HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2024 August 14.

Published in final edited form as:

Nat Genet. 2024 August ; 56(8): 1701–1711. doi:10.1038/s41588-024-01745-3.

GAGE-seq concurrently profiles multiscale 3D genome organization and gene expression in single cells

Tianming Zhou[1], **Ruochi Zhang**[1,5], **Deyong Jia**[2], **Raymond T. Doty**[3], **Adam D. Munday**[3], **Daniel Gao**[4,6], **Li Xin**[2,4], **Janis L. Abkowitz**[3,4], **Zhijun Duan**[3,4,#], **Jian Ma**[1,#]

[1]Ray and Stephanie Lane Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA

[2]Department of Urology, University of Washington, Seattle, WA 98195, USA

[3]Division of Hematology and Oncology, Department of Medicine/Fred Hutch Cancer Center, University of Washington, Seattle, WA 98195, USA

[4]Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle, WA 98109, USA

[5]Present address: Eric and Wendy Schmidt Center, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

[6]Present address: Department of Chemistry, Pomona College, Claremont, CA 91711, USA

Abstract

The organization of mammalian genomes features a complex, multiscale three-dimensional (3D) architecture, whose functional significance remains elusive due to limited single-cell technologies that can concurrently profile genome organization and transcriptional activities. Here, we introduce GAGE-seq, a scalable, robust single-cell co-assay measuring 3D genome structure and transcriptome simultaneously within the same cell. Applied to mouse brain cortex and human bone marrow CD34+ cells, GAGE-seq characterized the intricate relationships between 3D genome and gene expression, showing that multiscale 3D genome features inform cell type-specific gene expression and link regulatory elements to target genes. Integration with spatial transcriptomic data revealed *in situ* 3D genome variations in mouse cortex. Observations in human hematopoiesis unveiled discordant changes between 3D genome organization and gene expression, underscoring a complex, temporal interplay at the single-cell level. GAGE-seq provides a powerful, cost-effective approach for exploring genome structure and gene expression relationships at the single-cell level across diverse biological contexts.

[#]Correspondence: zjduan@uw.edu (Z.D.) and jianma@cs.cmu.edu (J.M.).

Author contributions

Z.D. and J.M. conceived and oversaw the project. Z.D. conceived and developed the GAGE-seq protocol with critical contributions from T.Z. and J.M.. T.Z. developed the computational workflow and performed all the data analysis with assistance from R.Z., under the supervision of Z.D. and J.M.. D.J. provided mice and dissected the mouse brain tissues, under the supervision of L.X.. R.T.D. and A.M. prepared human PBMCs for method optimization, under the supervision of J.L.A.. D.G. performed experiments under the supervision of Z.D. T.Z., Z.D., and J.M. wrote the manuscript with input from all authors.

Competing interests

Z.D. is listed as the inventor on a provisional patent application that covered the GAGE-seq experimental protocol filed by the University of Washington. No competing interests are declared by the other authors.

INTRODUCTION

Connecting genotype to phenotype remains a challenge due to the complex principles governing genome functions. Mammalian genomes are organized within the three-dimensional (3D) space of the cell nucleus[1], featuring architectural structures across genomic scales, including chromosome territories[2], A/B compartments[3], subcompartments[3,4], topologically associating domains (TADs)[5,6] and subTADs[7,8], and chromatin loops[9,10]. These structures play critical roles for gene regulation, cellular development, and disease progression[11–16]. Single-cell analyses provide unique insights into these processes, uncovering variability in 3D genome features in individual cells that bulk analyses might mask[14,17,18]. Yet, understanding how changes in multiscale 3D genome structure within a single cell influence its transcriptional program and cellular phenotypes remains a major challenge in epigenomics.

Cellular and molecular heterogeneity is pivotal in differentiation and tissue development. Advances in single-cell technologies, such as scRNA-seq and single-cell Hi-C (scHi-C), have deepened our understanding of cellular heterogeneity[19–21] and 3D genome organization[17,22–28]. To fully unravel the connections between 3D genome organization and transcriptional activity in individual cells, technologies that can concurrently measure both in the same cell are needed. Current computational approaches enable some integration of scHi-C and scRNA-seq[27,29,30], revealing connections between 3D genome organization and gene expression at cell-type level. However, such integration cannot capture the individual cell differences and cell-to-cell variation between structure and function, as it correlates data from separate cells. While imaging-based methods can provide simultaneous 3D genome organization and transcriptional activity within the same cells, they are constrained by low throughput and limited genomic coverage[31–34]. These limitations underscore the need for high-throughput genomic technologies capable of co-assaying 3D genome and gene expression in the same cell.

Here, we report GAGE-seq (genome architecture and gene expression by sequencing), a highly scalable and cost-effective method for simultaneously profiling of chromatin interactions and gene expression in single cells. GAGE-seq, thanks to its combinatorial barcoding strategy, offers higher methodological throughput, as well as greater efficiency and effectiveness than recent technologies such as HiRES[35]. We applied GAGE-seq to profile 9,190 cells across diverse mammalian cell lines and tissues, including mouse brain and human bone marrow. Specifically, we developed an experimental and analytical framework for elucidating the connections between multiscale 3D genome features and cell type-specific gene expression, as well as their spatial and temporal interplay.

RESULTS

Overview of GAGE-seq

GAGE-seq is a high-throughput, effective, and robust single-cell multiomics technology that simultaneously profiles the 3D genome and transcriptome in individual cells (Fig. 1a). GAGE-seq leverages the highly scalable “combinatorial indexing” paradigm previously employed in sci-Hi-C[22,36–38], as well as other single-cell methods[39–42] (Fig. 1a). The

procedure can be summarized as follows: (i) The RNA in cross-linked and permeabilized cells or nuclei is reverse transcribed (RT) with a biotinylated poly(T) or random hexamer primer containing DNA sequences, facilitating the ligation of the first-round barcoded cDNA adaptors (Supplementary Fig. 1, Supplementary Table 1); (ii) Cross-linked chromatins are efficiently fragmented (the first round chromatin fragmentation) using two 4-cut restriction enzymes (RE), CviQI and MseI, both producing the same adhesive DNA end 5'-TA, enabling the identification of chromatin interactions via proximity ligation; (iii) After a second round of chromatin fragmentation to introduce adhesive DNA ends for ligating the first-round barcoded DNA adaptors (Supplementary Fig. 1, Supplementary Table 1), cells/nuclei are distributed to a 96-well plate, where the first-round barcodes for DNA or cDNA are introduced through ligation of barcoded adaptors; (iv) Intact cells/nuclei are then pooled, diluted, and redistributed to a second 96-well plate, where the second-round barcodes for DNA or cDNA are introduced through ligation; (v) After reverse-crosslinking to release barcoded nucleic acids, all genomic DNA and cDNA are pooled, and biotinylated cDNA fragments are separated from genomic DNA with streptavidin beads; (vi) Sequencing libraries for scHi-C and scRNA-seq are separately generated and sequenced (Methods); and finally, (vii) Matched scHi-C and scRNA-seq profiles are identified according to the well-specific barcoding combinations (Fig. 1a, Supplementary Fig. 1, Supplementary Table 1, Methods). This combinatorial cellular indexing strategy can be further extended to achieve even larger throughput using additional rounds of ligation-mediated barcoding.

Quality validation and benchmarking of GAGE-seq

To assess the quality and specificity of GAGE-seq data, we performed experiments using a mixture of human (K562) and mouse (NIH3T3) cell lines (Fig. 1b–e, Methods, Supplementary Methods). Successful separation of human and mouse reads in both scHi-C and scRNA-seq data was demonstrated, identifying 683 human and 568 mouse cells out of 1,500 expected, along with 57 doublets observed in line with the expected 4.4% collision rate (Fig. 1b–e). Cells passing stringent quality criteria exhibited an average of 181,240 (K562, 39.2% duplicate rate) and 206,113 (NIH3T3, 38.0% duplicate rate) chromatin contacts (>1Kb intra-chromosomal) for scHi-C, as well as an average of 24,784 (K562, 35.7% duplicate rate) and 16,596 (NIH3T3, 31.2% duplicate rate) unique molecular identifiers (UMIs) from 3,699 (K562) and 2,256 (NIH3T3) genes per cell for scRNA-seq (Fig. 1, Supplementary Table 2). These robust results underscore GAGE-seq's ability to concurrently measure single-cell chromatin interactions and transcriptome with high sensitivity and accuracy. In addition, GAGE-seq's efficient fragmentation of crosslinked chromatin before proximity ligation, enabled by two four-cutters (Fig. 1a, Methods), allows for efficient detection of multi-way interactions, with >25% of all identified chromatin contacts in each scHi-C library (Supplementary Table 2).

Validating GAGE-seq in additional cell lines, GM12878 and MDS-L, further confirmed its robustness, specificity, sensitivity, and reproducibility (Fig. 2, Supplementary Fig. 2 and 3, Methods, Supplementary Methods). Whole-genome and whole-library level analysis showed GAGE-seq's chromatin interaction and gene expression profiles strongly correlating with published datasets (Fig. 2a–b). Low collision rate (Fig. 1b), binomial distribution of scHi-C reads (Fig. 1b, Supplementary Fig. 2a and 3a), typical chromatin contact decay curve (Fig.

2c), high *cis-trans* ratio (Fig. 1c, Supplementary Fig. 2c and 3c, Supplementary Table 2–4), and aggregated pseudobulk and single-cell chromatin contact maps (Fig. 2d, Supplementary Fig. 4 and 6), as well as pseudobulk and single-cell A/B compartment scores and insulation scores (Fig. 2e), further confirmed the specificity of the GAGE-seq scHi-C signals. The specificity of the GAGE-seq scRNA-seq signals was demonstrated through low collision rate (4.6% in the K562/NIH3T3 library) (Fig. 1d), binomial distribution of RNA reads (Fig. 1d, Supplementary Fig. 2d and 3d), and the fact that the majority of RNA reads (86%) mapped to the gene body (Fig. 2f), complemented by the pseudobulk and single-cell RNA signal distribution at individual gene loci (Fig. 2g, Supplementary Fig. 5). Notably, similar to SHARE-seq[43], GAGE-seq scRNA-seq reads were found to be 25%−50% intronic (Fig. 2f), indicating enriched nascent RNA. The high reproducibility across replicates was demonstrated at multiple levels (Fig. 2a,b,d,e,g,h,i), and its methodological resolution (library complexity) of scHi-C matched existing lower-throughput, unimodal methods, such as Dip-C[26,27], as well as sn-m3C-seq[44,45] (Fig. 2j). GAGE-seq scRNA-seq data quality was also comparable to existing methods (Fig. 2k). In line with previous scHi-C studies[23,36], GAGE-seq scHi-C data revealed cell cycle stages (Supplementary Fig. 6, Supplementary Methods). Compared to the recent HiRES method[35], GAGE-seq offers major advantages in throughput, efficiency, and cost-effectiveness (Fig. 2j–k, Supplementary Fig. 10 and 11), as well as in resolving rare cell types in complex tissues (see later section and Supplementary Results)

GAGE-seq reveals complex cell types in mouse cortex

To demonstrate the utility of GAGE-seq in unveiling complex cell types based on single-cell 3D genome features and gene expression within a tissue context, we turned our focus to the adult mouse brain cortex, known for its cell type diversity. Applying GAGE-seq on cells from the mouse cortex (8-9 weeks old), we generated 3,296 high-quality joint single-cell profiles of chromatin interactions and transcriptomes (Methods, Supplementary Methods). On average, each cell displayed 231,136 chromatin contacts (at ~50% duplication rate), with 20,160 UMIs and 1,883 genes per cell (~59% duplication rate), in line with the adult mouse whole brain data from the recently published HiRES data (Supplementary Fig. 7 and 9, Fig. 2j–k, Supplementary Table 5).

Our GAGE-seq scRNA-seq data identified 28 known cell types across three major lineages in the mouse cortex, including 15 excitatory neuron subtypes, 8 inhibitory neuron subtypes, and 5 glial cell subtypes, such as astrocytes and oligodendrocytes (Fig. 3a–b, Supplementary Fig. 12 and 14). These cell identities were confirmed by unique marker gene expressions (Fig. 3b). Notably, GAGE-seq scRNA-seq data enabled the delineation of many rare neuronal subtypes not identified by HiRES[35], such as L5 PT CTX, Sncg, and Meis2 (Fig. 3a–b, Supplementary Fig. 13–15). Reanalysis of HiRES mouse brain data with Fast-Higashi[46] further confirmed the superior performance of GAGE-seq in identifying complex cell subtypes, despite a lower sequencing depth in GAGE-seq (Supplementary Fig. 16 and 17; Supplementary Results). Although 3D genome features are known to encode cell identity[35,47], scHi-C often identified fewer cell types in complex tissues than scRNA-seq[27,44,45,48]. Utilizing Fast-Higashi for scHi-C embedding, GAGE-seq distinguished all 28 transcriptome-defined cell types, including the aforementioned L5 PT CTX, Sncg, and

Meis2 rare subtypes (Fig. 3c, Supplementary Fig. 13–15). The scHi-C-based delineation supports these cell types with distinct 3D genome features, with insulation scores surrounding gene bodies showing cell type-specific connection with gene expression (Fig. 3d; see later section with more analysis).

Spatial integration reveals *in situ* 3D genome variation

Using GAGE-seq to map the 3D genome and transcriptome of single cells, we explored the *in situ* variation of the 3D genome in the adult mouse cortex. We leveraged GAGE-seq scRNA-seq as a "bridge" for this analysis. Recently, the spatial transcriptomics method MERFISH successfully discerned the spatial organization of distinct cell populations in the mouse primary motor cortex[49]. We started by integrating our GAGE-seq scRNA-seq data with the MERFISH data using Seurat[50], allowing us to establish a connection between the two datasets (Methods).

We focused on the excitatory neuron cell types present in both GAGE-seq and MERFISH datasets. Within the integrated embedding space, cells primarily clustered by cell type, and cells from both datasets integrated cohesively, indicating high correlation between cell types identified by the two methods (Fig. 3e, Supplementary Fig. 18, Methods). We next characterized the *in situ* variation of both marker gene expression and 3D genome features of these maker gene loci in the mouse cortex. As a proof of principle, we investigated the *in situ* pattern of marker genes for L5 intratelencephalic (IT) CTX. The observed and inferred gene expression demonstrated a high degree of congruence, further supporting the reliability of the integration (Spearman's r=0.76, two-sided P=0; Supplementary Fig. 19b–c, j–k). Layer 5, where L5 IT CTX cells reside, corresponded with the highest expression level, scA/B value[27], gene body score (Supplementary Methods), and a low single-cell insulation score (Fig. 3f–g, Supplementary Fig. 19), reinforcing the overall correlation between expression and 3D genome structure. Interestingly, despite consistently low expression levels and gene body scores in more superficial layers, the scA/B value increased and the single-cell insulation score decreased slightly around layer 3, a cortical layer containing the L2/3 IT CTX cells that are not adjacent to the tissue boundary, suggesting potential discrepancies of expression and various 3D genome features at finer spatial resolution (highlighted by arrows in Fig. 3f–g, Supplementary Fig. 19).

Impact of 3D genome on gene expressions in single cells

We next rigorously examined the relationship between gene expression and various multiscale 3D genome features in single cells, including A/B compartments, TAD-like domains, and chromatin loops.

Our analysis of the 3,461 genes expressed in inhibitory neurons (n=508) or excitatory neurons (n=1,938) revealed a strong correlation between cell type-specific gene expression and scA/B value, reflecting compartmentalization variations[27,29] (Fig 4a, top panels). Inhibitory neurons, for instance, showed a much higher expression for 432 genes which corresponded to a higher scA/B value (t-test P=1.1e-46; Fig. 4a, top middle panel). Most of the 391 genes with a higher scA/B value in inhibitory neurons also snowed notably higher expression levels in these cells compared to excitatory neurons (t-test P=7.5e-26, Fig. 4a, top

right panel). Overall, there is a significant correlation between differential gene expression and differential scA/B value (Pearson's r=0.38, P<1e-100, Fig. 4a, top left, Supplementary Fig. 20). At the chromatin domain level, we identified a negative correlation between cell type-specific gene expression and the associated single-cell insulation score across cell types (Fig. 4a, bottom panels, Supplementary Fig. 21), suggesting that TAD-like domain variations around the gene body are accompanied with changes in transcriptional activity of the gene. This phenomenon, aligning with previous findings at the cell type level[29], may be attributed to domain melting noted in highly expressed long genes in mouse hippocampus and midbrain neurons[47].

We subsequently examined the relationship between single-cell insulation score surrounding the gene body and the potential occurrence of domain melting within our diverse collection of cell types revealed by GAGE-seq (Supplementary Methods). We focused on the four genes (*Grik2*, *Dscam*, *Rbfox1*, and *Nrxn*) known to undergo domain melting[47], profiling their scA/B value, single-cell insulation score, and single-cell gene expression. Notably, these genes manifested high expression across almost all 28 cell subtypes revealed by GAGE-seq, with the exception of *Dscam* and *Grik2* in VLMC and Micro cells (Supplementary Fig. 22, Fig. 3d). As expected, *Dscam*, *Rbfox1*, and *Nrxn3* were predominantly in the active A compartment in the majority of cell subtypes (Fig. 3d, Supplementary Fig. 22), while the *Grik2* locus was in a weak B compartment across all the cells, despite its high expression (Supplementary Fig. 22). Aggregated single-cell insulation scores varied across the gene body, with most cell subtypes showing lower scores correlating with elevated gene expression (Fig. 3d, Supplementary Fig. 22). The aggregated chromatin contact maps indicate potential occurrence of domain melting around these gene bodies (Fig. 3d, Supplementary Fig. 23). A similar phenomenon was also detected for the *Rbfox1* locus across different excitatory neurons (Fig. 3d, low panels).

We next further confirmed the above observed connection between multiscale 3D genome features and gene expression at single cell resolution. Higher gene expression in a cell often corresponded to a higher scA/B value and lower single-cell insulation score in the same cell (Fig. 4b, Supplementary Fig. 24). For instance, of the 432 genes showing a significantly elevated scA/B value in inhibitory neurons, most displayed higher expression in these neurons than in excitatory neurons (Spearman's r=0.22, P=7.4e-28, n=2446 cells; Fig. 4b, top panel). At the chromatin domain level, the 198 genes expressed highly in Pvalb cells exhibited notably lower single-cell insulation scores than in other inhibitory neurons (Spearman's r=0.45, P=1.5e-26, n=508 cells; Fig. 4b, low panel). Thus, the connection between multiscale 3D genome features and gene expression is evident at the single-cell resolution.

We then confirmed our observations on single loci. As a proof of principle, we focused on the Pvalb inhibitory subtype (including both Pvalb a and Pvalb b). We first selected genes that have 1) significantly higher scA/B values and expression in inhibitory neurons compared to excitatory neurons (Fig. 4a, top panels, Supplementary Fig. 25), and 2) significantly higher expression and lower single-cell insulation scores in Pvalb compared to other inhibitory neurons (Fig. 4a, bottom panels, Supplementary Fig. 26, Supplementary Methods). This approach led us to the *Erbb4* gene. The *Erbb4* gene plays a pivotal role

in the central nervous system and has been linked to schizophrenia[51]. As expected, we observed differential A/B compartment states correlated with cell type-specific expression of the *Erbb4* gene (Fig. 4c), and differential single-cell insulation score that suggests domain melting in the gene locus (Fig. 4d, low panel). The TAD-like domain structure of the *Erbb4* gene body in Sst and Meis2 cells appears to be melted in Pvalb cells (i.e., less pronounced), which is again accompanied with high gene expression in Pvalb cells (Fig. 4d, top panel). Additionally, it appears that the *Erbb4* gene body interacts more frequently with the downstream two small TAD-like domains in Pvalb cells than in Sst and Meis2 cells (Fig. 4d, top panel). On a finer scale, we also observed a cell type-specific putative enhancer-promoter chromatin loop at the TSS of the *Erbb4* gene in Pvalb cells (Fig. 4e–g, Supplementary Methods). Moreover, when integrating with chromatin accessibility, the putative enhancer region exhibits differential chromatin accessibility that correlates with the cell type-specific expression of the *Erbb4* gene (Fig. 4f).

Integrative analysis of GAGE-seq and chromatin accessibility

We next aimed to demonstrate how integrating GAGE-seq with chromatin accessibility data enhances the connection between CREs and target genes. For this, we integrated GAGE-seq with Paired-seq data (from the same mouse cortex region)[52] (Supplementary Results, Supplementary Methods). Overall, genes with distinct contributions from 3D genome and chromatin accessibility show varied functions (Supplementary Fig. 27–28) and integrating 3D genome and chromatin accessibility data markedly improves gene expression prediction accuracy (Supplementary Fig. 29).

Our integrative analysis of GAGE-seq and chromatin accessibility enhances the connection of CREs to their target genes. The gene expression and transcription start site (TSS)-CRE interaction frequency correlation decreases with greater genomic distance between TSS and CRE (Fig. 5a, Supplementary Methods). Also, overlaps between Paired-seq-identified gene-CRE pairs and those identified by other approaches generally decrease with increasing genomic distance between TSS and CRE (Supplementary Fig. 30). However, refining with GAGE-seq data markedly improved this overlap, particularly for long-range (>100kb) gene-CRE pairs (Fig. 5b, Supplementary Fig. 30, Supplementary Results), highlighting the advantage of GAGE-seq in revealing CRE-gene pairs (Supplementary Results).

We also explored the joint regulation of gene expression by 3D genome and chromatin accessibility at individual gene loci. A strong correlation was found between *Epha4* gene expression and the chromatin interaction frequency with a distal CRE, as well as between *Epha4* gene expression and chromatin accessibility at the TSS and the distal CRE in different excitatory neuron subtypes (Fig. 5c–e). Motif analysis of chromatin accessibility peaks identified potential binding sites for transcription factors Twist2 (Spearman's P=1e-289) and Arx (Spearman's P=2e-132) (Fig. 5f). However, no significant differences were noted for A/B compartment value, insulation score, and gene body score of the *Epha4* locus across neuron subtypes (Supplementary Fig. 31), indicating that fine-scale CRE-chromatin looping instead of changes in the large-scale 3D features may be responsible for the cell type-specific *Epha4* expression (Supplementary Results).

Developmental stages of human hematopoiesis

Hematopoiesis is a classic model system with well-characterized cell type changes and their associated gene expression signatures, making it an ideal model for exploring the dynamic relationship between 3D genome structure and gene expression. We generated GAGE-seq profiles of 2,815 human bone marrow (BM) CD34+ cells after stringent quality filtering (Supplementary Fig. 32–34, Supplementary Table 6, Methods, Supplementary Methods), obtaining an average of 265,336 chromatin contacts (at ~50% duplication rate) and detecting on average 5,504 UMIs and 985 genes per cell (at ~63% duplication rate), which is in line with the publicly available scRNA-seq datasets (Supplementary Fig. 32–34, Supplementary Table 6). To mitigate the potential impact of 3D genome's cell-cycle dynamics[23], we restricted our analysis to high-quality G0/G1 phase cells (837 cells) (Supplementary Methods).

Unsupervised clustering of GAGE-seq scRNA-seq data revealed six clusters (five clusters with continuous diffusion and one distinct cluster), each displaying unique gene signatures (Fig. 6a–b). Based on the gene expression signatures and known marker genes[53], we annotated these clusters into known cell types: hematopoietic stem cell (HSC), multipotent progenitor (MPP), lymphoid-primed MPP (LMPP), multi-lymphoid progenitor (MLP), megakaryocyte-erythroid progenitor (MEP), and B lymphocyte natural killer cell progenitors (B-NK) (Fig. 6a–b). These clusters, representing all three major blood cell lineages, showed a lymphoid lineage preference. Our GAGE-seq scHi-C data also successfully resolved these six cell types (Fig. 6a–b), further demonstrating the ability of the 3D genome to encode cell type information.

Focusing on four of the six identified cell types (HSC, MPP, MLP and B-NK), which represent early B-NK lineage, we used GAGE-seq to reconstruct the developmental trajectory, demonstrating the dynamic interplay between genome structure and gene expression along this trajectory. Transcriptome and 3D genome-based pseudotime trajectories, inferred from GAGE-seq data, were highly congruent (Fig. 6c, Supplementary Fig. 35–36, Methods), suggesting that global 3D genome temporal variations overall mirror transcriptional changes and differentiation progression. Further, we created an integrated pseudotime trajectory (Fig. 6d, Methods), which was confirmed by the accurate alignment of the four cell types along the differentiation pseudotime and the observation that earlier-stage progenitors (e.g., HSCs) decrease while later-stage cells (e.g., B-NK) increase along the pseudotime (Fig. 6d–e).

Temporal interplays between 3D genome and gene expression

Comparisons between marker gene expression and 3D genome features in individual cell types during differentiation pseudotime suggest complex temporal interplay between both scA/B values and single-cell insulation scores with marker gene expressions (Supplementary Results, Supplementary Fig. 37–43).

We then performed an unsupervised clustering to further unravel relationships between gene expression and 3D genome features in the B-NK differentiation, based on all genes expressed in at least twenty single cells in the trajectory (Methods). We identified 11 distinct

gene clusters (Fig. 6f). Notably, 5 of these 11 clusters showed a negative correlation between the changes in gene expression and scA/B value over pseudotime (Fig. 6g left panel, Supplementary Fig. 41). We closely examined gene cluster 9, where expression increases while scA/B value decreases. We selected two genes, *JAK1* and *ITPR1*, which exhibit the highest similarity with the average temporal patterns of this gene cluster. Their scA/B value at the gene bodies indeed decreases over pseudotime without A/B compartment switches (Fig. 6h left panels).This analysis identified gene groups with varied temporal patterns, including discordant patterns in expression and scA/B value, as reported previously[27], during differentiation.

Regarding chromatin domains, a uniform temporal trend was observable in the aggregated single-cell insulation scores across all gene clusters, mirroring the pattern seen in the marker gene sets (Fig. 6g, Supplementary Results, Supplementary Fig. 41), indicating global 3D genome changes, manifested by widespread TAD-like domain re-organizations, in B-NK cells. For *JAK1* and *ITPR1,* the single-cell insulation scores increased abruptly from MLP to B-NK, correlating with gene expression (Fig. 6h right panels), supported by aggregated contact maps (Fig. 6i, Supplementary Fig. 42–43). Additionally, we found that genes of different sizes appear to have distinct patterns with respect to single-cell insulation scores (Supplementary Fig. 44).

DISCUSSION

Our high-throughput multiomic single-cell technology, GAGE-seq, delivers an integrative approach to co-assay 3D genome structure and gene expression in individual cells with high resolution. We show that GAGE-seq can reveal complex cell types from complex tissues not identified by other existing methods. Additionally, its data integration with spatial transcriptomic data points to great potential to reach a deeper understanding of 3D genome variation within complex tissues. Importantly, GAGE-seq also facilitates the reconstruction of differentiation trajectories based on 3D genome features, transcriptomes, or both. Our integration of GAGE-seq with single-cell chromatin accessibility data further highlights the advantage of GAGE-seq in linking CREs and their target genes. The high congruence between these modalities underscores the intimate connection between the temporal variations of the 3D genome and transcriptional rewiring during cell differentiation. Notably, GAGE-seq has revealed much more nuanced relationships between 3D genome features and gene expression during bone marrow B-NK lineage differentiation, creating a resource for future studies to disentangle causal gene regulatory changes in differentiation through the lens of 3D genome in single cells.

GAGE-seq is characterized by its efficiency, scalability, robustness, cost-effectiveness, and adaptability. We envision that GAGE-seq, along with our analytical tools, could significantly enhance the current toolkit for single-cell epigenomics. With wide-ranging applications, GAGE-seq can deepen our understanding of genome structure and function, providing insights into normal development and disease pathogenesis. Future refinements, such as enhancing barcoding strategy for higher throughput and improving detection of chromatin contacts, may allow GAGE-seq to construct high-resolution cell atlases and assess the role of pathogenic noncoding single-nucleotide variants on chromatin loops[54]

in a massively parallel manner. Additionally, we anticipate a future application where GAGE-seq will be integrated with spatial labeling technologies, producing spatially-resolved scHi-C and scRNA-seq data. Such advancements will likely open up new avenues of investigation, such as exploring the role of the 3D genome in various tissue development and disease progression. Ultimately, GAGE-seq may offer the opportunity to integrate different molecular features in single cells, leading to a more comprehensive understanding of genome structure, cellular function, and their spatiotemporal variability.

METHODS

Ethics statement

The present study complies with all pertinent ethical regulations. All the mice used in this study received humane care in compliance with the principles stated in the Guide for the Care and Use of Laboratory Animals, NIH Publication, 1996 edition, and the protocols were approved by the Institutional Animal Care Committee (IACUC) at the University of Washington.

Cell lines used

K562 (#CCL-243, ATCC), GM12878 (#GM12878, Coriell) and NIH3T3 cells (CRL-1658, ATCC) were purchased from the respective vendors. The myelodysplastic cell line MDS-L was a gift from Dr. Kaoru Tohyama (Kawasaki University of Medical Welfare). See Supplementary Methods for cell culture details.

GAGE-seq experimental details

Preparation of the 96-well plates of barcoded adaptors.—Two separate barcoding rounds of ligation reactions are used in GAGE-seq (as detailed in Supplementary Table 1). The design of the scRNA-seq part barcodes resembles that of Split-seq[42] and SHARE-seq[43] (Supplementary Table 1). The molecular structure of the scHi-C part barcodes is depicted in Supplementary Fig. 1.

Cell lysis.—Crosslinked cells of K562, NIH3T3, GM12878, MDS-L, human bone marrow Cd34+ cells (as detailed in the Supplementary Methods) were thawed from −80°C or liquid nitrogen. 0.2 ml of high-salt lysis buffer 1 (50 mM HEPES pH 7.4, 1 mM EDTA pH 8.0, 1 mM EgTA pH 8.0, 140 mM NaCl, 0.25% Triton X-100, 0.5% IGEPAL CA-630, 10% glycerol, and 1× proteinase inhibitor cocktail (PIC)) was added per 1×10^6 cells. The cell solution was mixed thoroughly and incubated on ice for 10 min. After this, cells were pelleted at 500xg for 2 min at 4°C and then resuspended in 0.2 ml of high-salt lysis buffer 2 (10 mM Tris-HCl pH 8, 1.5 mM EDTA, 1.5 mM EgTA, 200mM NaCl, 1× PIC). The solution was incubated on ice for 10 min. Following this, cells were then pelleted at 500xg for 2 min at 4°C and then resuspended in 200 μl of 1 × T4 DNA ligase buffer (NEB, B0202S) containing 0.2% SDS. They are then incubated at 58°C for 10 min. To quench the reaction, 200 μl ice-cooled 1x NWB and 10 μl 10% Triton X-100 (MilliporeSigma, 93443) were added to the tube. Finally, cells were spun at 500xg for 4 min at 4°C. For crosslinked mouse brain cortex cells, the treatment was simplified. The step involving high-salt lysis buffer 1 and high-salt lysis buffer 2 was omitted, and 0.1% SDS was used for cell lysis.

Reverse transcription.—SDS treated cells were resuspended in 400 μl of RT mix (final concentration of 1x RT buffer, 500 mM dNTP, 10 mM Biotinylated RT primers, 7.5% PEG 6000 (VWR, 101443-484), 0.4U/ml SUPERase•In™ RNase Inhibitor, and 25U/ml Maxima H Minus Reverse Transcriptase (ThermoFisher Scientific, EP0752)). The RT primers contain a poly dT tail, a biotin molecule, and a universal ligation overhang. The sample then underwent a series of heating cycles. Initially, it was heated at 50 oC for 10 minutes, then it went through 3 thermal cycles (8 °C for 12s, 15 °C for 45s, 20 °C for 45s, 30°C for 30s, 42 °C for 2 min and 50 °C for 3 min). Afterwards, the sample was again incubated at 50 oC for 10 minutes. After reverse transcription, 600 μl of 1x NWB was added, the sample was centrifuged at 500x g for 3 minutes, and the supernatant was then removed.

1st-round chromatin fragmentation, proximity ligation, and 2nd-round chromatin fragmentation.—Cells were resuspended in 400 μl of restriction enzyme (RE) digestion mix (1x T4 ligase buffer (NEB, B0202S), 500U MseI (NEB, R0525M), 240U CviQI (NEB, R0639L), 0.32 U/ml Enzymatics RNase Inhibitor, 0.05 U/ml SUPERase RNase Inhibitor), and incubated at room temperature (25 °C) for 2 hr. Cells were then centrifuged at 500x g for 3 minutes at 4 °C, and the supernatant was removed. The remaining cell pellet was washed twice with 300 μl of 1x NWB, and as much supernatant was removed as possible. Next, the pellet was resuspended in 200 μl of ligation mix (1x T4 ligation buffer (NEB, B0202S), 50 Units T4 DNA ligase (ThermoFisher Scientific, EL0012), 0.32 U/ml Enzymatics RNase Inhibitor, 0.05 U/ml SUPERase RNase Inhibitor) and incubated at 16 °C overnight. This was followed by adding 20 μl 10x T4 ligation buffer, 1 μl SUPERase RNase Inhibitor and 20 μl DdeI (NEB, R0175L). The sample was then incubated at 37 °C for 1 hr and centrifuged at 500x g for 3 minutes, with the supernatant removed afterwards.

Combinatorial cellular barcoding.—Cells were resuspended in 330 μl of ligation mix (1x T4 ligase buffer (NEB, B0202S), 100 Units T4 DNA ligase (ThermoFisher Scientific, EL0012), 0.25 mg/ml BSA (ThermoFisher Scientific, AM2618), 5% PEG-4000 (ThermoFisher Scientific, EL0012), 0.32 U/ml Enzymatics RNase Inhibitor, 0.05 U/ml SUPERase RNase Inhibitor) and distributed into each well (3 μl/well) of the first-round barcoding plate, which already contained 2 μl of CARE-seq 1st-round adaptors in each well. This barcoding plate was then incubated at 25 °C for 3 hr. Afterwards, cells from all 96 wells were pooled into three 1.5 ml tubes, and 5 μl of 10% NP-40 (ABCam, ab142227) was added to each tube. This is followed by centrifuging at 500x g for 3 minutes at 4 °C. The supernatant was then removed and cells were resuspended in 300 μl 1x NWB containing 0.033% SDS and combined into one 1.5 ml tube. Cells were then pelleted at 500x g for 2 minutes at 4 °C. After three additional rounds of washing with 300 μl 1x NWB containing 0.033% SDS, cells were resuspended in 200 μl 1x NWB containing 0.1% SDS and filtered with with 10 μm or 20 μm cell ministrainer (PluriStrainer, 43-10010-50 or 43-10020-40). Cells were inspected under a microscope and counted with a hemocytometer. Approximately 7,500 cells were diluted with 1.25 ml of a dilution buffer containing 0.4x NEBuffer 2 (NEB, B7002S), 2 mg/ml BSA (ThermoFisher Scientific, AM2618), and 0.08 μM RNA ligation-1 block, and distributed into each well (3 μl/well) of a 96-well plate (the 2nd-round barcoding

plate). Then, 2 μl of cell lysis buffer (5x NEBuffer 2, 0.625% SDS) were then added to each well of the 2nd-round barcoding plate. The plate was incubated at 60 °C for at least 24 hr.

For the 2nd-round barcoding, 1.5 μl of pre-mixed GAGE-seq adapters (0.2 μM Hi-C-AD2 and 0.17 μM RNA-AD2) were added to the plate, followed by 23.5 μl of ligation mix (3 μl 1x T4 ligase buffer (NEB, B0202S), 0.15 μl 50 mg/ml BSA (ThermoFisher Scientific, AM2618), 1 μl 10% Triton X-100 (MilliporeSigma, 93443), 0.03 μl 20 μM 5'-P-TNA-Nextera-P5-AD, 0.03 μl 20 μM 5'-P-TA-Nextera-P5-AD, 0.03 μl 10 μM RNA ligation-1 block, and 0.8 μl T4 DNA ligase (ThermoFisher Scientific, EL0012)). The ligation was carried out at 25 °C for 24 hr, and then stopped by adding 2 μl of proteinase K digestion mix (0.2 μl proteinase K (ThermoFisher Scientific, AM2546), 0.5 μl 10% SDS and 1.8 μl water) to each well. A reverse crosslinking was carried out at 60 °C for 20 hr.

Reverse crosslinking and separation of scHi-C and scRNA-seq libraries.—After reverse crosslinking, the sample in each 96-well plate was pooled into 12 DNA low-binding 1.5 ml tubes (Eppendorf, 022431021). Genomic DNA (gDNA) and cDNA were precipitated by adding 66 μl 3M Sodium Acetate Solution (pH 5.2) (MilliporeSigma,127-09-3), 1 μl GlycoBlue (ThermoFisher Scientific, AM9515) and 720 μl iso-propanol (MilliporeSigma, I9516) to each tube, followed by incubating at −80 °C for at least 1 hr. The samples were then centrifuged at 15000 rpm for 10 min and the pellet in each tube were resuspended in 30 μl 1x NEBuffer 2 containing 0.15% SDS. After incubation at 37 °C for 10 min, the samples were combined into one DNA low-binding tube. gDNA and cDNA were precipitated by adding 66 μl 3M Sodium Acetate Solution (pH 5.2) and 720 μl iso-propanol, followed by incubating at −80 °C for at least 1 hr. The sample was then centrifuged at 15000 rpm for 10 min and the pellet was resuspended in 100 μl buffer EB (Qiagen, 19086). For each sample of a 96-well plate, 5.5 μl of MyOne C1 Dynabeads were washed twice with 1x B&W-T buffer (5mM Tris pH 8.0, 1M NaCl, 0.5mM EDTA, and 0.05% Tween 20) and resuspended in μl of 2x B&W buffer (10mM Tris pH 8.0, 2M NaCl, and 1mM EDTA) and added to the sample tube. The mixture was incubated at room temperature for 60 min and put on a magnetic stand to separate supernatant and beads.

Library construction and sequencing.—Sequencing libraries were constructed as described in detail in the Supplementary Methods. Both scHi-C and scRNA-seq libraries were pooled and paired end sequencing (PE 150) were performed on the HiSeq, NextSeq, or NovaSeq platform (Illumina).

GAGE-seq data processing workflow

Demultiplexing.—DNA and RNA reads were assigned to wells based on the two rounds of barcodes. For DNA reads, only read 2 was used for demultiplexing, allowing at most 1 mismatch in each of the two rounds of barcodes. DNA reads with more than 5 mismatches in the region between the two rounds of barcodes (the 9th-23rd nt) were discarded. After demultiplexing, the first 12 nt were removed from read 1 and the first 35 nt were removed from read 2. For RNA reads, only read 1 was used for demultiplexing, allowing at most 1 mismatch in each barcode round. RNA reads with more than 6 mismatches in the region

between the two rounds of barcodes (the 19th-48th nt) or with more than 6 mismatches in the region downstream of the first round of barcode (the 57th-71th nt) were discarded.

The two reference genomes were combined into a single reference genome file used for all GAGE-seq libraries. For DNA reads, BWA (0.7.17)[68] was used for alignment. The combined reference genome was indexed using command bwa index -a bwtsw. Paired, trimmed DNA reads were aligned to the combined reference genome using command bwa mem -SP5M. For RNA reads, STAR (2.7.8a)[69] was used for alignment. The GENCODE annotation files for human (v36) and mouse (vM25) were downloaded and concatenated. The combined reference genome was indexed using command --runMode genomeGenerate --sjdbOverhang 100 with the combined gencode annotation file. Only read 2 of RNA reads was aligned with the command STAR --outSAMunmapped Within.

Identification of contact pairs from DNA reads.—Pairtools (0.3.1.dev1)[70] was used to identify contact pairs from paired DNA reads with command pairtools parse --walks-policy all --no-flip --min-mapq=10. After that, walk reads (i.e., DNA reads containing multiple ligation sites) were further processed. Briefly, we assumed that any pair of loci in the same DNA read forms a valid contact pair, and these contact pairs were included in the results.

Deduplication of contact pairs.—The contact pairs were deduplicated. We extract the genomic positions of the two ends of each contact pair. We define two contact pairs as directly duplicated if the two contact pairs' first ends lie within 500 nt apart and their second ends also within 500 nt. If two contact pairs are not directly duplicated, but are directly or indirectly duplicated with a third contact pair, we define the first two contact pairs as indirectly duplicated. Among each cluster (i.e., connected component) of (in)directly duplicated contact pairs, the one with the largest difference between its two ends' genomic positions was retained, and the rest were marked as duplicates.

Deduplication of RNA reads.—The RNA reads were deduplicated. Two RNA reads are defined as directly duplicated if there is at most 1 mismatch in their UMI and if their genomic positions differ by at most 5 nt. The rest of the process is similar to the deduplication of contact pairs. Only one RNA read from each duplicate cluster is retained.

GAGE-seq integrative analysis for mouse brain cortex.

Integration with MERFISH data.—Integration of GAGE-seq data and MERFISH data was done with Seurat (4.1.1)[50]. Only scRNA-seq profiles from the GAGE-seq data were used for this integration. In the GAGE-seq mouse brain cortex data, the following cell types of excitatory neurons were used: L2/3 IT CTX a, L2/3 IT CTX b, L2/3 IT CTX c, L4 IT CTX, L4/5 IT CTX, L5 IT CTX, L6 IT CTX, L6 CT CTX a, L6 CT CTX b, L5/6 NP CTX, and L6b CTX. In the MERFISH data, cells from L2/3 IT, L4/5 IT, L5 IT, L5/6 NP, L6 CT, L6 IT, and L6b were used. Each time, the selected cells from GAGE-seq were integrated with one slice from the MERFISH data. All genes detected and expressed in both GAGE-seq and MERFISH were used. The 'FindIntegrationAnchors' and 'IntegrateData' functions were used with default parameters, except that the number of dimensions was set to 10.

Inference of whole-transcriptome expression and 3D genome features for MERFISH cells.—The integrated single-cell expression profiles of GAGE-seq data and MERFISH data were scaled by the 'ScaleData' function from Seurat with default parameters, and the first 30 PCs were calculated by the 'RunPCA' function. A 50-nearest neighbor regressor was created to estimate whole-transcriptome expression and 3D genome features from the 30-dimensional PC space. The regressor was trained on GAGE-seq data and then applied to the MERFISH data. The Gaussian kernel was used as the weight function. For each MERFISH cell, the bandwidth was defined as the 0.3 quantile of the distances to the 50 nearest neighbors.

Integration with Paired-seq data.—The integration of GAGE-seq data with Paired-seq data[52] was done using Seurat. Only scRNA-seq profiles from the GAGE-seq data and the Paired-seq data were used for this integration. In the GAGE-seq mouse brain cortex data, we excluded three cell types: L2 IT RvPP, L2/3 IT RSP, and L5 IT RSP. In the Paired-seq data, cells from BR_NonNeu_Endothelial, HC_ExNeu_CA1, HC_ExNeu_CA23, HC_ExNeu_DG, HC_ExNeu_Subiculum, and HC_NonNeu_Ependymal were excluded. The 'SelectIntegrationFeatures', 'FindIntegrationAnchors' and 'IntegrateData' functions were used with default parameters.

Inference of accessibility for GAGE-seq cells.—The integrated single-cell expression profiles of GAGE-seq data and Paired-seq data were scaled by the 'ScaleData' function from Seurat with default parameters. The first 20 PCs were calculated by the 'RunPCA' function. To estimate whole-transcriptome expression and 3D genome features from the 40-dimensional PC space, we created a 50-nearest neighbor regressor, which was trained on Paired-seq data and then applied to the GAGE-seq data. The Gaussian kernel was used as the weight function. For each GAGE-seq cell, the bandwidth was set based on the 0.3 quantile of the distances to the 40 nearest neighbors.

GAGE-seq integrative analysis for bone marrow

Trajectory and pseudotime.—The pseudotime of human bone marrow cells was inferred by the 'sc.tl.diffmap' and 'sc.tl.dpt' function in Scanpy (1.9.3)[71], jointly from the paired scRNA-seq profiles and scHi-C profiles. Specifically, cells in the HSC, MPP, MLP, and B-NK clusters were included. The first 5 PCs of the scRNA-seq profiles were used for the scRNA-based pseudotime and the first 2 PCs of the Fast-Higashi embeddings of the scHi-C profiles were used for the scHi-C-based pseudotime. The 5 scRNA-seq PCs and the 2 scHi-C PCs were then concatenated and used for the joint pseudotime. The 'sc.pp.neighbors' function was used to construct the neighbor graph with 30 (scRNA-based and joint pseudotime) or 20 (scHi-C-based pseudotime) nearest neighbors per cell. The 'sc.tl.diffmap' and 'sc.tl.dpt' function was applied with 10 diffusion components to learn a latent representation focusing on the trajectory and to infer the pseudotime for single cells. The origin of the trajectory was set based on the average expression level of HSC marker genes previously identified[53].

Unsupervised clustering of genes.—The clustering of genes was based on the expression and scA/B value. Genes expressed in at least 20 cells were included. To generate

features for genes, 1) the expression levels and scA/B values were z-score normalized per gene among all cells. 2) cells were evenly divided into 10 bins based on the pseudotime, and 3) the average values of the expression and scA/B value in each bin were calculated for each gene. This process led to 20 features for each gene. The Louvain clustering algorithm was then applied to genes with 20 neighbors, a resolution of 1.5. The correlation was used as the distance metric.

Statistics and reproducibility

Boxplots in all figures show the median, first, and third quartiles, and whiskers extend no further than 1.5× interquartile range. The robustness and reproducibility of GAGE-seq were validated extensively by using multiple cell lines and primary tissue cells (both mouse and human). Blinding was not relevant to the study, thus data collection and analysis were not performed blind to the conditions of the experiments. No statistical method was used to predetermine sample size. The experiments were not randomized. Specifics on additional statistics used for each analysis are described in detail in the Methods section and Supplementary Methods.

Additional experimental methods, methods for quality control and benchmarking, methods for identifying single-cell 3D genome features, and other methods are described in the Supplementary Methods.

Data availability

All sequencing data from this study have been submitted to GEO under the accession # GSE238001. We use the following publicly available datasets in this work: *in situ* Hi-C datasets from Rao et al.[3] (GSE: GSE63525); scHi-C datasets from Nagano et al.[17] (GEO: GSE48262), Nagano et al.[23] (GEO: GSE94489), Ramani et al.[22] (GEO: GSE84920), Kim et al.[37] (4DN Data Portal: 4DNES4D5MWEZ, 4DNESUE2NSGS, 4DNESIKGI39T, 4DNES1BK1RMQ, and 4DNESTVIP977), Tan et al.[26] (GEO: GSE117876), Tan et al.[57] (GEO: GSE121791), Tan et al.[27] (GEO: GSE162511), Flyamer et al.[24] (GEO: GSE80006), Gassler et al.[60] (GEO: GSE100569), Stevens et al.[25] (GEO: GSE80280), Collombet et al.[59] (GEO: GSE129029), Lee et al.[44] (GEO: GSE124391), Liu et al.[45] (GEO: GSE132489), and Mulqueen et al.[58] (GEO: GSE174226); scRNA-seq datasets from Chen et al.[62] (GEO: GSE126074), Plongthongkum et al.[56] (GEO: GSE157660), Chen et al.[55] (GEO: GSE178707), Ma et al.[43] (GEO: GSE140203), Xu et al.[65] (ArrayExpress: E-MTAB-11264), Xiong et al.[66] (GEO: GSE158435), Zhu et al.[63] (GEO: GSE130399), Zhu et al.[52] (GEO: GSE152020), Cao et al.[61] (GEO: GSE117089), Mimitou et al.[64] (GEO: GSE126310), and Zhang et al.[53] (GEO: GSE137864); HiRES co-assayed scHi-C and scRNA-seq datasets from Liu et al.[35] (GEO: GSE223917); MERFISH spatial transcriptome datasets from Zhang et al.[49] (Brain Image Library: cf1c1a431ef8d021); Paired-seq co-assayed scRNA-seq and scATAC-seq from Zhu et al.[52] (GEO: GSE152020).

Code availability

The source code of the GAGE-seq data processing and analysis workflows can be accessed at: https://github.com/ma-compbio/GAGE-seq, which has also been deposited via Zenedo (https://doi.org/10.5281/zenodo.10888453)[72]. In our GitHub repository, we have provided

notebooks (https://github.com/ma-compbio/GAGE-seq/tree/main/scripts_analysis) that detail the integration between GAGE-seq and Paired-seq data for single-cell joint analysis of 3D genome structure, chromatin accessibility, and gene expression.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors would like to thank Yang Zhang for assistance with the figures. This work was primarily supported by the National Institutes of Health (NIH) grant R01HG012303 (J.M. and Z.D.), with additional funding, in part, provided by NIH Common Fund 4D Nucleome Program grants UM1HG011593 (J.M.) and UM1HG011586 (Z.D.), NIH Common Fund Cellular Senescence Network Program grant UG3CA268202 (J.M.), and NIH grants R01HG007352 (J.M.) and R61DA047010 (Z.D.). Z.D. was additionally supported by EvansMDS Discovery Research Grant 2019. J.M. received additional support from a Guggenheim Fellowship from the John Simon Guggenheim Memorial Foundation, a Google Research Collabs Award, and a Single-Cell Biology Data Insights award from the Chan Zuckerberg Initiative. R.Z. was supported by the Eric and Wendy Schmidt Center at the Broad Institute. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

References

1. Dekker J, Belmont AS, Guttman M, Leshyk VO, Lis JT, Lomvardas S, et al. The 4D nucleome project. Nature 2017;549:219–26. [PubMed: 28905911]
2. Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet 2001;2:292–301. [PubMed: 11283701]
3. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 2014;159:1665–80. [PubMed: 25497547]
4. Xiong K, Ma J. Revealing Hi-C subcompartments by imputing inter-chromosomal chromatin interactions. Nat Commun 2019;10:5069. [PubMed: 31699985]
5. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 2012;485:376–80. [PubMed: 22495300]
6. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 2012;485:381–5. [PubMed: 22495304]
7. Phillips-Cremins JE, Sauria MEG, Sanyal A, Gerasimova TI, Lajoie BR, Bell JSK, et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell 2013;153:1281–95. [PubMed: 23706625]
8. Beagan JA, Phillips-Cremins JE. On the existence and functionality of topologically associating domains. Nat Genet 2020;52:8–16. [PubMed: 31925403]
9. Salameh TJ, Wang X, Song F, Zhang B, Wright SM, Khunsriraksakul C, et al. A supervised learning framework for chromatin loop detection in genome-wide contact maps. Nat Commun 2020;11:3428. [PubMed: 32647330]
10. Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, et al. CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell 2015;163:1611–27. [PubMed: 26686651]
11. Marchal C, Sima J, Gilbert DM. Control of DNA replication timing in the 3D genome. Nat Rev Mol Cell Biol 2019;20:721–37. [PubMed: 31477886]
12. Ma J, Duan Z. Replication Timing Becomes Intertwined with 3D Genome Organization. Cell 2019:681–4. [PubMed: 30735630]
13. Zheng H, Xie W. The role of 3D genome organization in development and cell differentiation. Nat Rev Mol Cell Biol 2019;20:535–50. [PubMed: 31197269]

14. Misteli T The Self-Organizing Genome: Principles of Genome Architecture and Function. Cell 2020;183:28–45. [PubMed: 32976797]
15. Spielmann M, Lupiáñez DG, Mundlos S. Structural variation in the 3D genome. Nat Rev Genet 2018;19:453–67. [PubMed: 29692413]
16. Oudelaar AM, Higgs DR. The relationship between genome structure and function. Nat Rev Genet 2021;22:154–68. [PubMed: 33235358]
17. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature 2013;502:59–64. [PubMed: 24067610]
18. Zhou T, Zhang R, Ma J. The 3D Genome Structure of Single Cells. Annu Rev Biomed Data Sci 2021;4:21–41. [PubMed: 34465168]
19. Stuart T, Satija R. Integrative single-cell analysis. Nat Rev Genet 2019;20:257–72. [PubMed: 30696980]
20. Cao J, O'Day DR, Pliner HA, Kingsley PD, Deng M, Daza RM, et al. A human cell atlas of fetal gene expression. Science 2020;370:. 10.1126/science.aba7721. [PubMed: 32703862]
21. Calderon D, Blecher-Gonen R, Huang X, Secchia S, Kentro J, Daza RM, et al. The continuum of Drosophila embryonic development at single-cell resolution. Science 2022;377:eabn5800. [PubMed: 35926038]
22. Ramani V, Deng X, Qiu R, Gunderson KL, Steemers FJ, Disteche CM, et al. Massively multiplex single-cell Hi-C. Nat Methods 2017;14:263–6. [PubMed: 28135255]
23. Nagano T, Lubling Y, Várnai C, Dudley C, Leung W, Baran Y, et al. Cell-cycle dynamics of chromosomal organization at single-cell resolution. Nature 2017;547:61–7. [PubMed: 28682332]
24. Flyamer IM, Gassler J, Imakaev M, Brandão HB, Ulianov SV, Abdennur N, et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. Nature 2017;544:110–4. [PubMed: 28355183]
25. Stevens TJ, Lando D, Basu S, Atkinson LP, Cao Y, Lee SF, et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. Nature 2017;544:59–64. [PubMed: 28289288]
26. Tan L, Xing D, Chang C-H, Li H, Xie XS. Three-dimensional genome structures of single diploid human cells. Science 2018;361:924–8. [PubMed: 30166492]
27. Tan L, Ma W, Wu H, Zheng Y, Xing D, Chen R, et al. Changes in genome architecture and transcriptional dynamics progress independently of sensory experience during post-natal brain development. Cell 2021;184:741–758.e17. [PubMed: 33484631]
28. Li G, Liu Y, Zhang Y, Kubo N, Yu M, Fang R, et al. Joint profiling of DNA methylation and chromatin architecture in single cells. Nat Methods 2019;16:991–3. [PubMed: 31384045]
29. Zhang R, Zhou T, Ma J. Multiscale and integrative single-cell Hi-C analysis with Higashi. Nat Biotechnol 2022;40:254–61. [PubMed: 34635838]
30. Luo C, Liu H, Xie F, Armand EJ, Siletti K, Bakken TE, et al. Single nucleus multi-omics identifies human cortical cell regulatory genome diversity. Cell Genom 2022;2:. 10.1016/j.xgen.2022.100107.
31. Cardozo Gizzi AM, Cattoni DI, Fiche J-B, Espinola SM, Gurgo J, Messina O, et al. Microscopy-based chromosome conformation capture enables simultaneous visualization of genome organization and transcription in intact organisms. Mol Cell 2019;74:212–222.e5. [PubMed: 30795893]
32. Mateo LJ, Murphy SE, Hafner A, Cinquini IS, Walker CA, Boettiger AN. Visualizing DNA folding and RNA in embryos at single-cell resolution. Nature 2019;568:49–54. [PubMed: 30886393]
33. Su J-H, Zheng P, Kinrot SS, Bintu B, Zhuang X. Genome-scale imaging of the 3D organization and transcriptional activity of chromatin. Cell 2020;182:1641–1659.e26. [PubMed: 32822575]
34. Takei Y, Yun J, Zheng S, Ollikainen N, Pierson N, White J, et al. Integrated spatial genomics reveals global architecture of single nuclei. Nature 2021;590:344–50. [PubMed: 33505024]
35. Liu Z, Chen Y, Xia Q, Liu M, Xu H, Chi Y, et al. Linking genome structures to functions by simultaneous single-cell Hi-C and RNA-seq. Science 2023;380:1070–6. [PubMed: 37289875]
36. Ramani V, Deng X, Qiu R, Lee C, Disteche CM, Noble WS, et al. Sci-Hi-C: A single-cell Hi-C method for mapping 3D genome organization in large number of single cells. Methods 2020;170:61–8. [PubMed: 31536770]

37. Kim H-J, Yardımcı GG, Bonora G, Ramani V, Liu J, Qiu R, et al. Capturing cell type-specific chromatin compartment patterns by applying topic modeling to single-cell Hi-C data. PLoS Comput Biol 2020;16:e1008173. [PubMed: 32946435]
38. Bonora G, Ramani V, Singh R, Fang H, Jackson DL, Srivatsan S, et al. Single-cell landscape of nuclear configuration and gene expression during stem cell differentiation and X inactivation. Genome Biol 2021;22:279. [PubMed: 34579774]
39. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 2015;523:486–90. [PubMed: 26083756]
40. Cusanovich DA, Daza R, Adey A, Pliner HA, Christiansen L, Gunderson KL, et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science 2015;348:910–4. [PubMed: 25953818]
41. Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. Science 2017;357:661–7. [PubMed: 28818938]
42. Rosenberg AB, Roco CM, Muscat RA, Kuchina A, Sample P, Yao Z, et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science 2018;360:176–82. [PubMed: 29545511]
43. Ma S, Zhang B, LaFave LM, Earl AS, Chiang Z, Hu Y, et al. Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. Cell 2020;183:1103–1116.e20. [PubMed: 33098772]
44. Lee D-S, Luo C, Zhou J, Chandran S, Rivkin A, Bartlett A, et al. Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. Nat Methods 2019;16:999–1006. [PubMed: 31501549]
45. Liu H, Zhou J, Tian W, Luo C, Bartlett A, Aldridge A, et al. DNA methylation atlas of the mouse brain at single-cell resolution. Nature 2021;598:120–8. [PubMed: 34616061]
46. Zhang R, Zhou T, Ma J. Ultrafast and interpretable single-cell 3D genome analysis with Fast-Higashi. Cell Syst 2022;13:798–807.e6. [PubMed: 36265466]
47. Winick-Ng W, Kukalev A, Harabula I, Zea-Redondo L, Szabó D, Meijer M, et al. Cell-type specialization is encoded by specific chromatin topologies. Nature 2021;599:684–91. [PubMed: 34789882]
48. Heffel MG, Zhou J, Zhang Y, Lee D-S, Hou K, Alonso OP, et al. Epigenomic and chromosomal architectural reconfiguration in developing human frontal cortex and hippocampus. BioRxiv 2022:2022.10.07.511350. 10.1101/2022.10.07.511350.
49. Zhang M, Eichhorn SW, Zingg B, Yao Z, Cotter K, Zeng H, et al. Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. Nature 2021;598:137–43. [PubMed: 34616063]
50. Chidester B, Zhou T, Alam S, Ma J. SPICEMIX enables integrative single-cell spatial modeling of cell identity. Nat Genet 2023;55:78–88. [PubMed: 36624346]
51. Law AJ, Kleinman JE, Weinberger DR, Weickert CS. Disease-associated intronic variants in the ErbB4 gene are related to altered ErbB4 splice-variant expression in the brain in schizophrenia. Hum Mol Genet 2007;16:129–41. [PubMed: 17164265]
52. Zhu C, Zhang Y, Li YE, Lucero J, Behrens MM, Ren B. Joint profiling of histone modifications and transcriptome in single cells from mouse brain. Nat Methods 2021;18:283–92. [PubMed: 33589836]
53. Zhang Y, Xie X, Huang Y, Liu M, Li Q, Luo J, et al. Temporal molecular program of human hematopoietic stem and progenitor cells after birth. Dev Cell 2022;57:2745–2760.e6. [PubMed: 36493772]
54. Nasser J, Bergman DT, Fulco CP, Guckelberger P, Doughty BR, Patwardhan TA, et al. Genome-wide enhancer maps link risk variants to disease genes. Nature 2021;593:238–43. [PubMed: 33828297]
55. Chen AF, Parks B, Kathiria AS, Ober-Reynolds B, Goronzy JJ, Greenleaf WJ. NEAT-seq: simultaneous profiling of intra-nuclear proteins, chromatin accessibility and gene expression in single cells. Nat Methods 2022;19:547–53. [PubMed: 35501385]

56. Plongthongkum N, Diep D, Chen S, Lake BB, Zhang K. Scalable dual-omics profiling with single-nucleus chromatin accessibility and mRNA expression sequencing 2 (SNARE-seq2). Nat Protoc 2021;16:4992–5029. [PubMed: 34650278]

57. Tan L, Xing D, Daley N, Xie XS. Three-dimensional genome structures of single sensory neurons in mouse visual and olfactory systems. Nat Struct Mol Biol 2019;26:297–307. [PubMed: 30936528]

58. Mulqueen RM, Pokholok D, O'Connell BL, Thornton CA, Zhang F, O'Roak BJ, et al. High-content single-cell combinatorial indexing. Nat Biotechnol 2021;39:1574–80. [PubMed: 34226710]

59. Collombet S, Ranisavljevic N, Nagano T, Varnai C, Shisode T, Leung W, et al. Parental-to-embryo switch of chromosome organization in early embryogenesis. Nature 2020;580:142–6. [PubMed: 32238933]

60. Gassler J, Brandão HB, Imakaev M, Flyamer IM, Ladstätter S, Bickmore WA, et al. A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. EMBO J 2017;36:3600–18. [PubMed: 29217590]

61. Cao J, Cusanovich DA, Ramani V, Aghamirzaie D, Pliner HA, Hill AJ, et al. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. Science 2018;361:1380–5. [PubMed: 30166440]

62. Chen S, Lake BB, Zhang K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. Nat Biotechnol 2019;37:1452–7. [PubMed: 31611697]

63. Zhu C, Yu M, Huang H, Juric I, Abnousi A, Hu R, et al. An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome. Nat Struct Mol Biol 2019;26:1063–70. [PubMed: 31695190]

64. Mimitou EP, Cheng A, Montalbano A, Hao S, Stoeckius M, Legut M, et al. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. Nat Methods 2019;16:409–12. [PubMed: 31011186]

65. Xu W, Yang W, Zhang Y, Chen Y, Hong N, Zhang Q, et al. ISSAAC-seq enables sensitive and flexible multimodal profiling of chromatin accessibility and gene expression in single cells. Nat Methods 2022;19:1243–9. [PubMed: 36109677]

66. Xiong H, Luo Y, Wang Q, Yu X, He A. Single-cell joint detection of chromatin occupancy and transcriptome enables higher-dimensional epigenomic reconstructions. Nat Methods 2021;18:652–60. [PubMed: 33958790]

67. Yao Z, van Velthoven CTJ, Nguyen TN, Goldy J, Sedeno-Cortes AE, Baftizadeh F, et al. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. Cell 2021;184:3222–3241.e26. [PubMed: 34004146]

Methods-only References

68. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 2009;25:1754–60. [PubMed: 19451168]

69. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15–21. [PubMed: 23104886]

70. Goloborodko A, Abdennur N, Venev S, hbbrandao, gfudenberg. mirnylab/pairtools: v0.2.0. 2018.

71. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol 2018;19:15. [PubMed: 29409532]

72. Zhou T GAGE-seq analysis workflow. 2024. 10.5281/zenodo.10888453

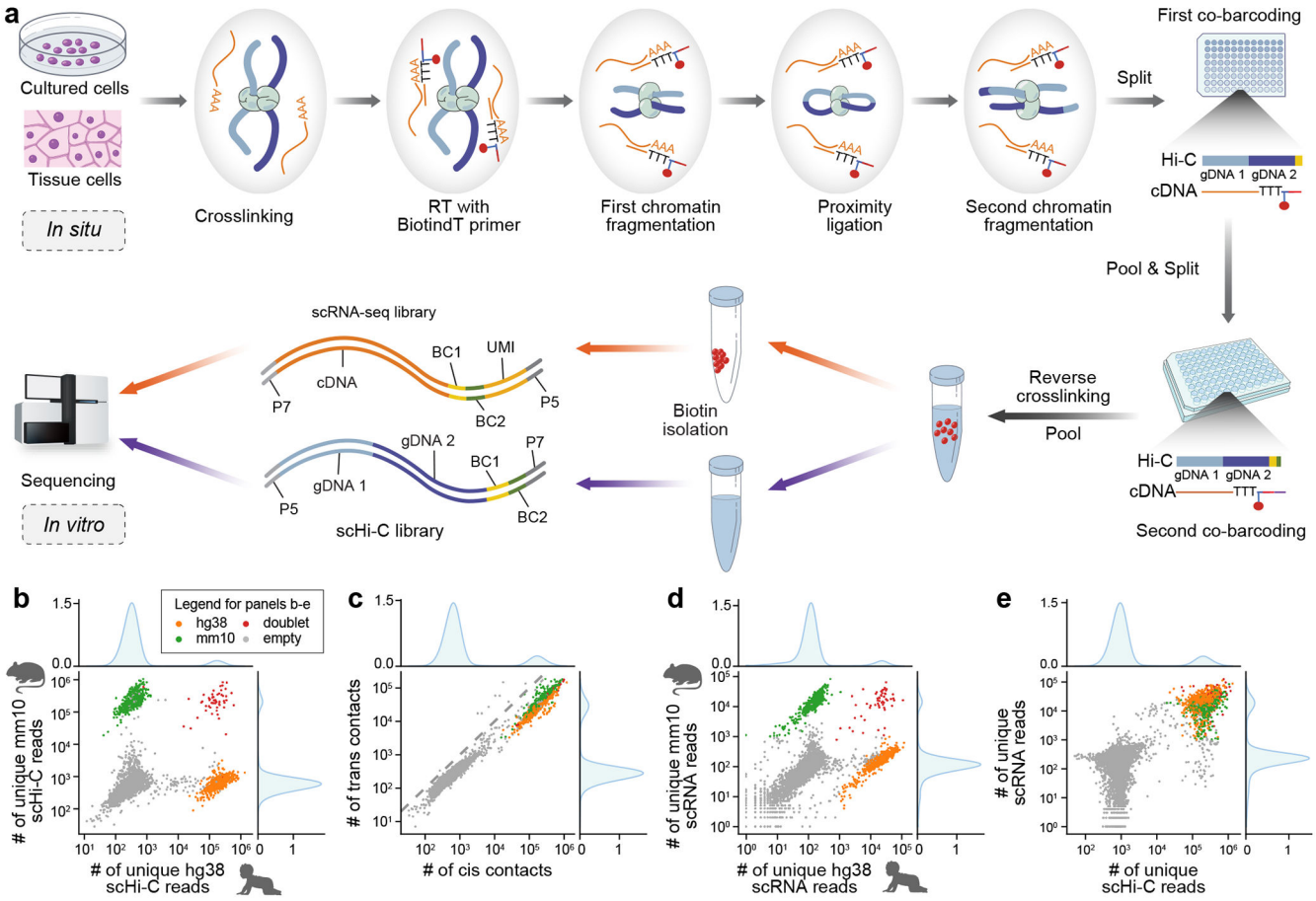

Figure 1. Overview and validation of GAGE-seq.

a. Schematic representation of the GAGE-seq workflow detailing the simultaneous single-cell profiling of 3D genome architecture and gene expression. **b-e**. Validations demonstrating the specificity of GAGE-seq using mixed experiments with the human (K562) and mouse (NIH3T3). **b** and **d**. Scatter plots showing the collision level in the GAGE-seq scHi-C (**b**) and scRNA-seq (**d**) libraries, and histograms showing the binomial distribution of reads mapped to hg38 (top) and mm10 (right). **c**. Scatter plot showing the cis:trans ratio of scHi-C reads. **e**. Scatter plot showing the well-separation of scHi-C and scRNA reads of valid cellular indices from that of empty indices. Mouse is colored in green, human in orange, collisions in red, and empty indices in gray.

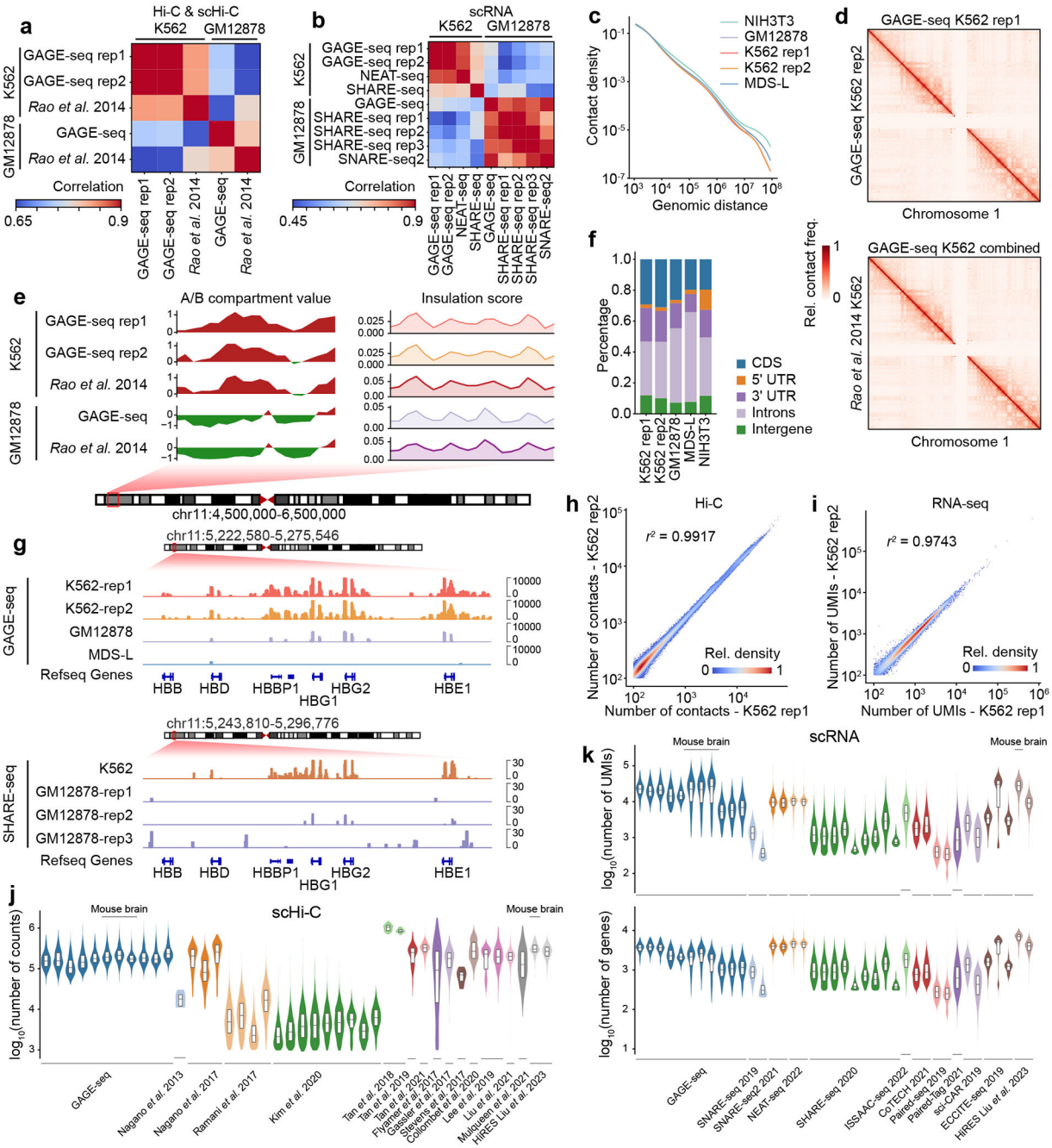

Figure 2. High-quality scHi-C and scRNA-seq data generated by GAGE-seq.

a. Pearson's correlation between the aggregated scHi-C profiles from GAGE-seq replicates and the bulk *in situ* Hi-C data[3]. **b**. Comparison of aggregated scRNA-seq profiles of GAGE-seq replicates with NEAT-seq[55], SHARE-seq[43], and SNARE-seq2[56]. Pearson's correlation is shown. **c**. Decay curves of chromatin contact for the GAGE-seq scHi-C libraries. **d**. Comparison of aggregated contact maps between two GAGE-seq K562 replicates (upper), and between the combined GAGE-seq K562 library and an *in situ* Hi-C library[3] (lower). **e**. Comparison of A/B compartments and TAD-like domain calling at the human beta-globin

locus between GAGE-seq (pseudo bulk) and *in situ* Hi-C[3]. **f**. RNA read distribution across gene bodies in the GAGE-seq scRNA libraries. **g**. Aggregated single-cell gene expression profiles at the GAPDH locus. Upper panel: scRNA-seq signals of GAGE-seq libraries of K562, GM12878, and MDS-L cells (hg38). Lower panel: scRNA-seq signals of SHARE-seq in GM12878 cells (hg19)[43]. **h.** Reproducibility between two biological replicates of GAGE-seq scHi-C libraries. **i**. Reproducibility between two biological replicates of GAGE-seq scRNA libraries. r^2 statistics are shown. **j**. Comparison of GAGE-seq scHi-C library size with published scHi-C[17,22–27,37,57–60] and co-assay methods[35,44,45]. **k**. Comparison of scRNA-seq library size (upper) and the number of detected genes (lower) with published co-assay methods[35,43,52,55,56,61–66].

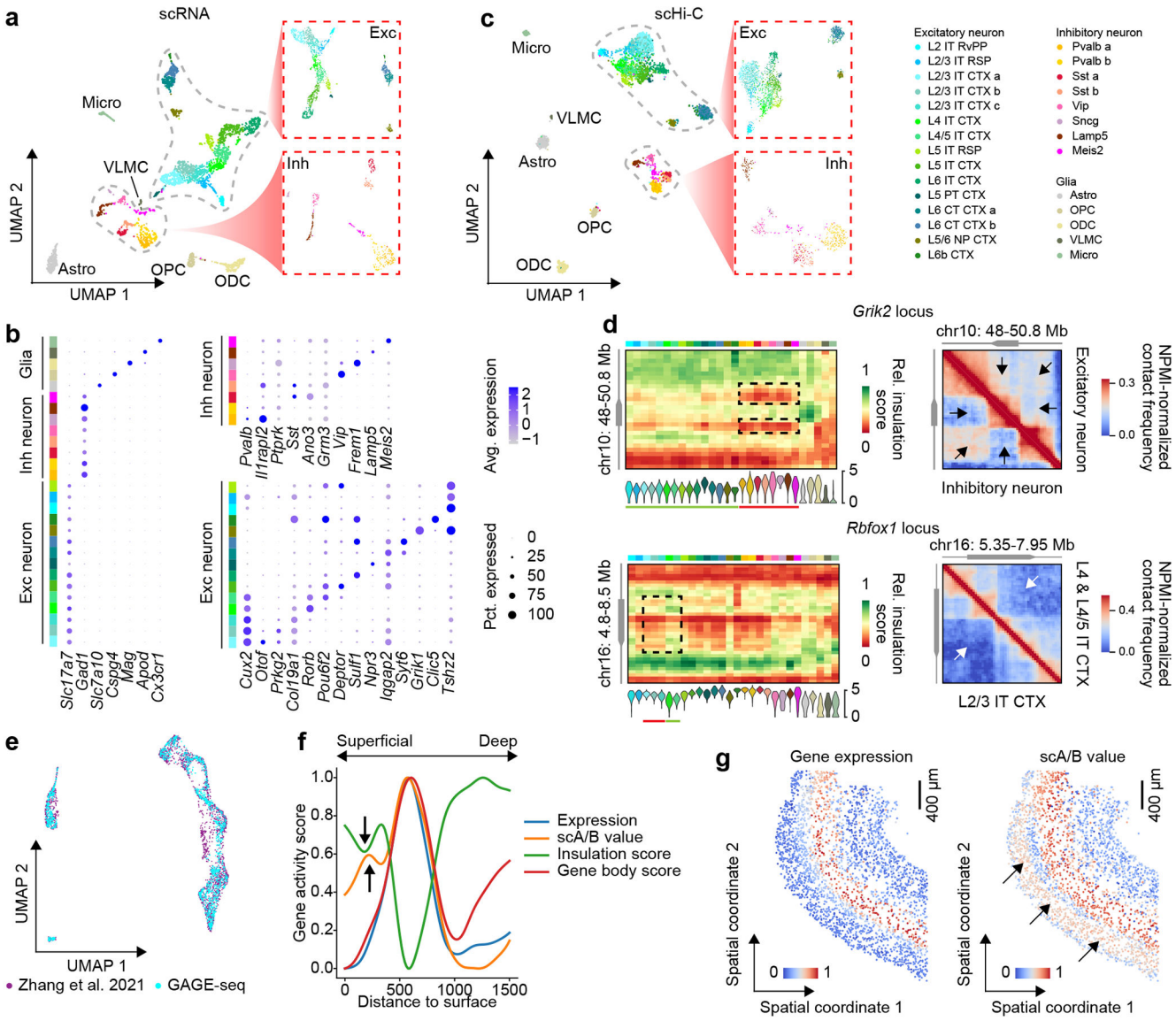

Figure 3. Cell types in mouse cortex characterized by GAGE-seq scHi-C and scRNA-seq.

a and **c**. UMAP visualization of mouse cortex scRNA-seq (a) and scHi-C profiles (c) from GAGE-seq. Insets: UMAP visualization of excitatory neuron subtypes (top) and inhibitory neuron subtypes (bottom). **b**. Cell type-specific expression (based on scRNA-seq in GAGE-seq) of known marker genes, including glial types, neuronal types, and neuron subtypes. **d**. Visualization of cell type-specific 3D chromatin architecture and gene expression at representative gene loci. Left: aggregated single-cell insulation score (100-Kb resolution, upper) and gene expression (lower) at the *Girk2* locus and the *Rbfox1* locus. Right: aggregated contact maps (50-Kb resolution) of the *Girk2* locus (top panel, excitatory vs inhibitory neurons) and the *Rbfox1* locus (low panel, L4 & L4/5 IT CTX vs L2/3 CTX). Cell types selected in the right panels are highlighted by green lines (higher expression) or red lines (lower expression) in the corresponding left panels. **e.** UMAP visualization of the integration of GAGE-seq and a MERFISH dataset[49]. **f**. Inferred spatial patterns of gene

expression and 3D genome features of L5 IT CTX marker genes. **g**. *In situ* plots of inferred single-cell gene expression (left) and scA/B value (right) for L5 IT CTX marker genes. Layer 3 was highlighted by black arrows in panels (f) and (g). The cell type abbreviations are based on the naming convention used in[67].

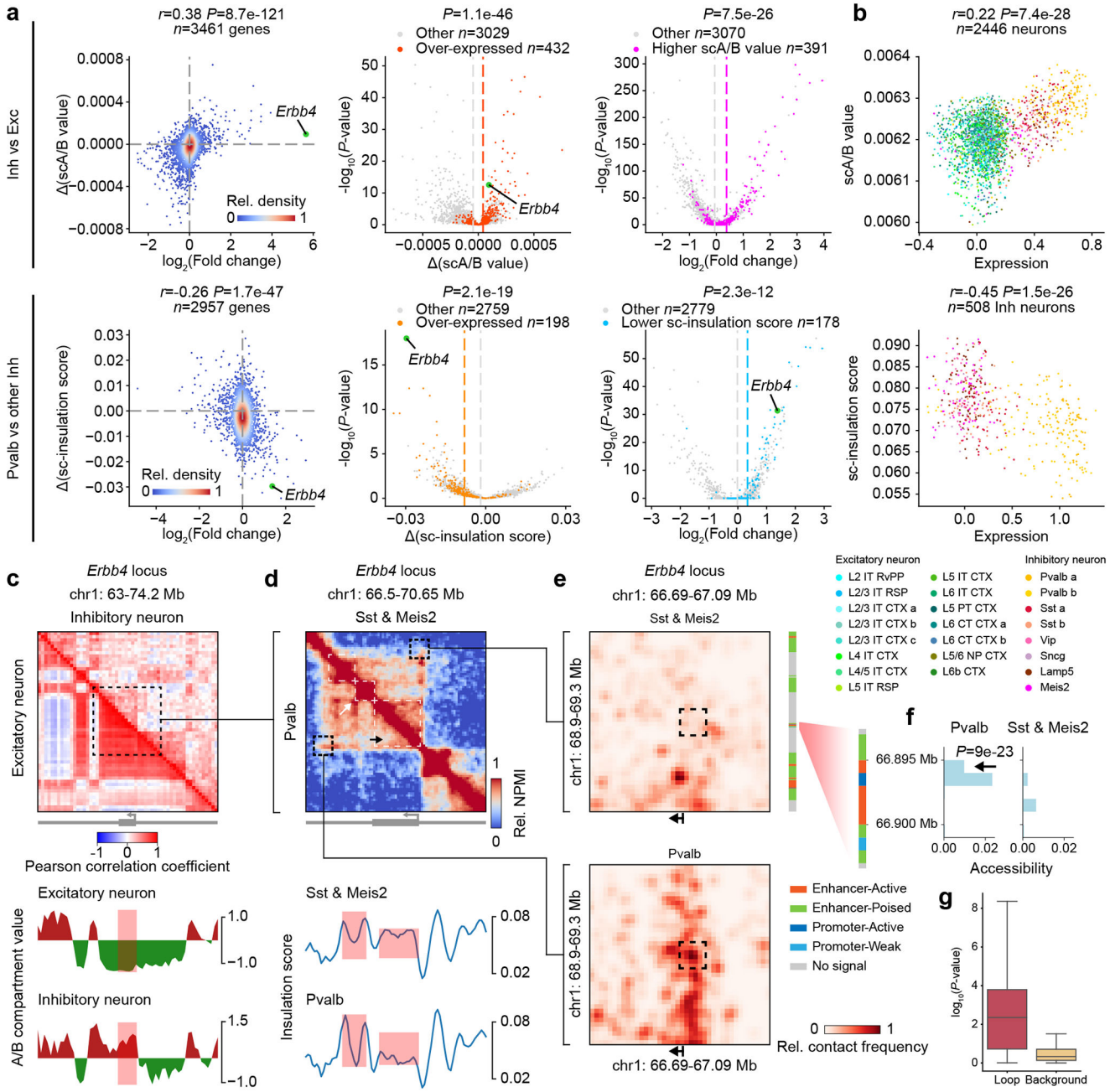

Figure 4. 3D genome features inform cell type-specific gene expressions in the mouse cortex.

a. Correlations between gene expression and 3D genome features across neuron cell types. Upper row: inhibitory (n=508) vs. excitatory (n=1938). Lower row: Pvalb (n=188) vs. other inhibitory (n=320). Left column: correlation between differential expression and differential 3D genome feature (Pearson's correlation coefficients and the *P*-values from one-sided tests for nonzero correlations shown). Middle column: volcano plot of differential scA/B value and single-cell insulation score; Right column: volcano plot of differential expression. *P*-values from one-sided t-tests with unequal variance are shown in middle and right columns. **b**. Single-cell level correlation of gene expression with scA/B value

(upper) or insulation score (lower) in inhibitory neurons (432 genes) and Pvalb (198 genes), respectively (Spearman's correlation coefficients and the *P*-values from one-sided tests for nonzero correlations shown). **c**. Comparison of A/B compartment (200-Kb resolution) of the *Erbb4* locus between inhibitory and excitatory neurons. Pearson's correlation matrices of aggregated contact maps (top) and the A/B compartment scoretracks (bottom) are shown. **d**. Comparison of the pseudo-bulk contact map (50-Kb resolution) of the *Erbb4* locus between Pvalb and other inhibitory subtypes. Pseudo-bulk contact maps (upper) and the insulation scores (bottom) are displayed. Two Pvalb-specific strides (white arrow) and melted TAD (black arrow) are shown in the top panel. The gene body is shown right under the contact matrices in (c) and (d), while the bottom panels highlight differential 3D genome features with light red boxes. **e**. Loop example in Pvalb (lower) and Sst and Meis2 (upper) inhibitory subtypes at 10-Kb resolution. Aggregated contact maps, regulatory element annotations[52] (right), and TSS of *Erbb4* (bittin arrow) are shown. **f**. Differential accessibility around the enhancer in Pvalb (left) vs. Sst and Meis2 (right), with a 1kb enhancer region highlighted (black arrow). The P-values of one-sided Mann-Whitney U tests are shown. **g**. Loop vs. non-loop contacts correlation with expression. *P*-values from two-sided tests for nonzero Spearman's correlation coefficients are shown (n=3,105 cells).

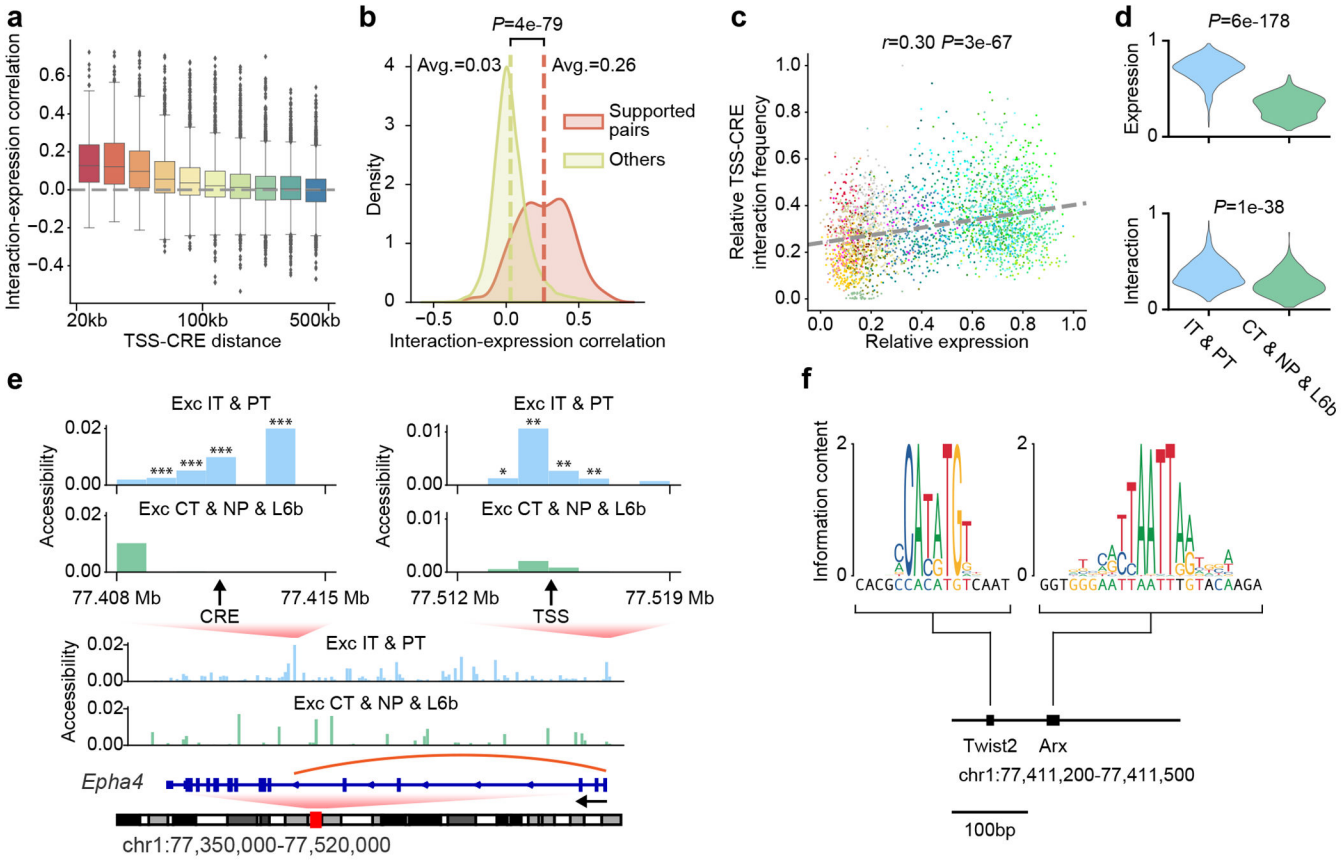

Figure 5. Integrative analysis of GAGE-seq and chromatin accessibility in the mouse cortex.

a. Correlation coefficient (n=3,105 cells) between expression and TSS-CRE interaction frequency for each gene-CRE pairs from Paired-seq data[63], grouped by genomic distance between TSS and CRE. **b**. Comparison between gene-CRE pairs corroborated by other sources (red) and those identified only from Paired-seq data[63] (yellow). The P-value of two-sided Mann-Whitney U test is shown. **c-e**. The combined effect of 3D genome and accessibility on expression at the *Epha4* locus. **c**. Correlation of interaction-expression for a specific gene-CRE pair at the *Epha4* gene, with dots representing single cells colored by cell type. **d**. Expression (upper) and TSS-CRE interaction frequency (lower) comparison among excitatory subtypes, revealing heightened levels in IT and PT subtypes. The P-values of one-sided Mann-Whitney U tests are shown. **e**. Accessibility comparison around the TSS and CRE (chr1: 77410959-77411960) of the *Epha4* gene among excitatory subtypes, showing higher accessibility IT and PT subtypes. The P-values of two-sided Mann-Whitney U tests are shown. IT and PT subtypes are compared against CT, NP, and L6b subtypes in **(d)** and **(e)**. In panel **(e)**, *: P<1e-3; **: P<1e-5; ***: P<1e-10; the P-values in the upper left plot are (from left to right): 2e-11, 7e-20, 8e-34, 7e-52; the P-values in the upper right plot are: 6e-4, 6e-8, 7e-6, 2e-6, 1e-4. **f**. Binding sites of transcription factors Twist2 and Arx at the CRE of the *Epha4* gene, depicting both the canonical motif (top) and the identified binding motif sequence (bottom) for each TF.

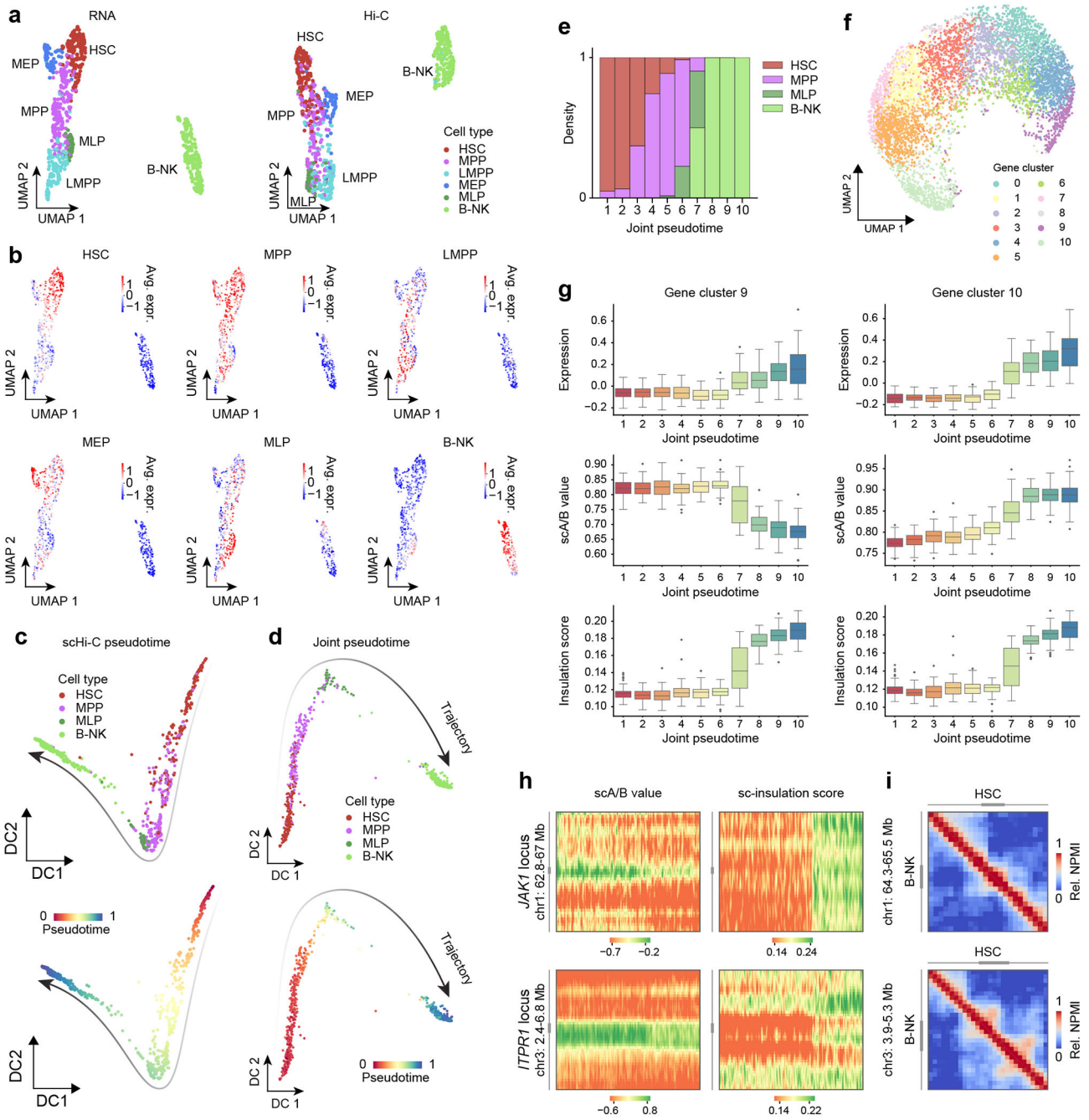

Figure 6. Interplay between 3D genome variation and gene expression changes in human bone marrow differentiation.

a. UMAP visualization of GAGE-seq scRNA-seq (left) and scHi-C profiles (right) of human bone marrow CD34+ cells. **b**. Average expression of known marker genes on the UMAP plot. The 6 panels include n=124, 78, 24, 82, 126, and 356 genes for HSC, MPP, LMPP, MEP, MLP, and B-NK, respectively. **c**-**d**. Inferred B-NK lineage trajectory and pseudotime from scHi-C profiles (c) and jointly from scRNA-seq and scHi-C profiles (d), displayed by cell type (upper) and pseudotime (lower). **e**. Cell type compositions across 10 equally

divided pseudotime bins. **f**. UMAP visualization of gene clusters determined by the temporal trend of expression and scA/B value. **g**. Temporal trends of gene expression (upper row), scA/B value (middle row), and single-cell insulation score (lower row) of gene clusters 9 (left column) and 10 (right column). **h**. scA/B (left) and single-cell insulation score (right) of the *JAK1* (upper) and *ITPR1* (lower) loci (at 100-Kb resolution). Each row represents a cell, ordered by the joint pseudotime from left to right. Heat maps were smoothed by a Gaussian kernel with a receptive field of 10 neighboring cells and 1 neighboring bin in each direction. **i**. Pseudo-bulk contact maps (at 50-Kb resolution) of HSC and B-NK at the *JAK1* (upper) and *ITPR1* (lower) loci.