological interpretability compared to the raw CFmap, offering a clearer window into the functional architecture of the 3D genome. To elucidate the biological relevance of single-cell chromatin loop signatures, we established an analytical framework that connects loop features to gene expression, cell state, and spatial organization. Application of CellLoop across diverse single-cell 3D genome datasets revealed distinct chromatin loop signatures that uncover subtle biological heterogeneity among cells. These signatures facilitated refined cell-type classification, resolution of finer subpopulations, and spatial domain interpretation, demonstrating the potential of 3D chromatin loops to serve as robust, orthogonal markers for cell identity and function.

A key advantage of CellLoop lies in its unsupervised nature, enabling de novo discovery of chromatin loops without prior annotations or reliance on specific interaction patterns. Nonetheless, several challenges and opportunities remain. First, while CellLoop combines intra-cellular topological density and inter-cellular background support to enhance loop detection robustness, its performance remains dependent on data quality. In particular, extremely sparse single-cell Hi-C contact maps may lack sufficient signal even with neighborhood re-voting, which limits reliable loop identification. Second, current single-cell 3D genomics data often suffer from sparsity and noise. Integrating multi-modal data, such as transcriptomics, epigenomics, and imaging, could help reconstruct more realistic nuclear topologies and facilitate the identification of complex, higher-order (multi-way) chromatin structures that extend beyond pairwise loops. Third, the interpretation of single-cell chromatin loop signatures remains heavily dependent on cellular context. Future efforts should focus on integrative computational frameworks that incorporate multi-omics, spatial transcriptomics, and live-cell imaging to better capture to elucidate intercellular associations and establish a robust biological context for single cells[34,35]. As high-resolution spatial 3D genome technologies continue to evolve[36], mapping spatially constrained chromatin loops will be instrumental in linking chromatin organization to tissue architecture, intercellular communication, and functional zonation in complex organs such as the brain. Lastly, understanding how alterations in chromatin loop patterns contribute to development and disease represents a major frontier. Integrating CellLoop'results with datasets capturing DNA methylation, histone modifications, and chromatin accessibility will help elucidate the

https://doi.org/10.1038/s41467-026-71406-y



relationship between multi-scale 3D genome topology and gene regulatory mechanisms[37]. This will be particularly valuable in disease contexts, where aberrant chromatin looping may underlie dysregulated transcriptional programs.

In summary, with the rapidly increasing volume of single-cell 3D genomic data, CellLoop is poised to play a pivotal role in rapidly and accurately identifying chromatin loops at single-cell resolution in the context of diseases or specific biological states. By uncovering unique structural signatures, CellLoop facilitates refined cell annotation, reveals latent regulatory relationships, and opens new avenues for dissecting gene regulation and cellular function from a 3D genome perspective.

Fig. 5 | Application of CellLoop to the HiRES Dataset for Refining Cell Subtypes. **a** UMAP visualization of cells using cell embeddings from the Higashi method (top) and single-cell chromatin loop features (bottom). Cells are labeled by cell type, cell cycle, and cell stage from left to right. **b** Upper: UMAP visualization of cells using single-cell chromatin loop features, with cells labeled by clustering via the Leiden algorithm (resolution = 5). Bottom: UMAP visualization of clusters 9 and 28. **c** Gene enrichment analysis for two gene sets associated with loop-gene pairs in clusters 9 and 28. These two cell clusters serve as reference sets for identifying loop-gene pairs. Enrichment analysis was performed using a hypergeometric test, and *P* values (Pval) were adjusted for multiple comparisons using the Benjamini-Hochberg method to control the false discovery rate (FDR). log represents the base-10 logarithm of the *P* value and fold change. **d** Example of a specific chromatin loop (Chr2:1690_1700) in cluster 28. Upper left: UMAP visualization of the Chr2:1690_1698/1699/1700 chromatin loop across cells. Upper middle and right: CFmap and LFmap surrounding the specific chromatin loop in the two clusters. Bottom: WashU Epigenome Browser showing the CTCF signal and LFmap for both clusters around the chromatin loop. Small squares mark the chromatin loops of interest. **e** Example of a loop-gene pair in cluster 28. Upper left: UMAP visualization of the chromatin loop and corresponding gene expression. Upper middle and right: CFmap and LFmap surrounding the chromatin loop in the two clusters. Bottom: WashU Epigenome Browser displaying epigenetic signals and LFmap around the chromatin loop. Statistical significance was assessed using a one-sided fisher's exact test.

Methods

Identification of single-cell chromatin loops

The CellLoop algorithm consists of two primary components: the preference prediction module and the peak calling module (Fig. 1a). The inputs to the algorithm are raw single-cell contact maps and initial embedding matrices (Methods). To demonstrate its workflow, we focus on intra-chromatin interactions and illustrate the algorithm using a representative chromatin region from single cell v.

In the first component, we first construct two types of graphs for each single cell v: 1. Intra-cellular graph: we constructed an undirected graph $A^v \in \{0,1\}^{n\times n}$ with bins as nodes, where each node corresponds to genomic bin. An edge is defined between bins i and j if the raw contact map of cell v records an interaction ($A^v_{ij}=1$); otherwise, $A^v_{ij}=0$. (Fig. 1a) 2. Inter-cellular graph: we then construct a weighted undirected k-nearest neighbors (k-NNs) graph $W^v \in R^{m\times m}$, where nodes represent single cells. cell v is connected to its k-NNs based on Euclidean distances in the initial embedding space (Fig. 1a). Using these graphs, we compute the enhanced contact map for cell v:

$$S^v = \frac{\sum_{u\in N(v)} w_{uv} A^u}{|N(v)|} \tag{1}$$

where $N(v)$ denotes the set of neighboring cells of v with edge weights $w_{uv} > 0.2$. Details on the default choice of k and its effect on algorithm robustness are provided in the Supplementary Methods.

Next, we predicted the preference for chromatin interaction formation ρ by integrating interaction strengths from multiple levels (Fig. 1a, bottom), defined as:

$$\rho^v(i,j) = \mathrm{IS}^v_{\mathrm{inter}}(i,j) \times \mathrm{IS}^v_{\mathrm{intra}}(i,j) \times A^v_{ij} \tag{2}$$

where $\mathrm{IS}^v_{\mathrm{inter}}(i,j)$ and $\mathrm{IS}^v_{\mathrm{intra}}(i,j)$ represent inter-cellular and intra-cellular interaction strengths between bins i and j, respectively. $\mathrm{IS}^v_{\mathrm{inter}}(i,j)$ is defined as:

$$\mathrm{IS}^v_{\mathrm{inter}}(i,j) = f\left(\mathrm{LD}^v(i,j)\right) * f(\mathrm{SNR}^v(i,j)) \tag{3}$$

where $\mathrm{LD}^v(i,j)$ denotes the local interaction density around focal pair(i,j) based on neighboring inter-cellular maps:

$$\mathrm{LD}^v(i,j) = \frac{B^v_{ij}}{(2l_1+1)^2} \tag{4}$$

where $B^v_{ij} = \sum_{i=-l_1}^{l_1}\sum_{j=-l_1}^{l_1} S^v_{ij}$ (default $l_1 = 1$; see Supplementary Fig. 28). The signal-to-noise ratio $\mathrm{SNR}^v(i,j)$ quantifies the relative strength of loop-associated contacts compared to their local background at the single-cell level:

$$\mathrm{SNR}^v(i,j) = \frac{S^v_{ij} \times \left((2l_3+1)^2 - (2l_2+1)^2\right)}{D^v_{ij} - C^v_{ij}} \tag{5}$$

where $C^v_{ij} = \sum_{i=-l_2}^{l_2}\sum_{j=-l_2}^{l_2} S^v_{ij}$ and $D^v_{ij} = \sum_{i=-l_3}^{l_3}\sum_{j=-l_3}^{l_3} S^v_{ij}$ (default $l_2 = 2, l_3 = 5$; see Supplementary Fig. 28). To enhance contrast between meaningful high-intensity contacts and weak background signals, a sigmoid-like normalization function was applied: $f(x') = \frac{1}{1+e^{-5x'}}, x' = \frac{x-x_{\min}}{x_{\max}-x_{\min}}$. The intra-cellular interaction strength $\mathrm{IS}^v_{\mathrm{intra}}(i,j)$ is defined as:

$$\mathrm{IS}^v_{\mathrm{intra}}(i,j) = \frac{h^v_i \cdot h^v_j}{||h^v_i||\,||h^v_j||} \tag{6}$$

where h^v_i represents the embedding vector of bin i, obtained using the node2vec graph representation learning algorithm applied to the undirected graph A^v based on a biased random walk procedure[38]. This embedding captures the local and global topological properties of the chromatin contact network within each cell.

The second component of the CellLoop algorithm distinguishes chromatin loops from the preference map ρ by formulating the task as a two-dimensional peak calling problem, with its core concept adapted and refined from a fast peak-calling method[18]. We retained chromatin interactions with $\rho > 0$ as candidates for peak calling. For each candidate interaction, we further defined a metric, the minimum distance $\delta^v(i,j)$ to higher-strength chromatin interactions within a range of $\mathrm{Dis}_{\max}$ (default $\mathrm{Dis}_{\max} = 100$):

$$\delta^v(i,j) = \min_{(i',j'):\rho^v(i',j')>\rho^v(i,j)} \sqrt{(i'-i)^2 + (j'-j)^2} \tag{7}$$

where $\delta^v(i,j) < \mathrm{Dis}_{\max}$. The overall score for chromatin interaction (i,j) is then:

$$\eta^v(i,j) = \rho^v(i,j) * \delta^v(i,j). \tag{8}$$

where the formulation follows the validated framework[18]. Candidate chromatin interactions were ranked in descending order of η^v. Both η^v and their rank r^v are normalized by $\eta^v_{\mathrm{std}}(i,j) = \eta^v(i,j)/\eta^v_{\max}$ and $r^v_{\mathrm{std}}(i,j) = r^v(i,j)/r^v_{\max}$, respectively. To distinguish true chromatin loops from background interactions, we identify the inflection point $\left(i_{\mathrm{bp}}, j_{\mathrm{bp}}\right)$ by minimizing the sum of squares of the normalized scores:

$$\left(i_{\mathrm{bp}}, j_{\mathrm{bp}}\right) = \underset{(i,j):j>i}{\mathrm{argmin}} \left(\eta^v_{\mathrm{std}}(i,j)^2 + r^v_{\mathrm{std}}(i,j)^2\right) \tag{9}$$

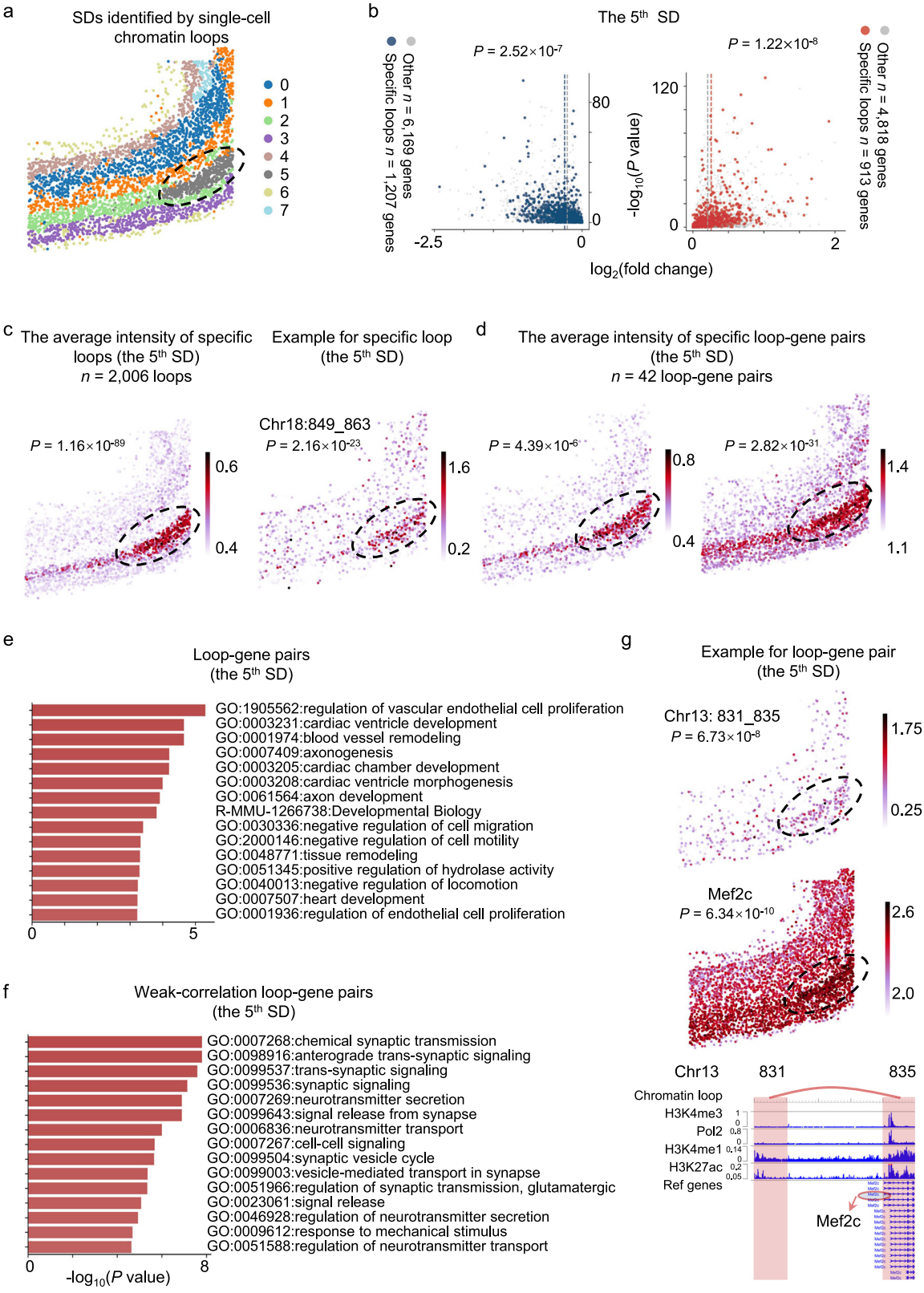
where the core idea here follows the fast peak-calling algorithm[18]. Finally, chromatin loops were defined as the set of interactions whose normalized scores exceeded that of the inflection point:

$$\mathrm{loopsets}^v = \left\{(i,j) | \eta^v_{\mathrm{std}}(i,j) > \eta^v_{\mathrm{std}}\left(i_{\mathrm{bp}}, j_{\mathrm{bp}}\right)\right\}. \tag{10}$$

The pipeline for dividing single-cell chromatin loops into different types

To facilitate integrated analysis across multiple cell types, we developed a computational pipeline to classify detected chromatin loops

https://doi.org/10.1038/s41467-026-71406-y



into distinct subtypes (Supplementary Fig. 12). The pipeline inputs include the single-cell chromatin loops matrix, $M \in \{0,1\}^{n\times m}$ and cell type annotation vector of single cells $V \in \{1,2,\cdots,N\}^n$, where n represents the number of single cells and m denotes the number of chromatin loops. M indicates the presence (1) or absence (0) of specific chromatin loops in individual cells, while V provides the corresponding cell type annotations.

Prior to classification, we applied a quality control filter that retained only chromatin loops detected in at least ten single cells. For a given chromatin loop i and cell type j, we defined the positive sample set (Pos) as the occurrence vector of loop i in cells annotated as type j. The negative sample set (Neg) was generated by the occurrence vector of loop i in cells annotated as other cell types.

Fig. 6 | Application of CellLoop to the GAGE-seq and MERFISH Datasets for Depicting Spatial Domains. a Spatial domains (SDs) identified by single-cell chromatin loops. The 5th SD is highlighted with an ellipse. **b** Left: Volcano plot of differential expression. Blue circles denote low-expressed genes within the anchor regions of chromatin loops specific to the 5th SD, compared to other cells. Grey circles represent other low-expressed genes. Right: Red circles denote high-expressed genes within the anchor regions of chromatin loops specific to the 5th SD, while grey circles represent other high-expressed genes (*P* values derived from one-sided *t*-tests). **c** Left, in situ plot of the average intensity of the 5th SD specific chromatin loops. Right, in situ plot of the example chromatin loop Chr18:849_863. **d** In situ plots of the average intensity of loop-gene pairs specific to the 5th SD (left: chromatin loops; right: genes). **e** Gene enrichment analysis of the gene sets associated with loop-gene pairs in the 5th SD. **f** Gene enrichment analysis of the gene sets associated with weak-correlation loop-gene pairs in the 5th SD. In e and f, enrichment analysis was performed using a hypergeometric test, and *P* values were adjusted for multiple comparisons using the Benjamini-Hochberg method to control the false discovery rate (FDR). **g** Upper and middle, In situ plots of example loop-gene pair. Bottom, WashU Epigenome Browse shows epigenetic signals around the chromatin loop. In **c**, **d**, and **g** statistical significance *P* values (*P*) were assessed using a one-sided *t*-test.

We then calculated the enrichment score (ES) and statistical significance (*P* value) for chromatin loop *i* in cell type *j*. The ES was defined as the ratio of the proportion of loop occurrences in Pos to that in Neg. Statistical significance was assessed using a one-sided fisher's exact test applied to the corresponding contingency table, and *P* values were adjusted for multiple testing using the Benjamini-Hochberg procedure[39].

In a specific case, during the joint analysis of GAGE-seq and MERFISH datasets, chromatin loop strengths were inferred and normalized using the Tangram algorithm, resulting in continuous-valued loop intensity measures. For this analysis, statistical significance was assessed using a one-sided *t*-test, and the resulting *P* values were adjusted for multiple testing using the Benjamini-Hochberg method[39]. Based on these metrics, each chromatin loop was categorized into one of four specificity levels for a given cell type, ranging from highest to lowest specificity: strong specific (strong-loop), weak specific (weak-loop), non-specific (noSpe-loop), and rare (rare-loop) (Supplementary Fig. 12). Here, A rare-loop is defined as a chromatin loop that occurs in fewer than five cells or in less than 2% of the given cell type. Based on this, chromatin loops are defined as noSpe-loops with ES < 1.2, those with ES ≥ 1.2 and a *P* value > 0.0005 are defined as weak-loops, and those with ES ≥ 1.2 and *P* value ≤ 0.0005 are defined as strong-loops (Detailed descriptions of this procedure are provided in Supplementary Fig. 12).

Furthermore, based on the overall distribution of these specificity levels across all cell types, each chromatin loop was further categorized into one of three broader classes: specific to an individual cell type (Indi-Spe), specific to multiple cell types (Multi-Spe), or common (Com) (Supplementary Fig. 12). Among Indi-Spe loops, we further subcategorized them by examining the expression status of genes located at their loop anchors relative to other cell types, defining three subclasses: gene specifically highly expressed (Indi-Spe-highgene), gene specifically lowly expressed (Indi-Spe-lowgene), and gene non-specifically expressed (Indi-Spe-nonSpegene). The most specific gene within each loop anchor was used for downstream analysis.

For analyses targeting specific cell types or subtypes, we employed the Scanpy package to identify specific chromatin loops and genes by specifying the reference cell set for comparison using a one-sided *t*-test[40].

The definition of loop-gene pairs

In single-modality analyses, we define a loop-gene pair as a chromatin loop that exhibits specificity and encompasses within its anchor regions a gene that is highly and specifically expressed (selecting the most specific gene). For dual-modality data, an additional criterion is applied: the Pearson correlation coefficient between the single-cell expression level of the gene and the normalized occurrence score of the chromatin loop, specific to the pertinent cell type or subtype, must exceed 0.1. Pairs exhibiting a correlation coefficient of 0.1 or below are categorized as weak-correlation loop-gene pairs.

The initial embedding matrices of single cells

The CellLoop algorithm initializes cell embeddings through distinct approaches tailored to specific data modalities. Notably, CellLoop requires reliable embeddings to ensure robust performance by providing each single cell with stable and biologically meaningful neighbors.

Single-modality 3D genomic data: For datasets comprising solely 3D genomic information, CellLoop employs Higashi[14], an integrative framework based on hypergraph representation learning, to derive initial cell embeddings.

Dual-modality data integrating 3D genomics and other omics: In scenarios where 3D genomic data is complemented by additional omics layers, CellLoop utilizes UMAP to generate initial cell embeddings, leveraging the supplementary omics data (refer to Methods: UMAP Visualization and Clustering of Single Cells).

Looking ahead, advancements in spatial 3D genomics are anticipated to enable the construction of initial cell embeddings by incorporating spatial coordinates, thereby enhancing the contextual understanding of chromatin architecture[36].

Quantitative performance evaluation metrices

We employed three metrics, GSD, CPC, and CDC scores, to quantitatively assess the performance of detected chromatin loops in comparison to control interactions. For each chromatin loop, control chromatin interactions were constructed by fixing one anchor of the loop and positioning the other anchor at the same distance in the opposite direction, equal to the length of the chromatin loop. These control interactions were required to be present in the raw single-cell contact map.

GSD score: High-quality 3D genome structures for single cells were generated at a 20 KB resolution, as previously described[2,41]. The GSD score for each chromatin loop was computed by measuring the Euclidean distance between the two loop anchors in the reconstructed 3D genomic space of each cell. A lower GSD score indicates higher-quality chromatin loops.

CPC score: The scA/B values were computed from raw contact maps of single cells using the "dip-c color2" function[2,41]. The CPC score for a given single cell was defined as the Pearson correlation coefficient between the scA/B values at the two anchor regions of the chromatin loop identified within that cell. A higher CPC score reflects better-quality chromatin loops in the single-cell analysis.

CDC score: We calculated the CpG density surrounding each loop anchor and computed the Pearson correlation coefficient between the two CpG density profiles. The resulting value reflects the degree of chromatin state similarity between loop anchors. A higher CDC scores reflects better-quality chromatin loops in the single-cell analysis.

Aggerated peak score (APS)

We conducted the APS analysis on two matrices at a 20 Kb resolution: the loop frequency map (LFmap) and the contact frequency map (CFmap). To quantify the aggregate enrichment of a set of chromatin loops in these matrices, we calculated the average of a series of submatrices extracted from both maps. Each submatrix is a 220 Kb × 220 Kb square, centered on a single chromatin loop located in the upper triangle of the matrices. The APS of a set of chromatin loops was determined by the ratio of the central pixel value to the mean of the pixels in the lower-left corner of the submatrices that exceed the average value.

The enrichment analysis of ChIP-seq peak on chromatin loops

We retrieved ChIP-seq peaks for various epigenetic factors from ChIP-Atlas (https://chip-atlas.org/)[42]. Enrichment analysis was performed at a 20 Kb resolution. To quantify the aggregate enrichment of a set of chromatin loops, we centered on each chromatin loop and computed the number of peaks associated with all chromatin interactions surrounding the loop. This yielded a peak enrichment submatrix for each chromatin loop. The final enrichment matrix was obtained by calculating the average of a series of submatrices derived from the set of chromatin loops. The matrix was visualized using the plot_surface function in the matplotlib package.

UMAP visualization and clustering of single cells

UMAP visualization and clustering of single-cell chromatin loop and RNA-seq matrices were performed using the Scanpy package[40]. For the single-cell chromatin loop matrices, we first filtered out chromatin loop features at a 100 Kb resolution that were observed in fewer than 10 cells. The filtered matrix was then normalized using the TF-IDF transform algorithm as previously described[43]. A lower-dimensional representation of the data was generated by incorporating the first 20 dimensions from the singular value decomposition (SVD) of the TF-IDF-transformed matrix. This reduced representation was subsequently used as input for the UMAP algorithm. Cell clusters were identified in the two-dimensional space using the Leiden clustering algorithm with default parameters.

For the single-cell RNA-seq matrix, cells with fewer than 100 genes and genes expressed in fewer than 5 cells were excluded. The remaining count data were normalized by multiplying each cell's library size by a scale factor (default value = 10,000). The normalized matrix was then log-transformed in preparation for downstream analysis. Principal component analysis (PCA) was performed to reduce the dimensionality, and the resulting representation was used as input for UMAP. The Leiden clustering algorithm with default parameters was then applied to detect cell clusters.

For the single-cell scA/B matrix, the visualization and clustering pipeline followed a similar procedure to that of the RNA-seq matrix, with the exception that filtering of cells and features was not performed.

Inference chromatin loop features for MERFISH cells

Our analysis focused on excitatory neuron cells present in both the GAGE-seq and MERFISH datasets. To infer the intensity of chromatin loop features in the MERFISH spatial data, we employed the Tangram algorithm[29]. First, we selected 254 genes detected in the MERFISH data as training genes to identify the optimal alignment between the transcriptome libraries of GAGE-seq and MERFISH, using the 'cells' mode. Subsequently, we used the project_cell_annotation function to transfer the chromatin loop features from GAGE-seq to the MERFISH dataset.

Identification of spatial domains using chromatin loop features for MERFISH cells

Chromatin loop features of MERFISH cells were analyzed using the STAGATE package[30] to identify spatial domains. Features detected in fewer than 50 cells were excluded from the analysis. The top 50 differentially expressed features for each cell type were then selected as input for the STAGATE algorithm. Count normalization for each cell was performed by scaling the library size with a factor of 10,000 (default value). The normalized matrix was subsequently log-transformed to prepare it for downstream analysis. To construct a cell spatial neighbor network, we set the rad_cutoff parameter to 60 and employed the train_STAGATE function to learn low-dimensional latent representations of cells. Finally, the Louvain algorithm was applied to cluster the spatial domains.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

In this work, we used several publicly available single-cell Hi-C datasets: Single-cell Dip-C for the developing mouse cortex and hippocampus[2] (GSE146397), single-cell double-modality data of mouse embryonic development using HiRES[3] (GSE223917), single cell double-modality data of mouse cortex using GAGE-seq[6] (GSE238001). For all scHi-C datasets, we kept only the high-quality coverage cells with more than 200,000 read pairs. After filtering, the Dip-C dataset contains 1812 cells, whereas the HiRES dataset contains 4273 cells and the GAGE-seq dataset contains 1925 cells. In single-cell maps of 20 Kb or higher resolution, CellLoop identifies chromatin loops within a 10 Mb genomic distance; while in those of 10 Kb or lower resolution, it identifies such loops within 5 Mb. The cell type annotation of Dip-C and HiRES datasets were provided in the original study, and cell types in GAGE-seq were determined using the CellTypist algorithm[44] using single-cell RNA-seq for mouse brain cortex from Tasic et al as reference[45]. MERFISH spatial transcriptome dataset was downloaded from (https://download.brainimagelibrary.org/cf/1c/cf1c1a431ef8d021/[35]). Single-cell RNA-seq datasets were downloaded from (https://figshare.com/articles/dataset/sc_mouse_cortex/13677259/1?file=26404781). Source data are provided with this paper.

Code availability

The software package implementing the CellLoop algorithm has been deposited at Github (https://github.com/YusenYe/CellLoop[46]).

References

1. Nagano, T. et al. Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61–67 (2017).
2. Tan, L. et al. Changes in genome architecture and transcriptional dynamics progress independently of sensory experience during post-natal brain development. *Cell* **184**, 741–758.e717 (2021).
3. Liu, Z. et al. Linking genome structures to functions by simultaneous single-cell Hi-C and RNA-seq. *Science* **380**, 1070–1076 (2023).
4. Tian, W. et al. Single-cell DNA methylation and 3D genome architecture in the human brain. *Science* **382**, eadf5357 (2023).
5. Tan, L. et al. Lifelong restructuring of 3D genome architecture in cerebellar granule cells. *Science* **381**, 1112–1119 (2023).
6. Zhou, T. et al. GAGE-seq concurrently profiles multiscale 3D genome organization and gene expression in single cells. *Nat. Genet.* **56**, 1701–1711 (2024).
7. Chang, L. et al. Droplet Hi-C enables scalable, single-cell profiling of chromatin architecture in heterogeneous tissues. *Nat. Biotechnol.* **43**, 1694–1707 (2024).
8. Wu, H. et al. Simultaneous single-cell three-dimensional genome and gene expression profiling uncovers dynamic enhancer connectivity underlying olfactory receptor choice. *Nat. Methods* **21**, 974–982 (2024).
9. Lee, L. et al. SnapHiC-D: a computational pipeline to identify differential chromatin contacts from single-cell Hi-C data. *Brief. Bioinforma.* **24**, bbad315 (2023).
10. Salameh, T. J. et al. A supervised learning framework for chromatin loop detection in genome-wide contact maps. *Nat. Commun.* **11**, 3428 (2020).
11. Kaul, A., Bhattacharyya, S. & Ay, F. Identifying statistically significant chromatin contacts from Hi-C data with FitHiC2. *Nat. Protoc.* **15**, 991–1012 (2020).
12. Li, X., Zeng, G., Li, A. & Zhang, Z. DeTOKI identifies and characterizes the dynamics of chromatin TAD-like domains in a single cell. *Genome Biol.* **22**, 217 (2021).

13. Ye, Y., Zhang, S., Gao, L., Zhu, Y. & Zhang, J. Deciphering hierarchical chromatin domains and preference of genomic position forming boundaries in single mouse embryonic stem cells. *Adv. Sci.* **10**, 2205162 (2023).
14. Zhang, R., Zhou, T. & Ma, J. Multiscale and integrative single-cell Hi-C analysis with Higashi. *Nat. Biotechnol.* **40**, 254–261 (2022).
15. Xiong, K., Zhang, R. & Ma, J. scGHOST: identifying single-cell 3D genome subcompartments. *Nat. Methods* **21**, 814–822 (2024).
16. Ye, Y., Gao, L. & Zhang, S. Circular trajectory reconstruction uncovers cell-cycle progression and regulatory dynamics from single-cell Hi-C maps. *Adv. Sci.* **6**, 1900986 (2019).
17. Yu, M. et al. SnapHiC: a computational pipeline to identify chromatin loops from single-cell Hi-C data. *Nat. Methods* **18**, 1056–1059 (2021).
18. Rodriguez, A. & Laio, A. Clustering by fast search and find of density peaks. *Science* **344**, 1492–1496 (2014).
19. Yaffe, E. & Tanay, A. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat. Genet.* **43**, 1059–1065 (2011).
20. Rao, S. uhasS. P. et al. A 3D Map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
21. Zhang, Y. et al. Computational methods for analysing multiscale 3D genome organization. *Nat. Rev. Genet.* **25**, 123–141 (2024).
22. Mach, P. et al. Cohesin and CTCF control the dynamics of chromosome folding. *Nat. Genet.* **54**, 1907–1918 (2022).
23. Hsieh, T.-H. S. et al. Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Mol. Cell* **78**, 539–553.e538 (2020).
24. Van Hoecke, A. et al. EPHA4 is a disease modifier of amyotrophic lateral sclerosis in animal models and in humans. *Nat. Med.* **18**, 1418–1422 (2012).
25. Sanborn, A. L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci.* **112**, E6456–E6465 (2015).
26. Snijders Blok, L. et al. Heterozygous variants that disturb the transcriptional repressor activity of FOXP4 cause a developmental disorder with speech/language delays and multiple congenital abnormalities. *Genet. Med.* **23**, 534–542 (2021).
27. Benowitz, L. I. & Routtenberg, A. GAP-43: an intrinsic determinant of neuronal development and plasticity. *Trends Neurosci.* **20**, 84–91 (1997).
28. Zhang, M. et al. Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature* **598**, 137–143 (2021).
29. Biancalani, T. et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat. Methods* **18**, 1352–1362 (2021).
30. Dong, K. & Zhang, S. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nat. Commun.* **13**, 1739 (2022).
31. Li, W.-K. et al. Whole-brain in vivo base editing reverses behavioral changes in Mef2c-mutant mice. *Nat. Neurosci.* **27**, 116–128 (2024).
32. Bos, P. D. et al. Genes that mediate breast cancer metastasis to the brain. *Nature* **459**, 1005–1009 (2009).
33. Faruqi, M. EPHA4 inhibition rescues neurodegeneration in ALS. *Nat. Rev. Drug Discov.* **11**, 747–747 (2012).
34. Takei, Y. et al. Spatial multi-omics reveals cell-type-specific nuclear compartments. *Nature* **641**, 1037–1047 (2025).
35. Zhuang, X. Spatially resolved single-cell genomics and transcriptomics by imaging. *Nat. Methods* **18**, 18–22 (2021).
36. Ahanger, S. H. et al. Spatial 3D genome organization controls the activity of bivalent chromatin during human neurogenesis. *bioRxiv*, 2024.2008.2001.606248 (2024).
37. Dubois, F., Sidiropoulos, N., Weischenfeldt, J. & Beroukhim, R. Structural variations in cancer and the 3D genome. *Nat. Rev. Cancer* **22**, 533–546 (2022).
38. Grover, A., Leskovec, J. node2vec: Scalable feature learning for networks. *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 855–864 (2016).
39. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **57**, 289–300 (1995).
40. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
41. Tan, L., Xing, D., Daley, N. & Xie, X. S. Three-dimensional genome structures of single sensory neurons in mouse visual and olfactory systems. *Nat. Struct. Mol. Biol.* **26**, 297–307 (2019).
42. Zou, Z., Ohta, T. & Oki, S. ChIP-Atlas 3.0: a data-mining suite to explore chromosome architecture together with large-scale regulome data. *Nucleic Acids Res.* **52**, W45–W53 (2024).
43. Cusanovich, D. A. et al. A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell* **174**, 1309–1324 (2018).
44. Xu, C. et al. Automatic cell-type harmonization and integration across Human Cell Atlas datasets. *Cell* **186**, 5876–5891 (2023).
45. Tasic, B. et al. Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018).
46. Ye, Y. et al. CellLoop: Identifying single-cell 3D genome chromatin loops. *Zenodo* https://doi.org/10.5281/zenodo.18784293 (2026).

Acknowledgements

This work was supported by National Natural Science Foundation of China grant (no. 62573335 to Y.Y., nos. 62132015, 62350087 and 62550005 to L.G., no. 62422318 to H.C., no. 62002277 to Y.H.), and a grant from the Fundamental Research Funds for the Central Universities (no. QTZX23074 to Y.Y.), National Key Research and Development Program of China (no. 2023YFF0725500 to H.C.) and the State Key Laboratory of Medical Proteomics (SKLP-K202407 to H.C.). We thank L. Zhang and W. Min for helpful discussions. We also thank the Key Laboratory of Computational Bioinformatics of Xi'an at Xidian University.

Author contributions

Y.Y. conceived and designed the study. Y.Y. designed the CellLoop algorithm. Y.Y. implemented the CellLoop algorithm. L.G., H.C., Y.H., L.Y., and W.W. provided additional input during the development of the methods. Y.Y. performed the data analysis. Y.Y. provided support for the development of the software packages. Y.Y. supervised the study. Y.Y. wrote the manuscript. Y.Y., S.Z., Y.W. and S.L. revised and supplemented the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-71406-y.

Correspondence and requests for materials should be addressed to Yusen Ye, Hebing Chen or Lin Gao.

Peer review information *Nature Communications* thanks Zhen Miao and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.