Simultaneous profiling of chromatin architecture and transcription in single cells

Check for updates

Jiale Qu[1,2,3,7], Jun Sun[1,2,3,7], Cai Zhao[1,2,3], Xinyi Liu[1,2,3], Xinyao Zhang[1,2,3], Shaoshuai Jiang[1,2,3], Chao Wei[1,2,3], Haopeng Yu[4], Xiaoxi Zeng [4] ✉, Lili Fan [5] ✉ & Junjun Ding [1,2,3,4,6] ✉

The three-dimensional structure of chromatin plays a crucial role in development and disease, both of which are associated with transcriptional changes. However, given the heterogeneity in single-cell chromatin architecture and transcription, the regulatory relationship between the three-dimensional chromatin structure and gene expression is difficult to explain based on bulk cell populations. Here we develop a single-cell, multimodal, omics method allowing the simultaneous detection of chromatin architecture and messenger RNA expression by sequencing (single-cell transcriptome sequencing (scCARE-seq)). Applying scCARE-seq to examine chromatin architecture and transcription from 2i to serum single mouse embryonic stem cells, we observe improved separation of cell clusters compared with single-cell chromatin conformation capture. In addition, after defining the cell-cycle phase of each cell through chromatin architecture extracted by scCARE-seq, we find that periodic changes in chromatin architecture occur in parallel with transcription during the cell cycle. These findings highlight the potential of scCARE-seq to facilitate comprehensive analyses that may boost our understanding of chromatin architecture and transcription in the same single cell.

Dynamic folding of the chromatin architecture has essential biological functions in eukaryotes[1–6]. At the large scale, chromatin is divided into A and B compartments, which are enriched for transcriptionally active and inactive regions, respectively[7,8]. However, current approaches to capture the three-dimensional (3D) chromatin structure[7,9–12] and transcription[13–16] are mainly based on millions of cells methodology. As a result, the characteristics of each cell are concealed and the heterogeneity of cells makes it difficult to explain their relationship[17,18]. To overcome this problem, previous studies developed many single-cell unimodal omics methods, including single-cell chromatin conformation capture (scHi-C)[19–24] and single-cell transcriptome sequencing (scRNA-seq)[25–27]. Although these strategies are able to capture information from unpaired single cells, it is still not possible to explore the regulatory relationship between chromatin architecture and gene expression in the same single cell.

Recently, two strategies have been developed to link chromatin architecture and transcription. First are the computational strategies[28,29], which can infer cross-talk between the 3D chromatin structure and expression at the single-cell level using different models and algorithms. These strategies rely on simulations and correlations based on

[1]RNA Biomedical Institute, Sun Yat-sen Memorial Hospital, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China. [2]Advanced Medical Technology Center, The First Affiliated Hospital, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China. [3]Center for Stem Cell Biology and Tissue Engineering, Key Laboratory for Stem Cells and Tissue Engineering, Ministry of Education, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China. [4]West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu, China. [5]Guangzhou Key Laboratory of Formula-Pattern of Traditional Chinese Medicine, School of Traditional Chinese Medicine, Jinan University, Guangzhou, China. [6]Department of Histology and Embryology, School of Basic Medical Sciences, Guangzhou Medical University, Guangzhou, China. [7]These authors contributed equally: Jiale Qu, Jun Sun. ✉e-mail: zengxiaoxi@wchscu.cn; fanlili@jnu.edu.cn; dingjunj@mail.sysu.edu.cn
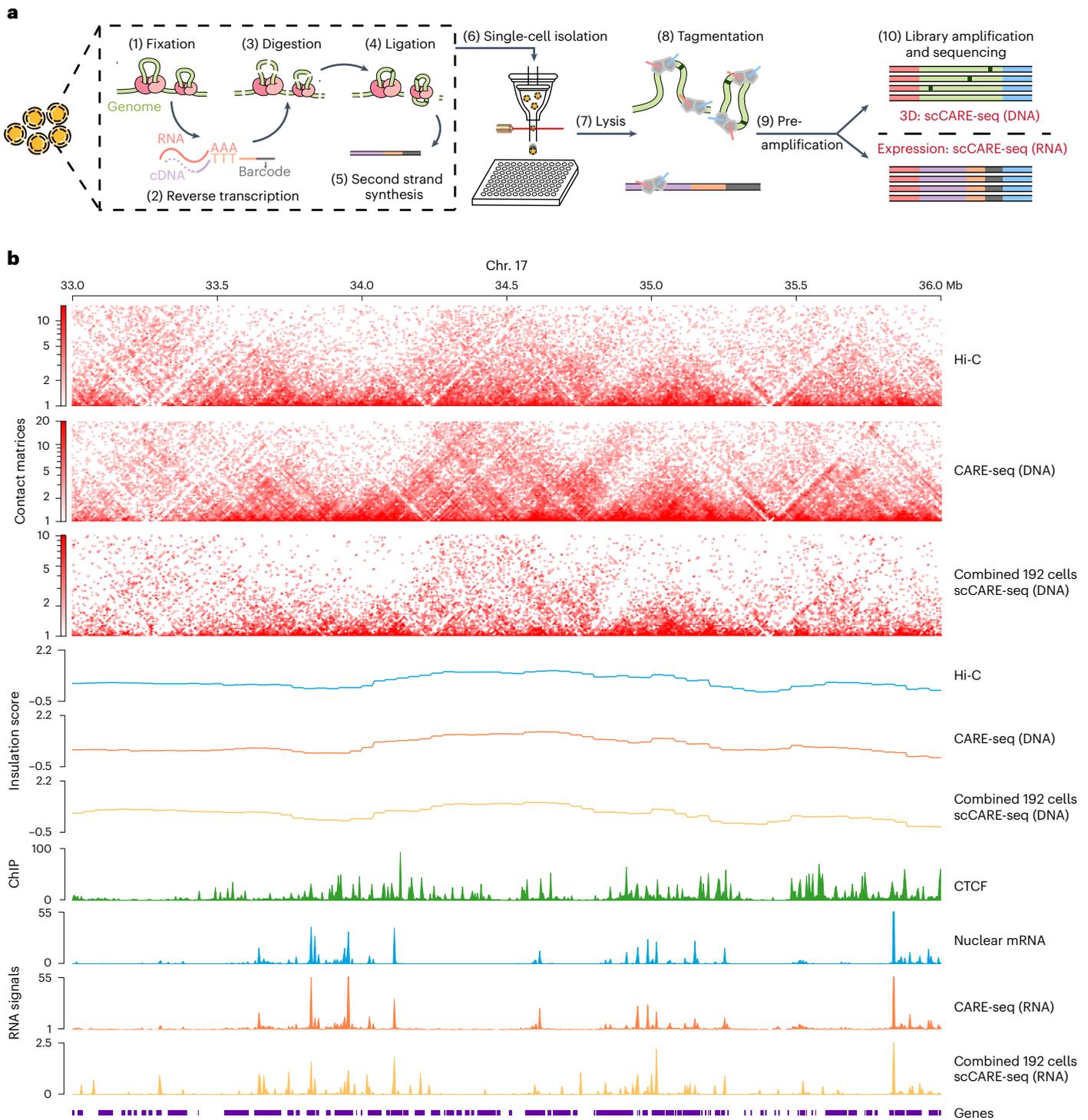
Fig. 1 | Overview of scCARE-seq. a, scCARE-seq workflow for measuring scHi-C and scRNA in the same single cell. mRNA reverse transcription, genome digestion and ligation are performed in intact nuclei. After flow cytometry sorting an individual cell into each well, single cells were lysed. Tn5 was then added to each well for tagmentation. Finally, following preamplification of each cell library, each well was divided into two halves for DNA and RNA library amplification, using a pair of common adapter primers and RNA index primers, respectively. **b**, Track view displaying both contact matrices, insulation score, CTCF ChIP signal and RNA signals from 33–36 Mb of chromosome 17 (Chr. 17).

information from unpaired cells, which cannot be compared with the actual relationship. Second, microscopy-based approaches[30–33] can directly reveal spatial information for the genome and transcripts in single cells. However, these methods rely on advanced optics equipment that limits their widespread application. As a complementary technology to above methods, single-cell multiomics based on next-generation

sequencing (NGS) is a promising approach that can generate integrated information from the same single cell and does not require special equipment, but research is still lacking.

Here, we developed an NGS-based method for the simultaneous detection of chromatin architecture and mRNA expression in single cells (scCARE-seq) by integrating Hi-C[24] and nuclei RNA-sequencing

(RNA-seq)[34] in the same cell. To demonstrate the utility of scCARE-seq, we used it to generate joint profiles of chromatin architecture and transcription from single mouse embryonic stem cells (mESCs) and to study their correlation during cell fate transition and the cell cycle. This new approach provides improved cell cluster separation compared with scHi-C and more perspectives on the regulatory programs underlying the connections between chromatin architecture and gene expression in diverse biological processes.

Results

Joint profiling of 3D genome and expression in single cells

The scCARE-seq approach integrates Hi-C and RNA-seq at the single cell level to evaluate the correlation between 3D chromatin structure and transcription in the same single cell (Fig. 1a, Extended Data Fig. 1a, Supplementary Table 1 and Methods). Briefly, mRNA reverse transcription, genome digestion and ligation were performed in intact nuclei, and individual cells were sorted by flow cytometry into separate wells and lysed. After adding Tn5 transposase to each well for tagmentation, each cell library was preamplified and divided into two halves for DNA and RNA library amplification using their respective primer pairs. In addition, in accordance with a previously published diluted strategy to reduce frequency of nuclei multiplets to 7.4%[35], we utilized cross-linking in a preparation diluted to 1×10^5 nuclei per milliliter and selected nuclei using fluorescence-activated cell sorting (FACS). To assess the ability of scCARE-seq to distinguish single cells, we designed a mixed-species experiment using HEK293T and mESCs. Our results revealed that the proportion of interspecies nuclei multiplets was 2.56% for both DNA and RNA partitions in the scCARE-seq datasets (Extended Data Fig. 1b). For simultaneous detection of chromatin architecture and mRNA expression in bulk cells (CARE-seq), oligo-biotin-dT was used for reverse transcription and Tn5 tagmentation was performed after cell lysis without single-cell sorting. The libraries were constructed after biotin enrichment.

To demonstrate whether both bulk and single-cell CARE-seq have the ability to profile similar patterns as the previous single-omics protocols, we adopted Hi-C methods[9,10] and nuclei RNA-seq methods[34] as representative methodologies for comparison. Examination of specific chromatin regions (Fig. 1b and Extended Data Fig. 2) indicated that the chromatin architecture and transcription detected by CARE-seq at both the bulk and single-cell level exhibited patterns similar to that of typical single-omics methods. Collectively, the results roughly demonstrated that our methods are capable of simultaneously capturing valuable information about 3D chromatin structure and transcription.

Simultaneous capture of 3D genome and expression in mESCs

To further evaluate the reliability and validity of CARE-seq in a same-cell population, we examined the two aspects of chromatin architecture and RNA expression at the bulk level. Regarding partition of the 3D chromatin structure, we demonstrated that CARE-seq was reproducible and captured similar global contact information as the Hi-C method[9,10] (Pearson r, 0.96; Spearman ρ, 0.96; stratum-adjusted correlation coefficient (SCC)[36], 0.98) (Fig. 2a and Extended Data Fig. 3b). In addition, we observed a high degree of correlation at the compartment and topologically associated domain (TAD) levels (Fig. 2b and Extended

Data Fig. 3a,c,d). Similar interaction patterns were observed from a large-scale 1 Mb resolution to a fine-scale 10 kb resolution, and were quantified by SCC (Fig. 2c). The above results show that CARE-seq is comparable with single-omics Hi-C technology.

Regarding the expression partition, CARE-seq predominantly captured mRNA information from nuclei. Thus, we first demonstrated the relationship between the two typical methods and found that nuclei RNA-seq could represent cell RNA-seq (Fig. 2d), as confirmed by previous research[34,37,38]. Next, global analysis revealed a high correlation between the gene expression signals of CARE-seq and typical RNA-seq (Spearman ρ, 0.85; Fig. 2d and Extended Data Fig. 3g) and a high level of reproducibility (Spearman ρ, 0.92; Fig. 2d and Extended Data Fig. 3e,f). In addition, to determine whether the two methods can identity similar genes with different expression levels, we ranked genes based on their expression, and calculated the Jaccard similarities between different quantiles of the two ranked gene sets. The high similarity coefficient suggested that CARE-seq reproduces similar genes to the traditional single-omics nuclei RNA-seq method at different expression levels (Fig. 2e). In particular, we observed a strong concordance of the expressed profiles for the pluripotency genes *Nanog* and *Sox2*, and other expressed genes in mESCs (Fig. 2f and Extended Data Fig. 4). Therefore, CARE-seq is comparable with typical RNA-seq for the detection of transcription.

Altogether, from systematic comparisons with typical single-omics methodologies, we illustrated that CARE-seq can simultaneously and accurately capture chromatin architecture and mRNA expression in the same cell population.

The high-quality 3D genome and expression from scCARE-seq

We next systematically compared scCARE-seq data with other published methods to confirm the quality of DNA and RNA (Supplementary Table 2 and Extended Data Fig. 5a). Single-cell level validation was based on the number of contacts and *cis*-to-*trans* (intrachromosomal contacts to interchromosomal contacts) ratio in the 3D chromatin structure, and the numbers of expressed genes and unique molecular identifiers (UMIs) in transcription. Bulk-level validation is based on the correlation coefficient between merged scCARE-seq data and typical bulk methods.

To show the data quality of the 3D chromatin structure at the single-cell level, we confirmed that the distribution of valid pairs and *cis*-to-*trans* ratio of scCARE-seq were analogous to other published scHi-C methods[19–23,39–41] (Fig. 3a and Extended Data Fig. 5b). We also calculated *cis* long-range interactions (>20 kb) for each cell in the scCARE-seq data (median 52.04%; Extended Data Fig. 6a). Bulk-level data also showed a high degree of correlation between scCARE-seq and other methods (Fig. 3c and Extended Data Fig. 6b), including Hi-C (Pearson r, 0.92; Spearman ρ, 0.92; SCC, 0.91) and CARE-seq (Pearson r, 0.94; Spearman ρ, 0.94; SCC, 0.94). Similar patterns in the contact heatmap further revealed that scCARE-seq can reproduce chromatin structure information at the bulk level (Fig. 3d and Extended Data Fig. 6c). In addition, we calculated the relative contact probability for the genome distance of single cells, and observed heterogeneity among individual cells (Extended Data Fig. 6d). Although the decay of biotin-based CARE-seq was smaller than that of the nonbiotin scCARE-seq, the distribution pattern of merged single cells was still

---

Fig. 2 | CARE-seq provides an accurate method to simultaneously captures chromatin architecture and transcriptome in mESCs. a, Scatter plots show high correlation between contacts from CARE-seq versus Hi-C (left) and two technical replicates (right). Pairwise correlations between Hi-C matrices data of 500 kb were computed. The *x* and *y* axes show mean log1p values for the matrix. log1p means ln(1+x). **b,** Scatter plots show the correlation between the compartment score (left) and insulation score (right) for CARE-seq (DNA) versus Hi-C. The compartment and insulation scores were computed by Hi-C matrices data of 100 kb and 40 kb, respectively. **c,** Comparison of CARE-seq and Hi-C in the contact matrices of chromosome 1 at: 1 Mb resolution; 10–100 Mb at 250 kb

resolution; 56–76 Mb at 50 kb resolution; 72–75 Mb at 10 kb resolution (left to right). Matrix similarity is evaluated by HiCRep at the corresponding resolution. **d,** Scatter plots show a high correlation for gene expression signals from total mRNA versus nuclear mRNA, nuclear mRNA versus CARE-seq (RNA) and two technical replicates (left to right). Pairwise correlations between genes are shown in counts per million (c.p.m.). The *x* and *y* axes show mean log(c.p.m.) values. **e,** Jaccard analysis indicating a high similarity coefficient in different percentages of top genes. **f,** Representative regions showing a consistent pattern of gene expression across datasets.

similar to that of the cell population (Extended Data Fig. 6d). The above results suggest that the 3D chromatin structure data obtained using scCARE-seq were reliable.

To evaluate the quality of gene expression data at the single-cell level, in terms of UMIs and gene numbers, we also confirmed that scCARE-seq was not inferior to scRNA-seq in different single-cell

multimodal omics studies[42–47] (Fig. 3b and Extended Data Figs. 5c and 6e). Bulk-level analysis indicated a high correlation between the combined transcriptional data of scCARE-seq (Fig. 3e) and that of nuclei RNA-seq (Pearson r, 0.84; Spearman ρ, 0.81) and CARE-seq (Pearson r, 0.86; Spearman ρ, 0.84). Specifically, the expression profile of *Klf4* and *Tet1* in each cell illustrated both the consistency
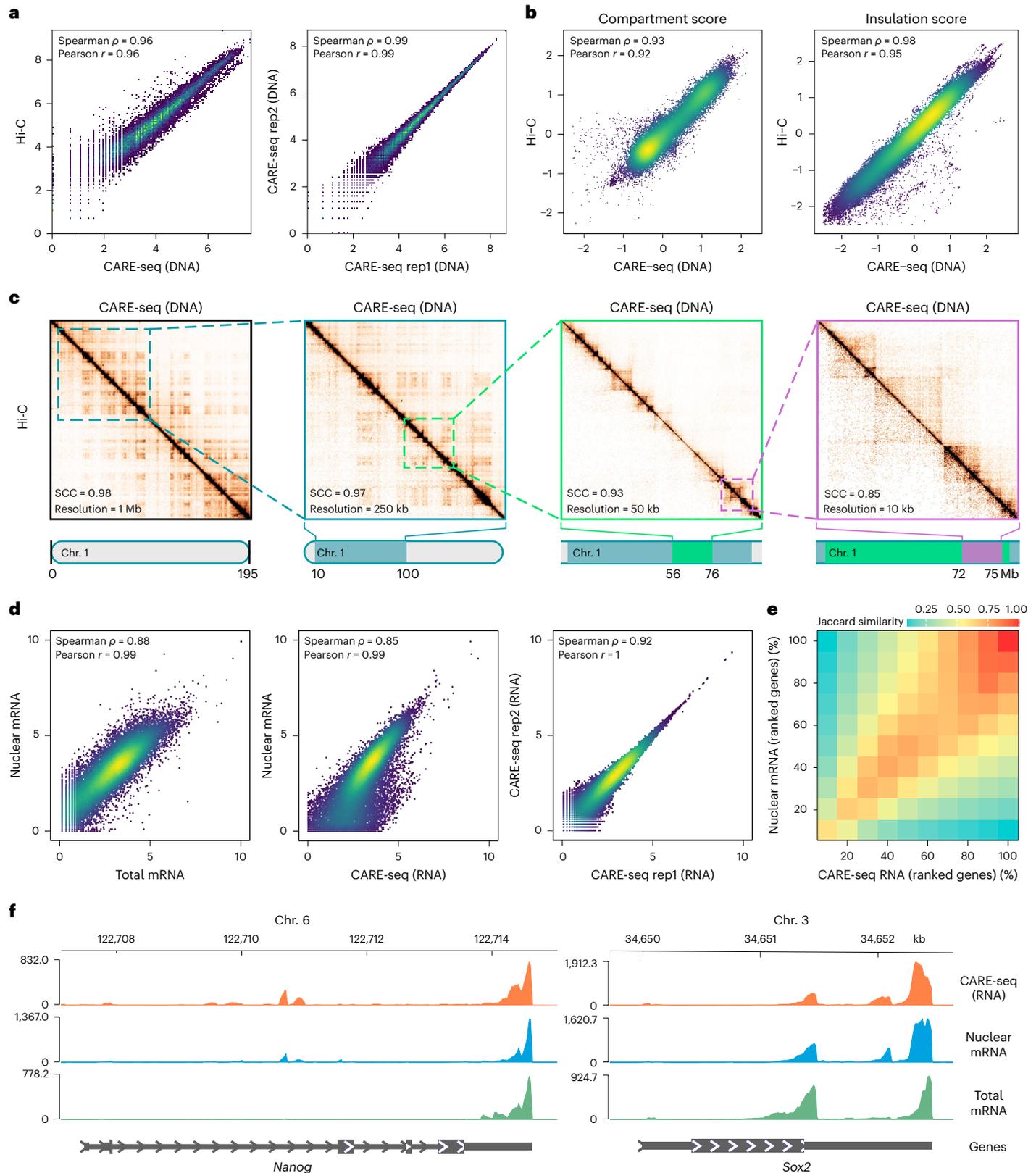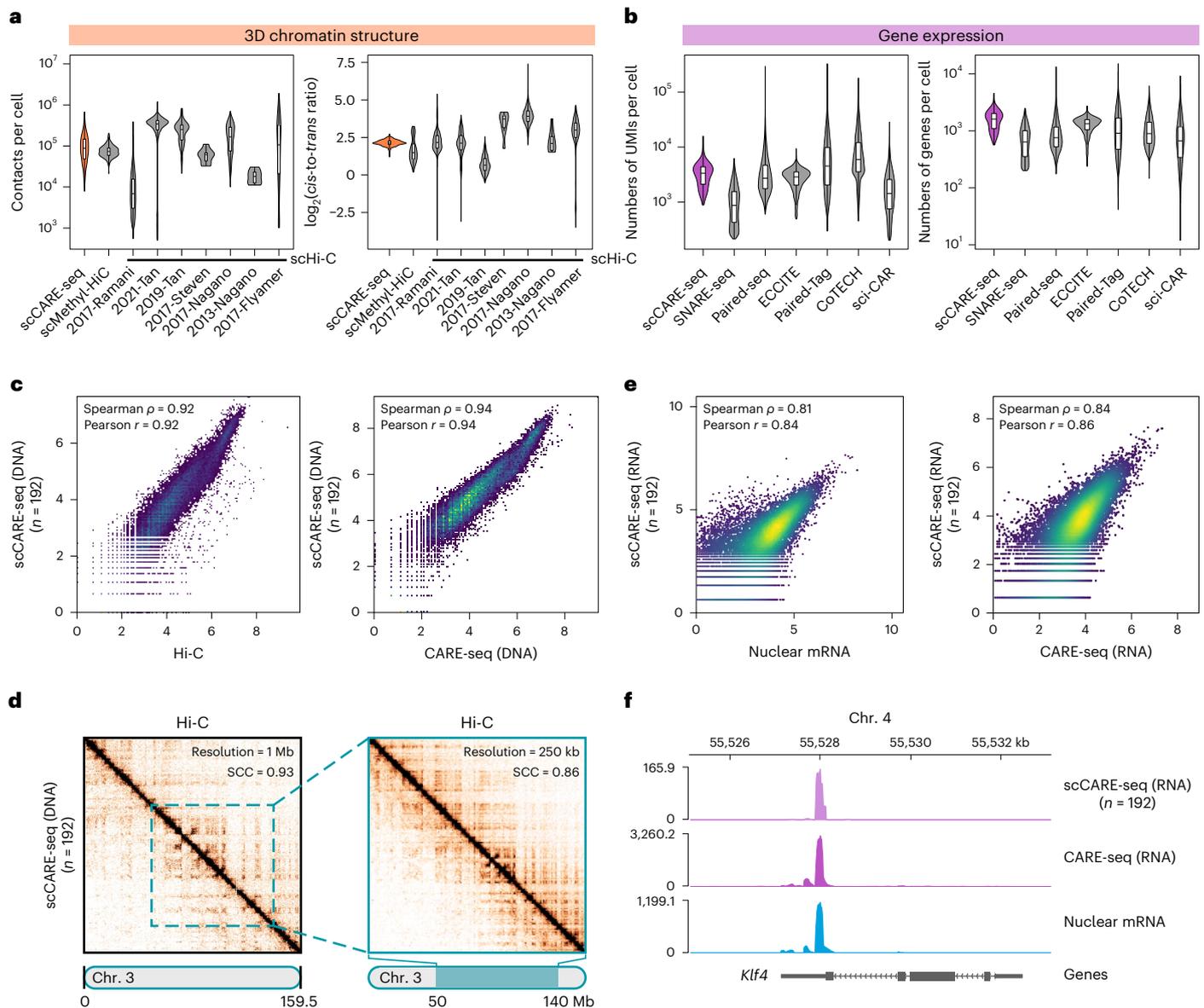
Fig. 3 | scCARE-seq simultaneously captures high-quality chromatin architecture and transcriptome data in the same single cell. a,b, Comparison of scCARE-seq with other single-cell methods. Plots show numbers of (**a**) contacts (left) and *cis*-to-*trans* ratio (right) (*n* = 384, 150, 6,998, 4,272, 409, 8, 4,098, 10 and 242 cells) and (**b**) UMIs (left) and expressed genes (right) (*n* = 192, 16,437, 14,095, 2,946, 50,754, 9,277 and 21,859 cells). Boxplots were drawn from the lower quartile (Q1) to the upper quartile (Q3), with the mid line denoting the median and whiskers a maximum of 1.5× interquartile range (IQR); outliers are not indicated. CoTECH, combined assay of transcriptome and enriched chromatin binding; ECCITE, expanded CRISPR-compatible cellular indexing of transcriptomes and epitopes by sequencing; sci-CAR, jointly profiles single-cell chromatin accessibility and mRNA; SNARE-seq, droplet-based single-nucleus chromatin accessibility and mRNA expression sequencing. **c,** Scatter plots

show the Spearman and Pearson correlation coefficients of contacts between merged scCARE-seq (*n* = 192) and bulk-level methods. Pairwise correlations were computed between Hi-C matrices data. The *x* and *y* axes show mean log1p values of the matrix. **d,** Comparison of contact heatmap of chromosome 3 between scCARE-seq and Hi-C, at 1 Mb resolution (left) and 50–140 Mb at 250 kb resolution (right). Matrix similarity is evaluated by HiCRep at the corresponding resolution. **e,** Scatter plots show the Spearman and Pearson correlation coefficients of gene expression between merged scCARE-seq (*n* = 192) and bulk-level methods. Pairwise correlations between genes are given in c.p.m. The *x* and *y* axes show mean log(c.p.m.) values. **f,** Representative region showing a consistent transcriptional profile for gene expression across datasets generated using scCARE-seq, CARE-seq and nuclei RNA-seq.

and heterogeneity of single-cell technology (Fig. 3f and Extended Data Fig. 6f).

Taken together, both single-cell and bulk-level validation proved that scCARE-seq can capture high-quality 3D chromatin structure and gene expression data in the same single cell.

The relationship between 3D genome and RNA by clustering
Previous studies revealed that 3D chromatin structure and transcription are altered during development or disease pathogenesis[4–6]. However,

the relationship between the 3D chromatin structure and transcription is difficult to interpret from bulk population and unpaired cell analysis because of heterogeneity among individual cells[17,18,28,29]. Here, we applied scCARE-seq to examine this relationship during the cell fate transition of 2i mESCs to serum mESCs. Single cells were clustered on the basis of their gene expression profiles (Fig. 4a), and consistent with previous reports, 2i and serum cells were clearly separated[41,48,49]. However, we did not observe distinct cell groups when these cells were clustered based on valid contacts (Fig. 4a). 3D chromatin structures
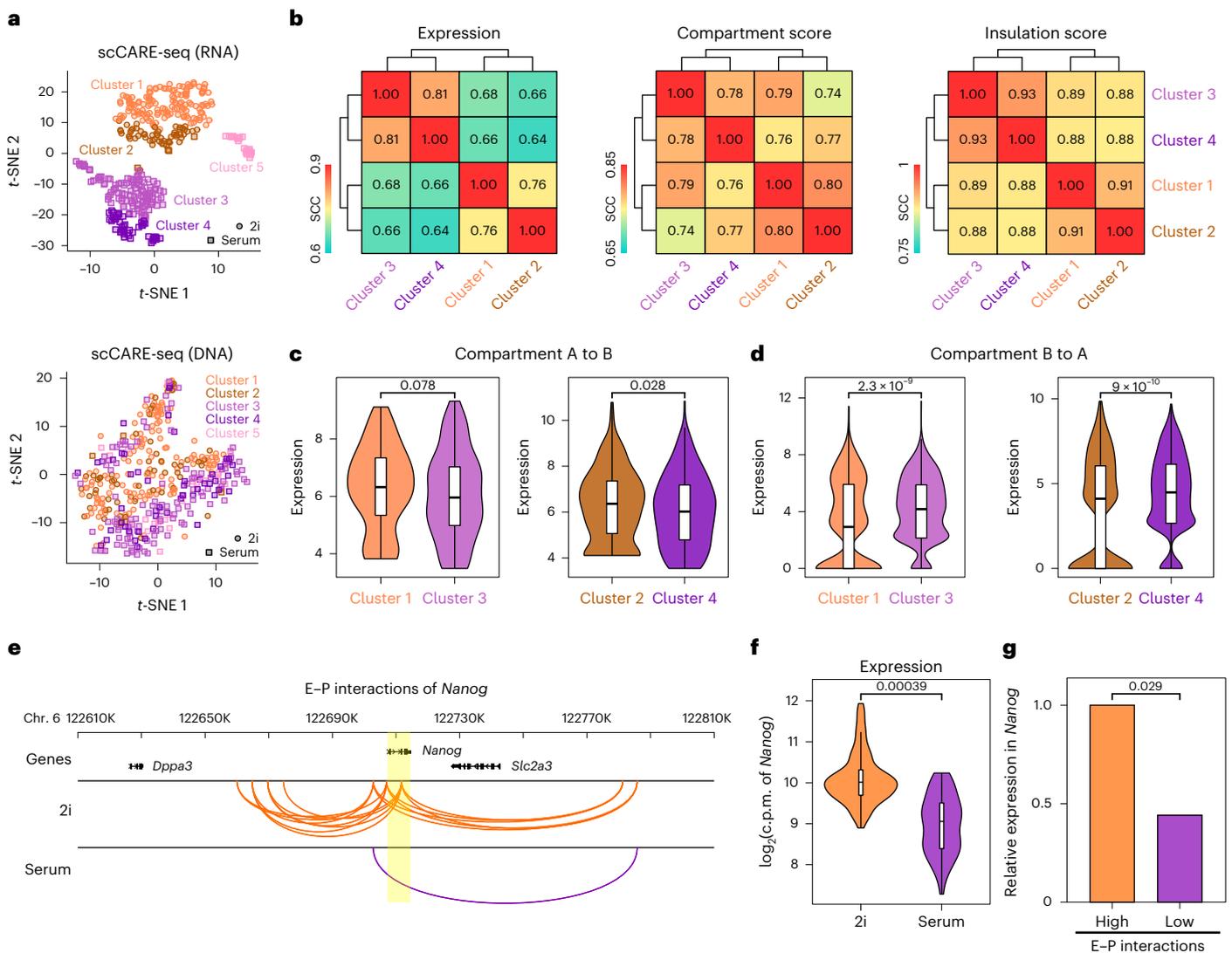
Fig. 4 | Transcriptional clustering reveals differences in 3D chromatin structure from 2i and serum single mESCs. a, Dimension reduction (*t*-SNE) visualization showing the clustering of single cells from scCARE-seq transcriptomic genes: c.p.m. values (upper) and scCARE-seq 3D chromatin structural valid pairs (lower). Each dot represents an individual cell and each color represents a cell cluster. **b**, Heatmaps showing the Spearman correlation coefficient of gene expression, compartment score and insulation score. **c**,**d**, Comparison of $\log_2$(c.p.m. + 1) values for genes in compartment A to B (**c**, $n = 114$ and 285, respectively) and compartment B to A (**d**, $n = 749$ and 1,363, respectively) with cluster 1 versus cluster 3 (left) and cluster 2 versus cluster 4 (right). *P* values determined by one-sided Wilcoxon signed-rank test:

0.078, 0.028, 2.3×10^{-9} and 9×10^{-10} from left to right. **e**, Interactions between enhancers and *Nanog* are represented by semi-circle connectors. **f**, Comparison of $\log_2$(c.p.m.) values of *Nanog* gene in each cell of the 2i condition ($n = 192$) and serum condition ($n = 192$). *P* value determined by one-sided Wilcoxon signed-rank test: 0.00039. Boxplots were drawn from the lower quartile (Q1) to the upper quartile (Q3), with the mid line denoting the median and whiskers indicating maximum 1.5× IQR; outliers are not indicated. **g**, Single cells were categorized into high and low groups based on the number of E–P interactions >1. The high group consisted of seventeen 2i mESCs with a higher number of E–P interactions, and the low group comprised twenty-eight 2i mESCs with a lower number of E–P interactions. *P* value, one-sided Wilcoxon signed-rank test: 0.029.

seem to have a weaker ability to phase cell types compared with transcription information. Therefore, we proposed that scCARE-seq can improve cell clustering in scHi-C.

To further demonstrate that scCARE-seq can better present subpopulation information of the 3D chromatin structure in single cells, we showed that the adjacent transcriptional clusters had a higher correlation coefficient at both the compartment and TAD levels (Fig. 4b). The chromatin contact matrices have a specific 3D chromatin structure in each cluster (Extended Data Fig. 7a,b). Moreover, we also observed that genes in compartment A exhibited higher activation than those in compartment B (Fig. 4c,d). We further explored whether the region of the compartment switch corresponded to changes in gene expression between clusters using published Hi-C data[50]. Our analysis

demonstrated that structural changes could indicate a corresponding increase or decrease in gene expression between clusters (Extended Data Fig. 7c,d). These results suggested that structural changes are also associated with changes in gene expression within single-cell data combined clusters. Because cluster 5 has an insufficient number of cells, we did not compare it with the others. These results revealed that scCARE-seq can highlight differences in the 3D chromatin structure of different clusters. In addition, we found that 2i mESCs had more enhancer–promoter (E–P) interactions for the *Nanog* gene than serum mESCs. Meanwhile, 2i mESCs also exhibited higher *Nanog* gene expression levels (Fig. 4e,f), which was consistent with previous findings[48]. To examine the relationship between the number of E–P interactions and the expression of *Nanog*, we utilized a sorting strategy based on
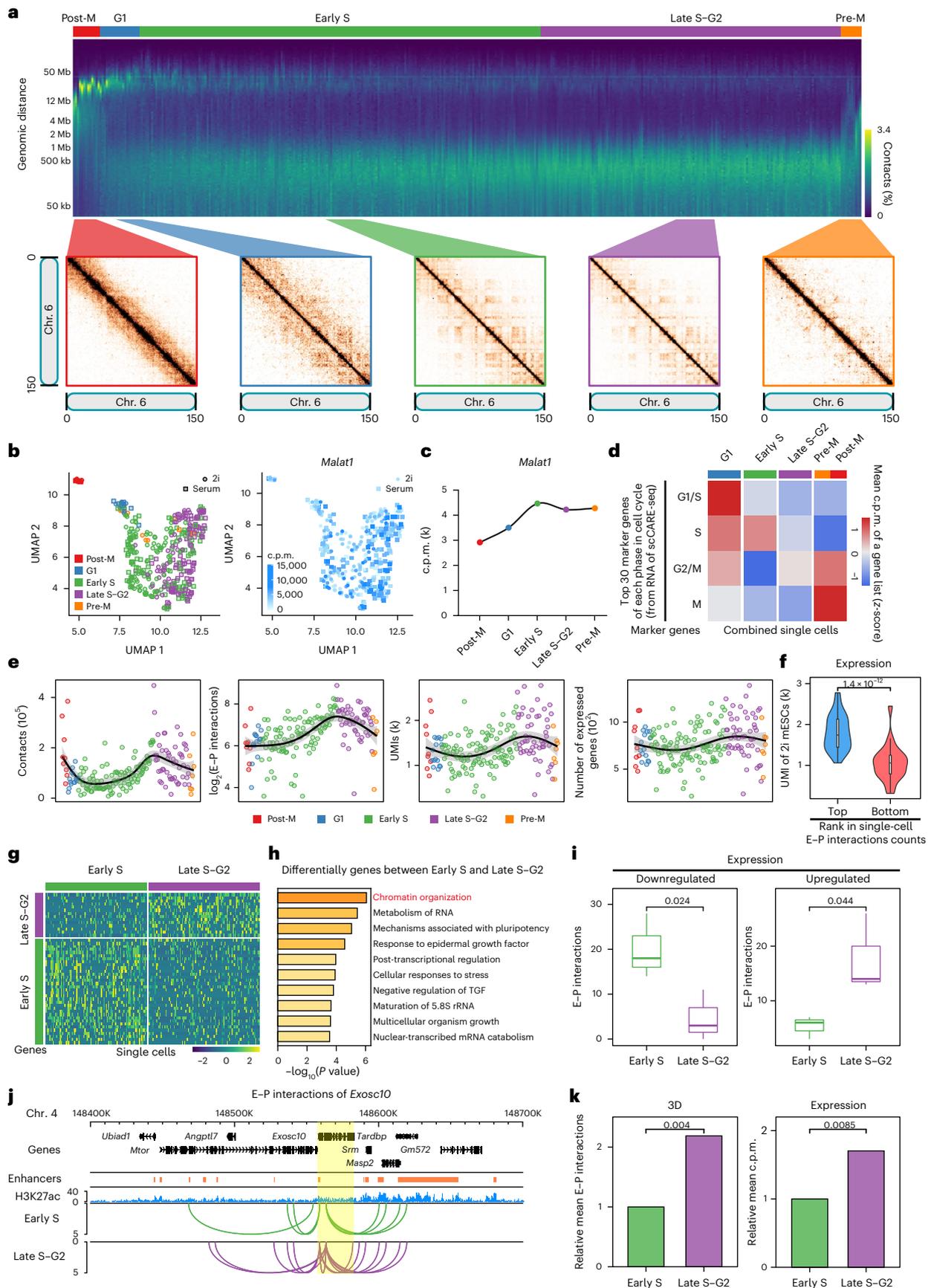
the E–P interaction count. Subsequently, we calculated the relative mean expression levels of *Nanog* in the two groups of single cells and found that single cells with more E–P interactions had higher expression levels of *Nanog* (Fig. 4g).

Overall, these results indicated that scCARE-seq can not only better distinguish subgroups of 3D chromatin structure, but also be used to study the relationship between 3D chromatin structure and gene expression from the same single-cell data.

Fig. 5 | 3D chromatin structure and expression have interrelated periodic changes in single cells during the cell cycle. a, Single-cell contact decay profiles ordered by in silico inferred cell-cycle phasing, with approximate cell-cycle phases shown at the top of the profiles. Each column represents a single cell (upper). Selected phased and pooled contact maps (lower). **b**, Uniform manifold approximation and projection embedding showing the cell-cycle phase of single cells from scCARE-seq and the *Malat1* gene expression level in each cell. Each dot represents an individual cell. **c**, Mean *Malat1* gene expression in the different cell-cycle phases. **d**, Mean gene expression of lists of cell-cycle specific marker genes averaged for each cell-cycle phase (*n* = 19, 195, 146 and 24 cells). **e**, Contacts, E–P interactions, UMIs and number of expressed genes per single 2i mESC. The black line represents the mean trend and shading represents confidence intervals of 0.95. **f**, Comparison of UMI in single cells of the upper and lower groups in 2i mESCs. Selection was based on the upper and lower groups of each of 48 single cells (25%) ranked by number of E–P interactions. Boxplots were drawn from the lower

quartile (Q1) to the upper quartile (Q3), with the mid line denoting the median and whiskers showing the maximum 1.5× IQR; outliers are not indicated. *P* value determined from one-sided Wilcoxon signed-rank test. **g**, Heatmap depicting RNA marker genes of Early S and Late S–G2 in serum mESCs. Columns represent single cells ordered by cell cycle, and rows represent different marker genes. **h**, The top GO terms were enriched using the marker genes in **g**. TGF, transforming growth factor; rRNA, ribosomal RNA. **i**, Comparison of E–P interactions for three upregulated and three downregulated marker genes between Early S and Late S–G2. Boxplots were drawn from the lower quartile (Q1) to the upper quartile (Q3), with the mid line denoting the median and whiskers showing the maximum 1.5× IQR; outliers are not indicated. *P* values were determined by one-sided *t*-test. **j**, E–P interactions of *Exosc10* represented by semicircular connectors. **k**, Comparison of relative mean E–P interactions (left) and relative mean expression levels (right) of *Exosc10* gene in single cells of Early S (*n* = 86) and Late S–G2 (*n* = 94). *P* value determined from one-sided Wilcoxon signed-rank test.

Periodic changes to 3D genome and expression in the cell cycle

Although the dynamics of the 3D chromatin structure and transcription during the cell cycle have been captured at the bulk and single-cell levels[20,51,52], their regulatory relationship is poorly understood. Therefore, we used scCARE-seq to study the correlation between chromatin architecture and transcriptional changes during the cell cycle. First, each cell was assigned to one of the five cell-cycle phases based on a previous strategy[20], which utilized the distribution of contact distance scales in single cells from the 3D chromatin structure of scCARE-seq data (Fig. 5a, Extended Data Fig. 8a and Supplementary Table 3). This was consistent with cell-cycle related changes in chromatin architecture reported previously[20]. Moreover, we clustered all the single cells based on the chromatin interaction of each cell. Next, we assigned cell-cycle phase information to these single cells and observed a continuous timescale for the cell cycle (Fig. 5b,c and Extended Data Fig. 8b,c), which suggested that clustering of the 3D chromatin structure is susceptible to the cell cycle. In addition, according to the cell-cycle timescale, we found that the number of contacts increased significantly in the S phase (Fig. 5e), which was attributed to DNA replication in this phase. We also observed a similar phenomenon[20]; namely, the fraction of interchromosomal contacts was lowest in the mitosis phase (Extended Data Fig. 8d), which might be related to the state transition from chromatin to chromosome. Therefore, we were able to use the chromatin architecture revealed by scCARE-seq to phase the cell cycle. The relationship between Hi-C-based cell-cycle phasing and gene expression was further confirmed by the high expression of similar marker genes (Fig. 5d and Supplementary Table 4).

To further explore the correlation between chromatin architecture and transcription during the cell cycle, we focused on the overall trend in their changes from scCARE-seq. The distribution of contacts and E–P interaction numbers show periodic changes and a maximum in the S phase (Fig. 5e). We speculate that the increase in E–P interactions is also related to DNA replication. Furthermore, we observed that the overall trends in E–P interactions and expression levels were similar during the cell cycle; UMIs and the number of expressed genes were also relatively higher in the S phase (Fig. 5e and Extended Data Fig. 8e). To evaluate the relationship between E–P interactions and UMI at a single-cell level, we selected single cells from the upper and lower 25% of E–P interactions. Subsequently, we examined the UMI counts in these two groups of cells in both 2i and serum mESCs. The results showed that cells with a higher number of E–P interactions also exhibited higher UMI counts in both cell types (Fig. 5f and Extended Data Fig. 8f). Therefore, the increase in overall transcription level in single cells is associated with an increase in E–P interactions during the cell cycle.

To investigate whether differences in the expression of the marker genes were associated with a difference in chromatin structure at specific phases of the cell cycle, we identified marker genes between the Early S and Late S–G2 cell-cycle phases (Fig. 5g). Subsequently, we performed a gene ontology (GO) analysis and found that these

genes were significantly enriched during 'chromatin organization' (Fig. 5h). By further aggregating the structural data for cells in each phase and examining the differences in structure of these 'chromatin organization'-related marker genes, we discovered a high correlation between the number of E–P interactions in these regions and the corresponding gene expression levels (Fig. 5i). To exemplify this correlation, we focused on the gene *Exosc10*, which participates in cell-cycle and mitotic progression[53]. Our analysis revealed that *Exosc10* exhibited more E–P interactions during the Late S–G2 phase, coinciding with its higher expression levels (Fig. 5j,k). Moreover, we filtered expressed genes in E–P interactions from single cells in the Early S and Late S–G2 phases (Extended Data Fig. 8g), and found that genes from a specific phase were also enriched with GO terms related to the cell cycle in a similar phase (Extended Data Fig. 8h). Therefore, the increase in overall transcription in single cells is associated with an increase in E–P interactions during the cell cycle. In conclusion, our findings indicated that differences in expression of the marker genes were associated with changes in the 3D chromatin structure at specific phases of the cell cycle .

Taken together, these results illustrated that periodic changes in 3D chromatin structure are in parallel with transcription during the cell cycle.

Discussion

Here we reported scCARE-seq, a single-cell multimodal omics method, that enabled the simultaneous detection of chromatin architecture and mRNA expression in single cells, and was comparable with previously established scHi-C and scRNA-seq methods. We also demonstrated the utility of scCARE-seq by applying it to 2i and serum mESCs, and found that our method was better able to reveal differences in 3D chromatin structure through transcriptional clustering. In addition, we revealed that periodic changes in chromatin architecture and transcription are interrelated in the cell cycle. These findings suggested that scCARE-seq is a promising tool for simultaneously identifying the cell fate and cell cycle of individual cells to solve the problem of interference from inconsistent cell-cycle phases during cell fate transition. In conclusion, scCARE-seq is an efficient method for joint profiling of 3D chromatin structure and transcription in single cells, and will provide further insight into embryonic development, cancer and other diseases. In addition, scCARE-seq can be used to identify rare cell types from different tissues[40].

There are several advantages associated with scCARE-seq. First, it can simultaneously detect 3D chromatin structure and transcription at the single-cell level, which reflects the global relationship between 3D chromatin structure and gene expression. Moreover, it does not require any special equipment and can be performed in most laboratories. Overall, scCARE-seq is an NGS-based technique for simultaneous detection of 3D chromatin structure and transcription in single cells, and can complement microscopy-based approaches.

However, as with existing single-cell multimodal omics methods[54], scCARE-seq data are still sparse. Therefore, the sensitivity of scCARE-seq needs to be improved and the number of single cells that can be detected needs to be scaled-up for more widespread application in biomedical research.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41594-023-01066-9.

References

1. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**, 292–301 (2001).
2. Dekker, J. & Mirny, L. The 3D genome as moderator of chromosomal communication. *Cell* **164**, 1110–1121 (2016).
3. Wei, C. et al. CTCF organizes inter-A compartment interactions through RYBP-dependent phase separation. *Cell Res.* **32**, 744–760 (2022).
4. Bonev, B. et al. Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572.e524 (2017).
5. Du, Z. et al. Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* **547**, 232–235 (2017).
6. Lupiáñez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
7. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
8. Dixon, J. R. et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
9. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
10. Huang, J., Jiang, Y., Zheng, H. & Ji, X. BAT Hi-C maps global chromatin interactions in an efficient and economical way. *Methods* **170**, 38–47 (2020).
11. Quinodoz, S. A. et al. Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell* **174**, 744–757.e724 (2018).
12. Kempfer, R. & Pombo, A. Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.* **21**, 207–226 (2020).
13. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).
14. Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* **18**, 1509–1517 (2008).
15. Gertz, J. et al. Transposase mediated construction of RNA-seq libraries. *Genome Res.* **22**, 134–141 (2012).
16. Lianoglou, S., Garg, V., Yang, J. L., Leslie, C. S. & Mayr, C. Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev.* **27**, 2380–2396 (2013).
17. Schwartzman, O. & Tanay, A. Single-cell epigenomics: techniques and emerging applications. *Nat. Rev. Genet.* **16**, 716–726 (2015).
18. Wen, L. & Tang, F. Single-cell sequencing in stem cell biology. *Genome Biol.* **17**, 71 (2016).
19. Nagano, T. et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).
20. Nagano, T. et al. Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61–67 (2017).
21. Stevens, T. J. et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017).
22. Flyamer, I. M. et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **544**, 110–114 (2017).
23. Ramani, V. et al. Massively multiplex single-cell Hi-C. *Nat. Methods* **14**, 263–266 (2017).
24. Tan, L., Xing, D., Chang, C.-H., Li, H. & Xie, X. S. Three-dimensional genome structures of single diploid human cells. *Science* **361**, 924–928 (2018).
25. Tang, F. et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
26. Svensson, V., Vento-Tormo, R. & Teichmann, S. A. Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* **13**, 599–604 (2018).
27. Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. *Nat. Rev. Genet.* **20**, 631–656 (2019).
28. Stuart, T. & Satija, R. Integrative single-cell analysis. *Nat. Rev. Genet.* **20**, 257–272 (2019).
29. Miao, Z., Humphreys, B. D., McMahon, A. P. & Kim, J. Multi-omics integration in the age of million single-cell data. *Nat. Rev. Nephrol.* **17**, 710–724 (2021).
30. Mateo, L. J. et al. Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**, 49–54 (2019).
31. Su, J.-H., Zheng, P., Kinrot, S. S., Bintu, B. & Zhuang, X. Genome-scale imaging of the 3D organization and transcriptional activity of chromatin. *Cell* **182**, 1641–1659.e1626 (2020).
32. Cardozo Gizzi, A. M. et al. Direct and simultaneous observation of transcription and chromosome architecture in single cells with Hi-M. *Nat. Protoc.* **15**, 840–876 (2020).
33. Takei, Y. et al. Integrated spatial genomics reveals global architecture of single nuclei. *Nature* **590**, 344–350 (2021).
34. Cao, J. et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).
35. Lee, D.-S. et al. Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. *Nat. Methods* **16**, 999–1006 (2019).
36. Yang, T. et al. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res.* **27**, 1939–1949 (2017).
37. Bakken, T. E. et al. Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PLoS ONE* **13**, e0209648 (2018).
38. Wu, H., Kirita, Y., Donnelly, E. L. & Humphreys, B. D. Advantages of single-nucleus over single-cell RNA sequencing of adult kidney: rare cell types and novel cell states revealed in fibrosis. *J. Am. Soc. Nephrol.* **30**, 23–32 (2019).
39. 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).
40. 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).
41. Li, G. et al. Joint profiling of DNA methylation and chromatin architecture in single cells. *Nat. Methods* **16**, 991–993 (2019).
42. Zhu, C. et al. An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome. *Nat. Struct. Mol. Biol.* **26**, 1063–1070 (2019).
43. Zhu, C. et al. Joint profiling of histone modifications and transcriptome in single cells from mouse brain. *Nat. Methods* **18**, 283–292 (2021).

44. 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* **18**, 652–660 (2021).

45. Chen, S., Lake, B. B. & Zhang, K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat. Biotechnol.* **37**, 1452–1457 (2019).

46. Cao, J. et al. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380–1385 (2018).

47. Mimitou, E. P. et al. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat. Methods* **16**, 409–412 (2019).

48. Ficz, G. et al. FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* **13**, 351–359 (2013).

49. Marks, H. et al. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590–604 (2012).

50. McLaughlin, K. et al. DNA methylation directs polycomb-dependent 3D genome re-organization in naive pluripotency. *Cell Rep.* **29**, 1974–1985.e1976 (2019).

51. Naumova, N. et al. Organization of the mitotic chromosome. *Science* **342**, 948–953 (2013).

52. Gibcus, J. H. et al. A pathway for mitotic chromosome formation. *Science* **359**, eaao6135 (2018).

53. Stuparević, I. et al. Regulation of the conserved 3′–5′ exoribonuclease EXOSC10/Rrp6 during cell division, development and cancer. *Biol. Rev. Camb. Philos. Soc.* **96**, 1092–1113 (2021).

54. Zhu, C., Preissl, S. & Ren, B. Single-cell multimodal omics: the power of many. *Nat. Methods* **17**, 11–14 (2020).

Methods

Cell culture

mESCs were derived from a mixed (129X1 × 129S1) mouse, which was a gift from S. Gao. The HEK293T cell line was a gift from J. Wang. The serum medium consists of high-glucose DMEM (Hyclone, catalog no. SH30022.01), 15% (v/v) fetal bovine serum (LONSERA, S711-001S), 1× GlutaMAX (Thermo Fisher Scientific, catalog no. 30050-061), 1× MEM nonessential amino acids (Thermo Fisher Scientific, catalog no. 11140-050), 0.1 mM β-mercaptoethanol (Sigma, catalog no. M6250), 1× penicillin/streptomycin (Thermo Fisher Scientific, catalog no. 15140-122), 1% (v/v) nucleoside mix (Sigma) and 1,000 U ml^{-1} recombinant leukemia inhibitory factor (Millipore, catalog no. ESG1107). The serum-to-2i transition was accomplished by washing mESCs twice in PBS (Hyclone, catalog no. SH30256.01) and then switching to 2i medium. For 2i medium, cells were grown under culture conditions comprising 50% Neurobasal (Thermo Fisher Scientific, catalog no. 21103-049), 50% DMEM/F12 (Thermo Fisher Scientific, catalog no. 11330-032), 1× N-2 supplement (Thermo Fisher Scientific, catalog no. 17502-048), 1× B-27 supplement (Thermo Fisher Scientific, catalog no. 12587-010), 1× GlutaMAX (Thermo Fisher Scientific, catalog no. 30050-061), 1× MEM nonessential amino acids (Thermo Fisher Scientific, catalog no. 11140-050), 0.1 mM β-mercaptoethanol (Sigma, catalog no. M6250), 1× penicillin/streptomycin (Thermo Fisher Scientific, catalog no. 15140-122), 1,000 U ml^{-1} recombinant leukemia inhibitory factor (Millipore, catalog no. ESG1107), 1 μM MEK inhibitor PD0325901 (Selleck, catalog no. S1036) and 3 μM GSK3 inhibitor CHIR-99021 (Selleck, catalog no. S1263). The HEK293T cell line was cultured in high-glucose DMEM (Hyclone, catalog no. SH30022.01), supplemented with 10% (v/v) fetal bovine serum (LONSERA, catalog no. S711-001S). All cells were cultured at 37 °C with 5% CO$_2$ and in feeder-free conditions on 0.1% gelatin-coated dishes.

Tn5 preparation

Tn5 preparation followed a previously described procedure with minor modifications[55,56]. Tn5 gene was amplified from pTXB1-Tn5 (Addgene, catalog no. 60240) and cloned into pET28a to construct the pET28a–6×His–Tn5 expression vector, which was transformed into BL21 (DE3) chemically competent cells. A single colony was inoculated and cultured in 30 ml of Luria–Bertani broth with kanamycin (50 μg ml^{-1}) at 37 °C and 200 r.p.m. overnight, and then diluted to 600 ml in the same medium and cultured at 37 °C and 200 r.p.m. to an optical density at 600 nm of 0.6–0.8. Bacterial pellets were collected and resuspended in bacterial lysis buffer (50 mM Tris–HCl pH 8.0, 300 mM NaCl, 20 mM imidazole, 0.1% Triton X-100, 1× EDTA-free protease inhibitor cocktails (Roche, catalog no. 04693132001) and 1 mg ml^{-1} lysozyme), on ice for 30 min and sonicated to lyse the bacteria. The lysates were spun at 3,800 g and 4 °C for 10 min to give the supernatant, and the solution was adjusted to contain 5 mM 2-mercaptoethanol, 1 mM phenylmethyl sulfonyl fluoride and 1 M NaCl. Cell lysates were then added to 2 ml of 50% Ni-NTA agarose (Qiagen, catalog no. 30210) with equilibrating (50 mM Tris–HCl pH 8.0, 1 M NaCl and 20 mM imidazole) and incubated for 1 h at 4 °C with rotation. After loading the sample, the Ni-NTA column was washed with 20–30 ml of wash buffer (50 mM Tris–HCl pH 8.0, 1 M NaCl, 20 mM imidazole, 0.1% Triton X-100). Finally, Tn5 was eluted with 4 ml of elution buffer (50 mM Tris–HCl pH 8.0, 1 M NaCl, 250 mM imidazole, 0.1% Triton X-100) and dialyzed in 1 L of 2× dialysis buffer (50 mM Tris–HCl pH 8.0, 0.2 M NaCl, 0.2 mM EDTA, 2 mM dithiothreitol, 0.2% Triton X-100, 20% glycerol). The dialyzed protein solution was concentrated using an Amicon Ultra 30-kDa centrifugal filter (Millipore, catalog no. UFC803024) to 25 μM, and 1× volume of glycerol was added before storage. Before the tagmentation reaction, purified Tn5 and annealed barcoded adapter were mixed in an equal molar ratio at a final concentration 12.5 μM and incubated at 25 °C for 1 h.

scCARE-seq experimental procedure

Cell fixation. Digested single cells were resuspended at a concentration of 0.1 million cells per milliliter of PBS before fixation, which largely eliminated multiple nuclei after FACS[35]. Cells were fixed in 1% formaldehyde (Sigma, catalog no. F8775) for 10 min at room temperature with rotation, and glycine (Sigma, catalog no. G7126) was added to quench the reaction at a final concentration of 125 mM for 5 min at room temperature with rotation. Cells were washed twice with ice-cold PBS containing 5 μl of RNaseOUT recombinant ribonuclease inhibitor (Invitrogen, catalog no. 10777-019) per milliliter via centrifugation at 4 °C, 1,000g for 5 min. One million cells were aliquoted into 1.5 ml RNase/DNase-free tube. The supernatant was discarded, and remaining processes were undertaken or the cell pellets were flash-frozen in liquid nitrogen for storage at −80 °C.

Reverse transcription. The process was similar to that described previously[34,46] with minor modifications. One milliliter of IGEPAL lysis buffer (10 mM Tris–HCl pH 7.5 (Invitrogen, catalog no. 15567-027), 10 mM NaCl (Sigma, catalog no. S5150), 3 mM MgCl$_2$ (Invitrogen, catalog no. AM9530G) and 0.1% IGEPAL CA-630 (Sigma, catalog no. I8896), 1% SUPERase In RNase Inhibitor (Invitrogen, catalog no. AM2696) and 1% BSA (Sigma, catalog no. A7906)) was incubated with cells on ice in a 1.5-ml RNase/DNase-free tube for 20 min, followed by centrifugation at 4 °C, 1,000g for 5 min to remove the supernatant. After washing twice in 100 μl of lysis buffer, nuclei were suspended in 500 μl of ice-cold lysis buffer (IGEPAL lysis buffer without IGEPAL CA-630), 8 μl was transferred to RNase/DNase-free PCR tubes and 8 μl of sterile nuclease-free water was added. For each tube, 1.6 × 10^4 nuclei were mixed with 8 μl of 25 μM oligo-dT primer (5′-ACGACGCTCTTCCGATCTNNNNNNNN[8 bp RNA barcode]TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN-3′, where 'N' is any base and 'V' is either 'A', 'C' or 'G') and 2 μl of 10 mM dNTP mix (Thermo Fisher Scientific, catalog no. R0191), denatured at 55 °C for 5 min and immediately placed on ice. Fourteen microliters of first-strand reaction mix, containing 8 μl of 5× Superscript IV first-strand buffer, 2 μl of 100 mM dithiothreitol, 2 μl of SuperScript IV reverse transcriptase (Invitrogen, catalog no. 18090050) and 2 μl of RNaseOUT was then added to each PCR tube. Reverse transcription was performed as follows: 4 °C for 2 min, 10 °C for 2 min, 20 °C for 2 min, 30 °C for 2 min, 40 °C for 2 min, 50 °C for 2 min and 55 °C for 10 min. The supernatant was removed by centrifugation at 4 °C, 1,000g for 5 min, leaving 3 μl.

Chromatin digestion. To each PCR tube was added 3 μl of 0.6% SDS at 62 °C for 10 min. SDS was quenched by the addition of 15 μl of doubly distilled H$_2$O (ddH$_2$O) and 3 μl of 10% Triton X-100 and incubation at 37 °C for 15 min. Then 3 μl of 10× NEBuffer 2 (NEB, catalog no. B7002S) and 3 μl of 25 U μl^{-1} MboI (NEB, catalog no. R0147M) were added and the chromatin was digested at 37 °C for 3 h.

Proximity ligation. The supernatant was removed, and 30 μl of 1× Ligation Master Mix without T4 DNA ligase (3 μl of 10× NEB T4 DNA ligase reaction buffer (NEB, catalog no. B0202S), 3 μl of 10% Triton X-100, 0.3 μl of 10 mg ml^{-1} BSA and 23.7 μl of ddH$_2$O) was added to resuspend the nuclei. After removing the supernatant, 27 μl of Ligation Master Mix (3 μl of 10× NEB T4 DNA ligase reaction buffer, 3 μl of 10% Triton X-100, 0.3 μl of 10 mg ml^{-1} BSA, 1.5 μl of T4 DNA ligase (NEB, catalog no. M0202S), and 19.2 μl of ddH$_2$O) was added. The sample was then incubated at 16 ˚C for 4 h. All supernatants were removed by centrifugation at 4 °C, 1,000g for 5 min, leaving 3 μl.

Second strand synthesis. After removing the supernatants, 65 μl of sterile nuclease-free water, 8 μl of mRNA Second Strand Synthesis buffer and 4 μl of mRNA Second Strand Synthesis enzyme (NEB, catalog no. E6111) were added to each tube, and second strand synthesis was carried out at 16 °C for 180 min.

FACS and cell lysis. The nuclei were then passed through cell strainers (40 μm) and stained with 4′,6-diamidino-2-phenylindole. Single nuclei were sorted into each well in 96-well plates and lysed in 3 μl of scReverse Crosslink Buffer (20 mM Tris–HCl pH 8.0 (Invitrogen, catalog no. 15568-025), 20 mM NaCl (Sigma, catalog no. S5150), 0.1% Triton X-100, 15 mM dithiothreitol, 1 mM EDTA pH 8.0 (Invitrogen, catalog no. 15575-020), 1 mg ml^{-1} Proteinase K (Invitrogen, catalog no. 25530-049) and 0.5 μM carrier single-strand DNA)) at 55 °C for 2 h, 65 °C for 12 h and then at 70 °C for 30 min.

Tagmentation. Each well was mixed with 4 μl of ddH$_2$O, 2 μl of 5× TAPS–DMF (50 mM TAPS-NaOH pH 8.5, 25 mM MgCl$_2$, 50% dimethylformamide) and 1 μl of 0.025 μM in-house Tn5, and incubated at 55 °C for 10 min, 72 °C for 5 min. Then, 2 μl of 0.5 mg ml^{-1} Proteinase K was added to each well using a multichannel pipette, and the reactions were incubated at 55 °C for 5 min, 65 °C for 25 min and 70 °C for 30 min.

Preamplification. Thirteen microliters of PCR mix, containing 0.5 μl of Mix preamplification primers (common primer, DNA primer, RNA primer, 40 μM each primer; Supplementary Table 1) and 12.5 μl of Q5 High-Fidelity 2× Master Mix (NEB, M0492L), was added to each well. Preamplification was carried out using the following program: 72 °C for 5 min, 98 °C for 30 s, 12–14 cycles of 98 °C for 10 s, 65 °C for 30 s, 72 °C for 1 min and a final 72 °C for 5 min. Previous primers were removed by the addition of 2 μl of 5 U μl^{-1} ExoI (NEB, M0293S) and incubation at 37 °C for 60 min and 80 °C for 20 min. Each well was split (12 μl each) and transferred to two new 96-well plates for chromatin architecture and expression analysis, respectively.

Hi-C library preparation. Each well was mixed with 8 μl of PCR Mix (2.2 μl of ddH$_2$O, 0.8 μl of Index DNA Primer Mix (10 μM Nextera index i5 and Nextera index i7), 5 μl of Q5 High-Fidelity 2× Master Mix) and then incubated at 98 °C for 30 s, four cycles of (98 °C for 10 s, 60 °C for 30 s, 72 °C for 1 min) and a final 72 °C for 5 min. Products were pooled and purified by a DNA Clean and Concentrator-5 column (Zymo, D4013) with a fivefold volume of DNA binding buffer and 0.8× AMPure XP beads (Beckmann, catalog no. A63881).

mRNA library preparation. PCR was performed by the addition of 8 μl of PCR Mix (2.2 μl of ddH$_2$O, 0.8 μl of Index RNA Primer Mix (10 μM Truseq index i5 and Nextera index i7), 5 μl of Q5 High-Fidelity 2× Master Mix) and incubation at 98 °C for 30 s, ten cycles of 98 °C for 10 s, 65 °C for 30 s, 72 °C for 1 min and a final 72 °C for 5 min. Libraries were pooled and purified using a DNA Clean and Concentrator-5 column with a fivefold volume of DNA binding buffer and 0.8× AMPure XP beads.

Quantification and sequencing. Both scHi-C and scRNA-seq libraries were quantified by Qubit 4 (Invitrogen, catalog no. Q33238) and quantitative real-time PCR. Libraries were sequenced on the Illumina platform.

CARE-seq
For cell populations, the main steps are similar to those of scCARE-seq. Specifically, oligo-biotin-dT (5′-ACGACGC[BiodT] CTTCCGATCTNNNNNNNNN[8 bp RNA barcode]TTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTVN-3′, where 'N' is any base and 'V' is either 'A', 'C' or 'G') was used for reverse transcription, and the restriction fragment ends were filled-in by addition of Fill-in Mix (0.4 mM bioticcan-14-ATP (Invitrogen, catalog no. 19524-016), 0.4 mM dCTP/dGTP/dTTP, 0.8 U μl^{-1} DNA Polymerase I, Large (Klenow) Fragment (NEB, catalog no. M0210L)) and incubation at 37 °C for 1 h. Tn5 tagmentation was performed after ligation and cell lysis without single-cell sorting, followed by direct library construction after biotin enrichment. The biotin enrichment step is the same as in previous methods[9]. Then 50 μl of PCR mix (18 μl of ddH$_2$O, 3 μl of 10 μM Nextera index i5 and 3 μl of 10 μM Nextera index i7),

1 μl of 10 mg ml^{-1} BSA, 25 μl of Q5 High-Fidelity 2× Master Mix) was added to the sample. PCR amplification was performed in a thermal cycler: 98 °C for 30 s, 10–14 cycles of 98 °C for 10 s, 65 °C for 30 s and 72 °C for 1 min, and a final 72 °C for 5 min. After PCR, the sample was purified using 0.6×/0.2× volume of AMPure XP beads.

Hi-C
Hi-C was performed according to a published protocol[9,10], but the library preparation was optimized. Briefly, after sonication, the purified DNA per 0.5 million cells was added to tagmentation buffer (27 μl of ddH$_2$O, 10 μl of 5× TAPS–DMF, 3 μl of 12.5 μM in-house Tn5). The sample was then enriched using Streptavidin T1 beads (Invitrogen, catalog no. 65602). The PCR amplification and purification steps are the same as CARE-seq procedures described above.

RNA-seq
An optimized RNA-seq was based on previous research[15]. For each tube, 3 μg of purified total RNA from mESCs was mixed with 4 μl of 12.5 μM oligo-dT primer and 1 μl of 10 mM dNTP mix, denatured at 55 °C for 5 min and immediately placed on ice. Then 7 μl of first-strand reaction mix, containing 4 μl of 5× Superscript IV First-Strand Buffer, 1 μl of 100 mM dithiothreitol, 1 μl of SuperScript IV reverse transcriptase and 1 μl of RNaseOUT was added to each PCR tube. Reverse transcription was performed as follows: 4 °C for 2 min, 10 °C for 2 min, 20 °C for 2 min, 30 °C for 2 min, 40 °C for 2 min, 50 °C for 2 min and 55 °C for 10 min. The supernatant was removed by centrifugation at 4 °C, 1,000g for 5 min, leaving 3 μl. Forty-eight microliters of sterile nuclease-free water, 8 μl of mRNA Second Strand Synthesis buffer and 4 μl of mRNA Second Strand Synthesis enzyme were added to each tube, and second strand synthesis was carried out at 16 °C for 180 min. The sample was purified using 2× volume of AMPure XP beads or DNA Clean and Concentrator-5 column and eluted in 10 μl of H$_2$O. The tagmentation reaction was initiated by mixing 27 μl of ddH$_2$O, 10 μl of 5× TAPS–DMF and 3 μl of 12.5 μM in-house Tn5, and incubated at 37 °C for 20 min and 72 °C for 5 min. After purification, the PCR amplification and size selection steps are the same as CARE-seq procedures described above.

Nuclei RNA-seq
The procedures were based on previously described methods[34,46]. After cell harvest, reverse transcription was the same as for the scCARE-seq procedures described above. Then 65 μl of sterile nuclease-free water, 8 μl of mRNA Second Strand Synthesis buffer and 4 μl of mRNA Second Strand Synthesis enzyme were added to the sample, and second strand synthesis was carried out at 16 °C for 180 min. Ten microliters of Reverse Crosslink Buffer (20 mM Tris–HCl pH 8.0, 20 mM NaCl, 0.1% Triton X-100, 15 mM dithiothreitol, 1 mM EDTA pH 8.0 and 1 mg ml^{-1} Proteinase K) was added for cell lysis, followed by tagmentation. After purification, the PCR amplification and size selection steps are the same as CARE-seq procedures described above.

Data analysis
Preprocessing of scCARE-seq and CARE-seq data. bcl files were converted to FASTQ using bcl2fastq (v.2.20.0). To separate Hi-C and RNA information, reads of RNA information were identified by custom scripts using RNA barcode. Adapters and low-quality reads of Hi-C and RNA information were then trimmed with Trim Galore (v.0.6.6, https://github.com/FelixKrueger/TrimGalore). For Hi-C partition, Hi-C paired-end reads were aligned to the mm10 reference genome and paired using HiC-Pro[57] (v.3.0.0, parameters: default settings). The data were then used to generate contact matrices and corrected with ice_norm[58] (part of HiC-Pro). For each chromosome, the ICE-normalized 10 kb, 40 kb, 50 kb, 100 kb, 250 kb, 500 kb, 1 Mb and 5 Mb resolution contact matrices were used for further analysis. For RNA partition, reads were aligned and counted using modified scRNA-seq pipeline of sci-CAR analysis[46] (https://github.com/JunyueC/sci-CAR_analysis),

and mm10 was used as reference genome. EdgeR was used to analyze differential count data between mESC groups[59]. In addition, Hi-C and other RNA methods data processing are similar to the above steps, excluding the separation of Hi-C and RNA information step. For species mixing experiments, nuclei with less than 90% of DNA reads and 75% of RNA reads mapped to one species were classified as multiple cells[35]. RNA reads were aligned to a combined hg19 and mm10 genome and only primary alignments were considered[60].

Hi-C correlation analysis. The correlation of Hi-C matrices was analyzed by hicCorrelate (part of hicexplorer[61], v.3.7.1, parameters: '−range 500000:3000000, --method = pearson, --log1p, --zMin 0, --zMax 1, --plotNumbers'). Regarding the compartment and insulation scores generated from GENOVA[62] (v.1.0.0, https://github.com/robinweide/GENOVA), the corresponding correlation was calculated by in-house R scripts using Spearman and Pearson methods.

RNA correlation analysis. The correlation of gene expression was analyzed by custom R scripts. We removed mitochondrial, ribosome-related genes and a few abnormal expressed genes. We then calculated the correlation coefficient using Spearman and Pearson methods in the remaining genes or highly expressed 25,000 genes of mESCs.

Visualization of chromatin architecture and expression. Track view was displayed using pyGenomeTracks[63] (v.3.6), and parameters for Hi-C matrix: 'depth = 700000 or 10000000, transform = log1p', parameters for RNA bigwig: 'summary_method = mean or max'. To generate Pearson's correlation matrices, all sampled Hi-C valid pairs were transferred to '.hic' format files using command of juicer[64], and visualization in Juicebox (v.1.11.08, https://github.com/aidenlab/Juicebox). In addition, matrix plots in two experiments, relative contact probability and Saddle-analyses were performed using GENOVA[62] to compare the 3D chromatin structure features.

Comparison of scCARE-seq chromatin architecture data with published single-cell Hi-C data. To assess the quality of chromatin architecture data generated by scCARE-seq, we compared the number of nonduplicated mapped pairs and the *cis*-to-*trans* ratio with published single-cell Hi-C technologies[19–22,39,40]. The published datasets were downloaded from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) under accession codes GSE48262, GSE80006, GSE94489, GSE80280, GSE121791 and GSE162511.

Comparison of scCARE-seq expression data with published single-cell multimodal omics expression data. To assess the reliability of the expression profile generated by scCARE-seq, we compared the UMI numbers and gene numbers with published single-cell multimodal omics technologies[42–46]. The published datasets were downloaded from the NCBI GEO under accession codes GSE117089, GSE126074, GSE158435, GSE152020 and GSE130399.

Compartment and TAD analysis. Principal component analysis was performed to detect active (A compartment) and inactive (B compartment) chromatin regions along the genome using the HOMER[65] (v.4.11) tool 'runHiCpca.pl' following the parameters '-res 1000000 -window 2000000 -genome mm10'. Next, the sign of the PC1 value of each bin was reassigned to A or B compartments according to the gene density. A positive PC1 value represents the A compartment, which has a higher gene density, whereas a negative PC1 value represents the B compartment. TAD boundaries were identified by TopDom[66] (v.0.0.2) with a window size of 5 based on the 40 kb ICE-normalized matrices in this study.

Identification of single-cell compartment and single-cell TAD. The contact maps of scCARE-seq were imputed using the Higashi-analysis pipeline[67] with default parameters. Single-cell A/B compartment detection and single-cell TAD identification were also utilized Higashi at 1 Mb and 40 kb resolution, respectively.

Identification of E–P interactions. The E–P interactions of different cell groups were imputed using the DeepLoop-analysis pipeline[68] with default parameters. Enhancers were defined by mESC H3K27ac chromatin immunoprecipitation (ChIP) seqencing[69].

Clustering of scCARE-seq data in mESCs. We utilized Higashi[67] software to calculate dimensionality reduction and clustering of the scHi-C data. We used Seurat[70] (v.4.1.0) to identify cell types after removing mitochondrial genes and using the 'FindVariableFeatures' function (parameters: selection.method = 'vst', nfeatures = 3000). Using the 'ScaleData' Function, we then performed principal component analysis with the 'RunPCA' function (parameters: features = VariableFeatures). Then, *t*-distributed stochastic neighbor embedding (*t*-SNE) was calculated using 'RunTSNE' (parameters: features = VariableFeatures, tsne.method = 'Rtsne', dims=c(1,2,3)), 'FindNeighbors' (parameters: dims = 1:3), and 'FindClusters' (parameters: resolution = 0.5) functions. We identified five clusters which were identified with the 'FindAllMarkers' function (parameters: only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25) in the R package Seurat.

In silico cell phasing over the cell cycle. Cell-cycle analysis was performed as described in a previous study[20]. Briefly, based on the original steps, we used UMAP information from Higashi[67] to correct the cell-phasing results. Next, we combined 3D and RNA information for each cell to show how that information changes during the cell cycle.

Reporting summary

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

Data availability

Sequencing data have been deposited at the NCBI GEO with accession number GSE211395, using mm10 reference genome. Other public datasets used in this study were downloaded from NCBI GEO with accession numbers as follows: ChIP-seq (GSE90895; CTCF and H3K27ac), in situ Hi-C (GSE124342), 2013-Nagano (GSE48262), 2017-Flyamer (GSE80006), 2017-Nagano (GSE94489), 2017-Steven (GSE80280), 2019-Tan (GSE121791), 2021-Tan (GSE162511), sci-CAR (GSE117089), SNARE-seq (GSE126074), CoTECH (GSE158435), Paired-Tag (GSE152020) and Paired-seq (GSE130399). Source data are provided with this paper.

Code availability

Custom scripts used in this study are available from https://github.com/jsun9003/scCARE-seq.

References

55. Picelli, S. et al. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).
56. Lai, B. et al. Trac-looping measures genome structure and chromatin accessibility. *Nat. Methods* **15**, 741–747 (2018).
57. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
58. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
59. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
60. Ma, S. et al. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell* **183**, 1103–1116.e1120 (2020).

61. Wolff, J. et al. Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* **48**, W177–W184 (2020).

62. van der Weide, R. H. et al. Hi-C analyses with GENOVA: a case study with cohesin variants. *NAR Genom. Bioinform.* **3**, lqab040 (2021).

63. Lopez-Delisle, L. et al. pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* **37**, 422–423 (2020).

64. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).

65. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

66. Shin, H. et al. TopDom: an efficient and deterministic method for identifying topological domains in genomes. *Nucleic Acids Res.* **44**, e70 (2015).

67. Zhang, R., Zhou, T. & Ma, J. Multiscale and integrative single-cell Hi-C analysis with Higashi. *Nat. Biotechnol.* **40**, 254–261 (2022).

68. Zhang, S. et al. DeepLoop robustly maps chromatin interactions from sparse allele-resolved or single-cell Hi-C data at kilobase resolution. *Nat. Genet.* **54**, 1013–1025 (2022).

69. Chronis, C. et al. Cooperative binding of transcription factors orchestrates reprogramming. *Cell* **168**, 442–459.e420 (2017).

70. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).

Author contributions

J.D. and J.Q. conceived and designed the study. J.Q. designed and performed all experiments. J.Q., J.S., X.L., C.Z. and X. Zhang performed the data analysis. J.Q., J.S. and C.Z. wrote the paper with input from all other authors. S.J. and C.W. discussed results and edited the manuscript. H.Y., X. Zeng and L.F. provided support. J.D. supervised the research.

Competing interests

J.D., J.Q. and J.S. are listed as inventors of a patent application titled 'Single-cell simultaneous detection of 3D chromatin structure and gene expression by sequencing'. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41594-023-01066-9.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41594-023-01066-9.

Correspondence and requests for materials should be addressed to Xiaoxi Zeng, Lili Fan or Junjun Ding.

Peer review information *Nature Structural & Molecular Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available. Primary Handling Editor: Dimitris Typas, in collaboration with the *Nature Structural & Molecular Biology* team.

Reprints and permissions information is available at www.nature.com/reprints.
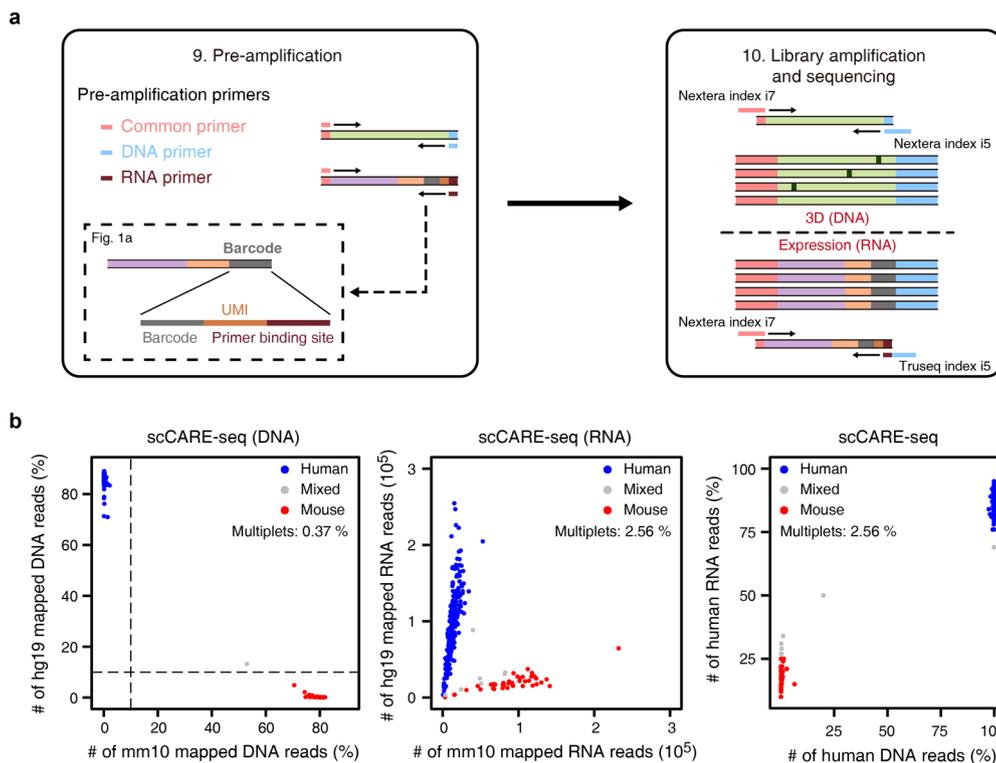
Extended Data Fig. 1 | Schematics for pre-amplification and construction of DNA/RNA libraries. a, Pre-amplification was achieved by primers mix to amplify nucleic acid numbers. DNA/RNA libraries construction was worked by specific DNA or RNA primers. **b**, Scatter plots showing the mapped DNA reads of Hi-C data (left) and mapped RNA-reads of RNA-seq data (middle) in hg19 and mm10 for each cell. And the fraction of human reads in DNA and RNA libraries for each cell (right). HEK293T refers to human embryonic kidney 293 T, and mESCs refers to mouse embryonic stem cells.
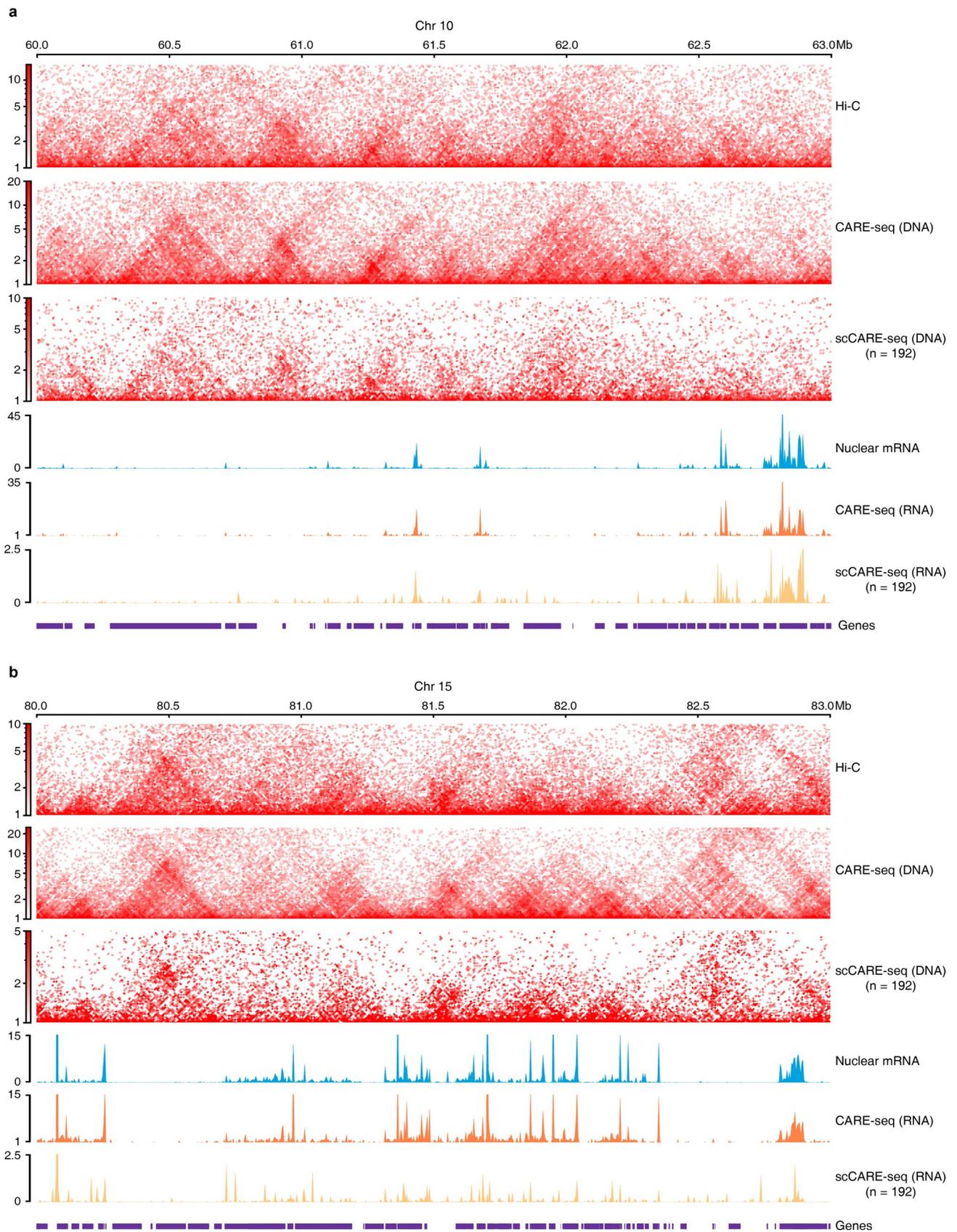
Extended Data Fig. 2 | Overview of our method to simultaneously detect 3D chromatin structure and transcription. a,b, Track view displaying both contact matrices and RNA signals from 60-63 Mb of chromosome 10 (**a**) and 80-83 Mb of chromosome 15 (**b**).
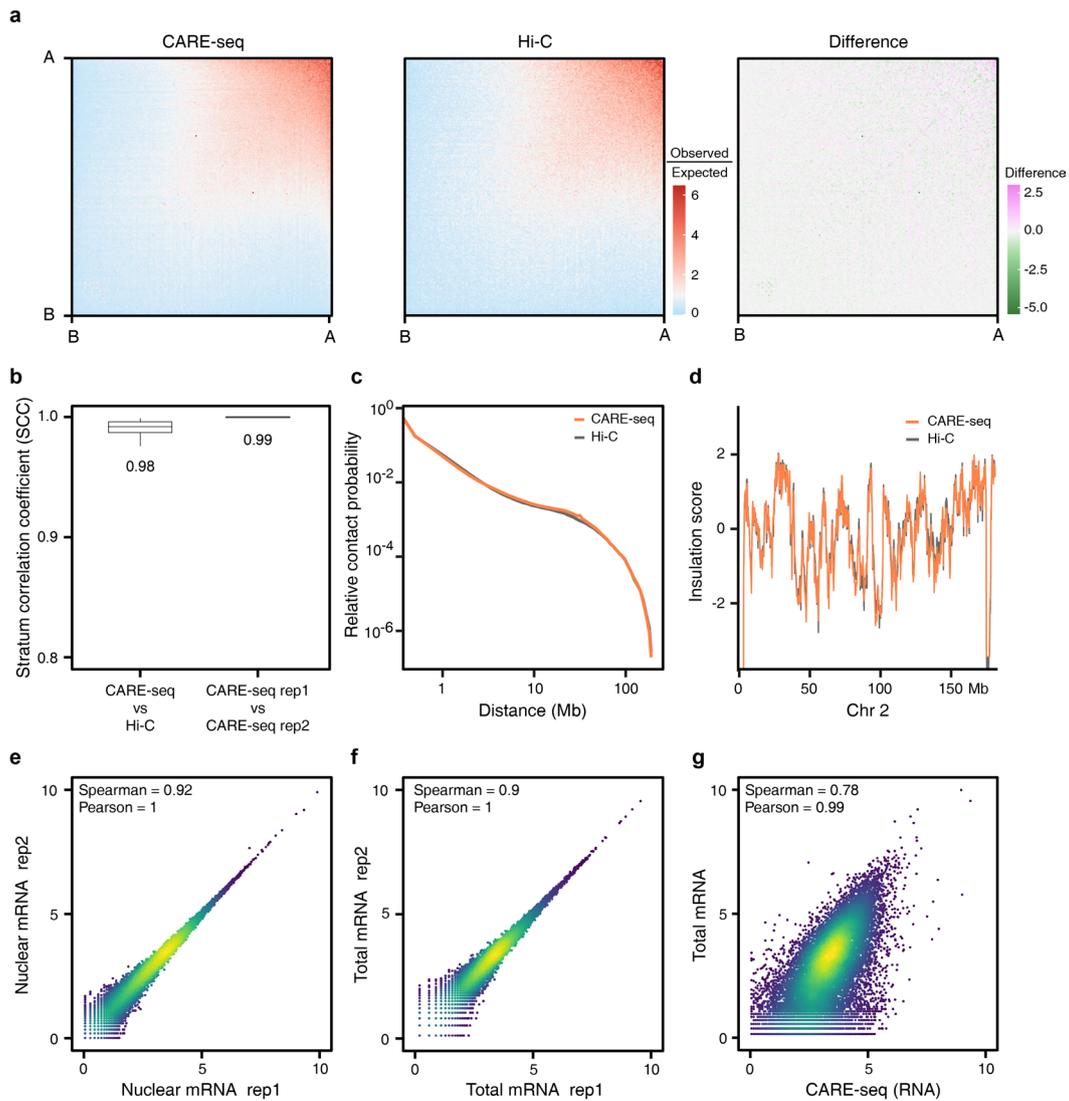
Extended Data Fig. 3 | Performance of CARE-seq in comparative analyses. **a**, Saddle plots: average contact enrichment between pairs of 500 kb regions arranged by their compartment scores and the difference was Hi-C compared to CARE-seq. The upper right quarter represents A-A interactions, the bottom left quarter represents B-B interactions. **b**, The similarity of different bulk 3D chromatin structure data was evaluated by HiCRep at 500 kb resolution per euchromosome (n = 19). The values represented the mean Stratum-adjusted Correlation Coefficient (SCC). The boxplots were drawn from lower quartile (Q1) to upper quartile (Q3), with the middle line denoting the median, whiskers with maximum 1.5 interquartile range (IQR) and outliers were not indicated. **c**, Dependence of contact probability on genomic separation for single cells from CARE-seq data (orange) and Hi-C data (black). **d**, Insulation profiles of CARE-seq and Hi-C over 40 kb bins in chromosomes 2. **e-g**, Scatter plots show the strong concordance of gene expression signals from two technical replicates in nuclear mRNA (**g**) and total mRNA (**f**), and total mRNA versus CARE-seq (**g**).
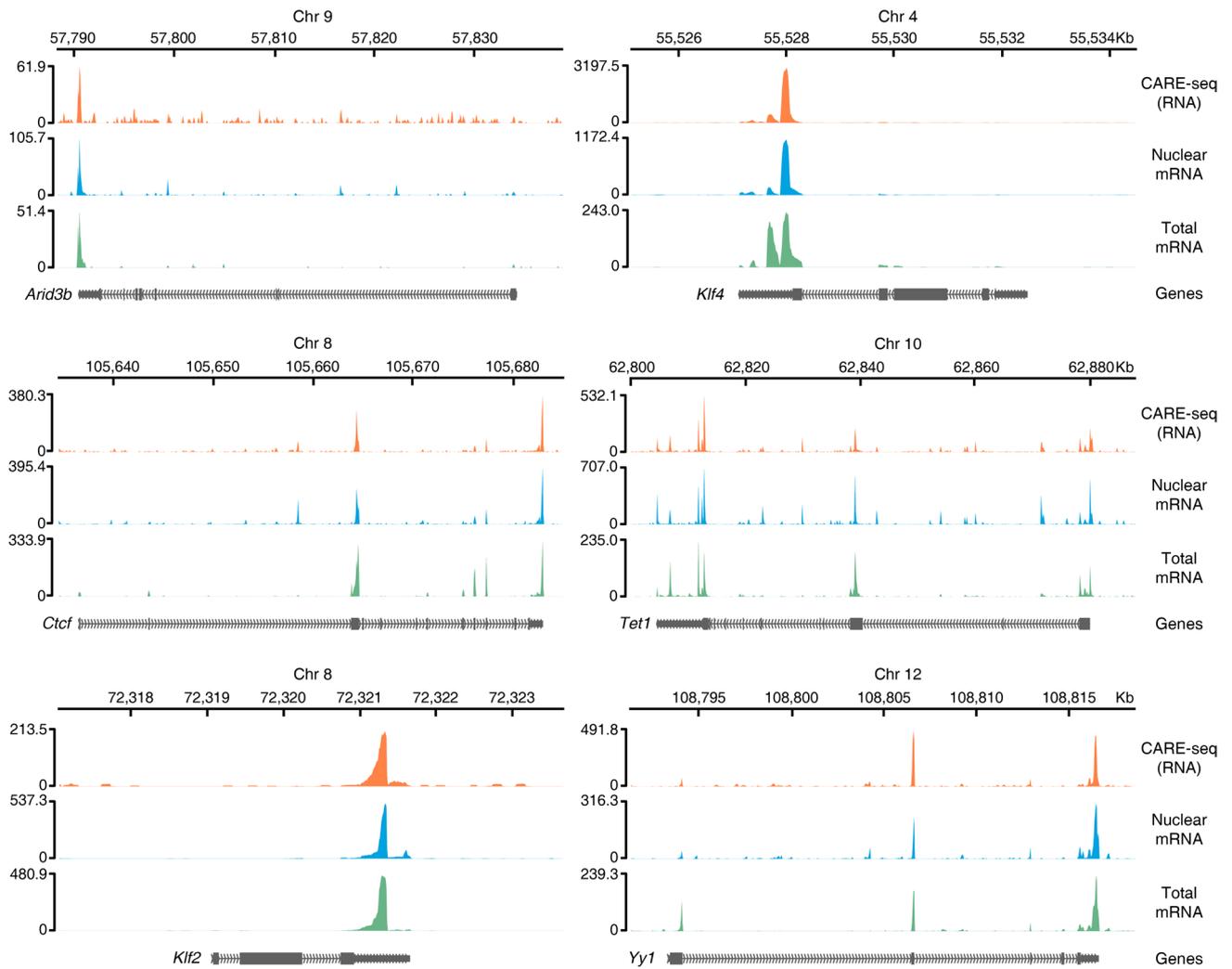
Extended Data Fig. 4 | Comparison of CARE-seq and typical RNA-seq in gene expression profiles. Representative regions showing a consistent pattern of gene expression across datasets.

a

Hi-C Sample	Total pairs	Mapped pairs	Mapping rate	Valid pairs	Valid pairs (rmDup)	Duplication ratio	Cis interaction	Trans interaction	Cis to trans ratio
Mean serum	2821078	2054800	70.16%	277499	114830	44.99%	93587	21243	4.35
Mean 2i	3649154	2731245	72.10%	337567	107393	56.62%	87855	19538	4.57

RNA sample	Total RNA reads	Uniquely aligned reads	Uniquely mapped rate	Remove duplicates	Duplication ratio	Percent of reads mapped gene bodies	Exon ratio	Intron ratio	UMI	GeneNum
Mean serum	197247	112783	56.62%	6794	90.90%	51.85%	52.10%	47.90%	3487	1610
Mean 2i	79224	46206	56.91%	2597	92.01%	54.58%	58.65%	41.35%	1400	772

b
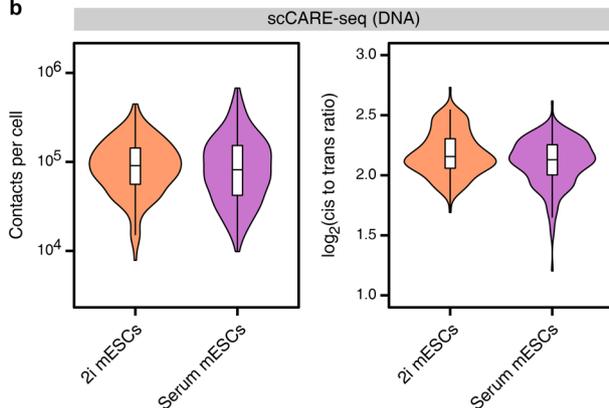
c
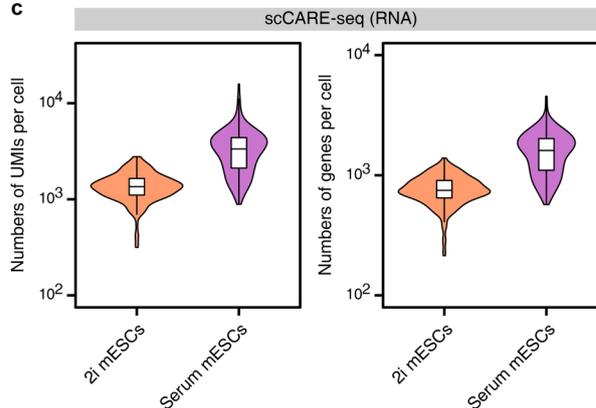
Extended Data Fig. 5 | Comparison of 2i and serum mESCs from scCARE-seq. a, The tables summarized the Hi-C (top) and RNA-seq data (bottom) of Supplementary Table 2, respectively. **b,c**, Comparison of scCARE-seq data from 192 2i mESCs and 192 serum mESCs. The contacts numbers (left) and cis-to-trans ratio (right) (**b**); UMIs (left) and expressed gene numbers (right) (**c**). The boxplots were drawn from lower quartile (Q1) to upper quartile (Q3), with the middle line denoting the median, whiskers with maximum 1.5 interquartile range (IQR) and outliers were not indicated.
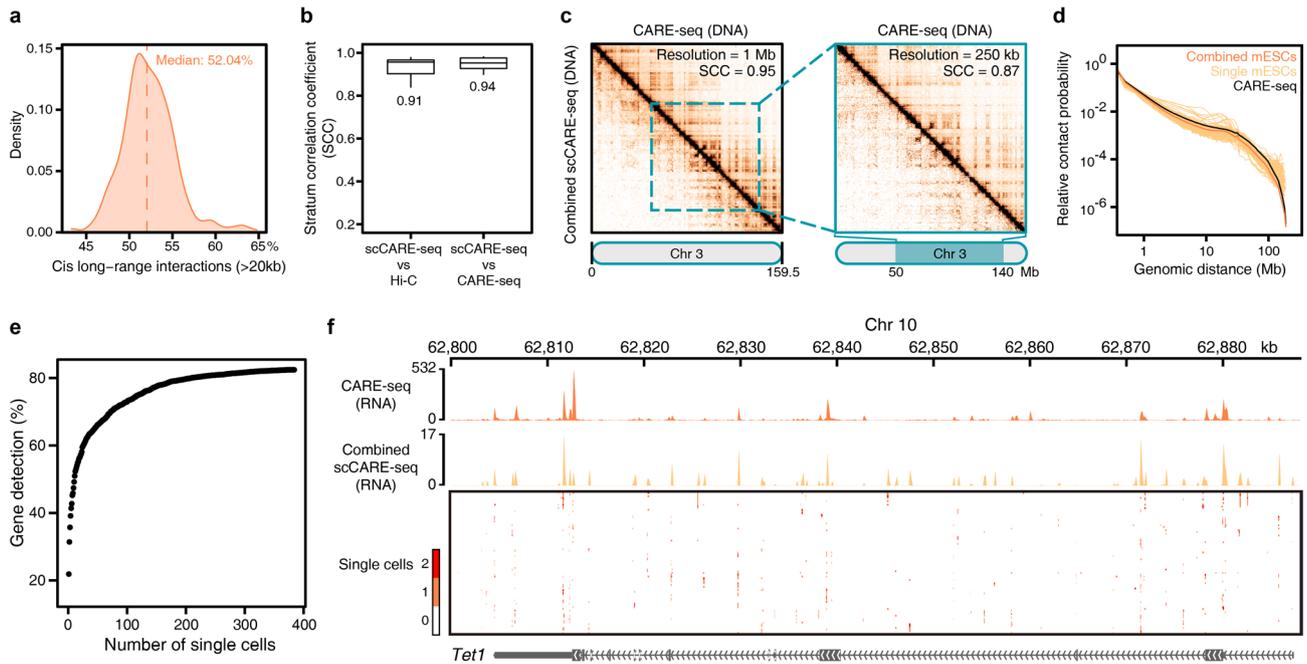
Extended Data Fig. 6 | scCARE-seq data quality in mESCs. a, Distribution of cis long-range interactions (>20 kb) in the scCARE-seq data (n = 384, median = 52.04%). **b**, The similarity between scCARE-seq Hi-C data and Hi-C or CARE-seq Hi-C data was evaluated by HiCRep at 500 kb resolution per euchromosome (n = 19). The values represented the mean Stratum-adjusted Correlation Coefficient (SCC). The boxplots were drawn from lower quartile (Q1) to upper quartile (Q3), with the middle line denoting the median, whiskers with maximum 1.5 interquartile range (IQR) and outliers were not indicated. **c**, Comparison of contact heatmap of chromosome 3 between scCARE-seq and CARE-seq, at 1 Mb resolution (left); 50–140 Mb/250 kb resolution (right). Matrix similarity is evaluated by HiCRep at the corresponding resolution. SCC, Stratum-adjusted Correlation Coefficient. **d**, Dependence of contact probability on genomic separation for single cells from scCARE-seq data (n = 192, yellow), combined scCARE-seq data from all cells (orange) and bulk CARE-seq data (black). **e**, Cumulative coverage percentage of genes detected in single cells compared to the bulk data. **f**, A representative region showing a consistent pattern of gene expression across datasets generated using scCARE-seq and CARE-seq. The transcriptional profiles are gene expression read counts from bulk (upper) and a total of 83 single cells (bottom).
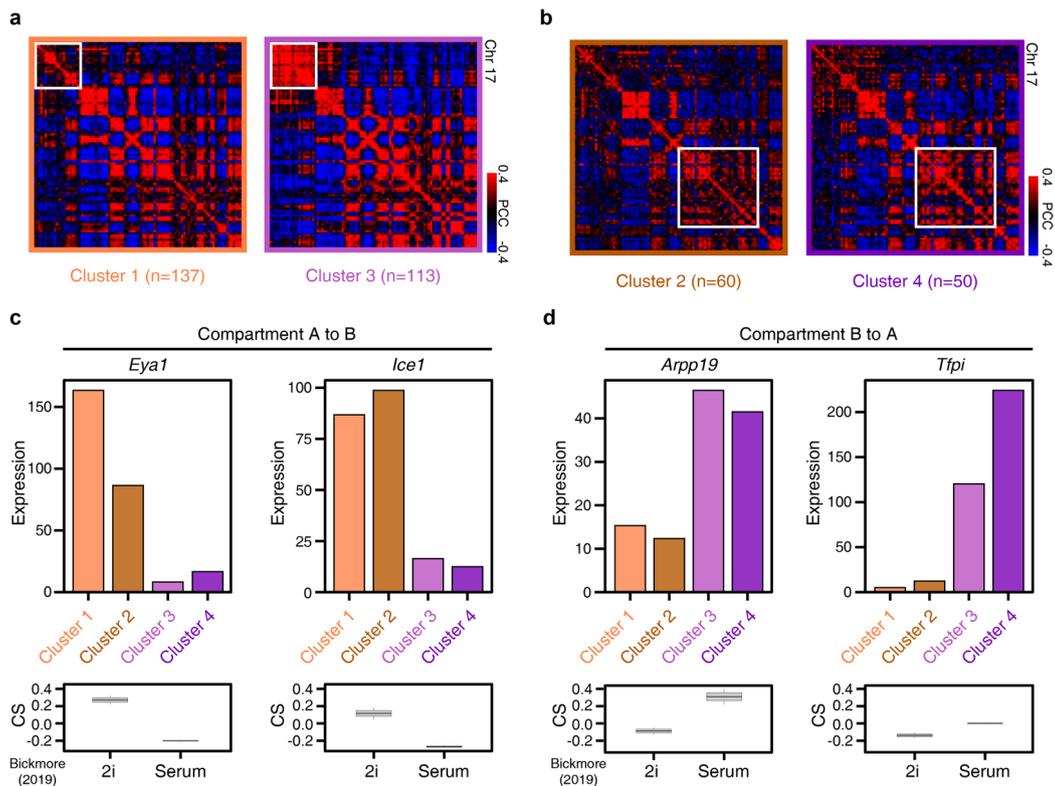
Extended Data Fig. 7 | The relationship between chromatin architecture and gene expression in the different cell clusters. a,b, Pearson's correlation matrixes from different cell clusters. Contacts numbers in different clusters were sampled to same numbers and plot the balanced matrixes in Juicebox (version 1.11.08). Pearson's correlation coefficient was calculated under 1-Mb resolution. White frame shows the difference regions. **c,d,** Showing change in expression of the different clusters (top) in the compartment A to B (**c**) and compartment B to A (**d**), where the compartment switch was defined by published Hi-C data (bottom). 2i and serum have two replications, respectively. CS represents log2(compartment scores+1). The boxplots were drawn from lower quartile (Q1) to upper quartile (Q3), with the middle line denoting the median, whiskers with maximum 1.5 interquartile range (IQR) and outliers were not indicated.
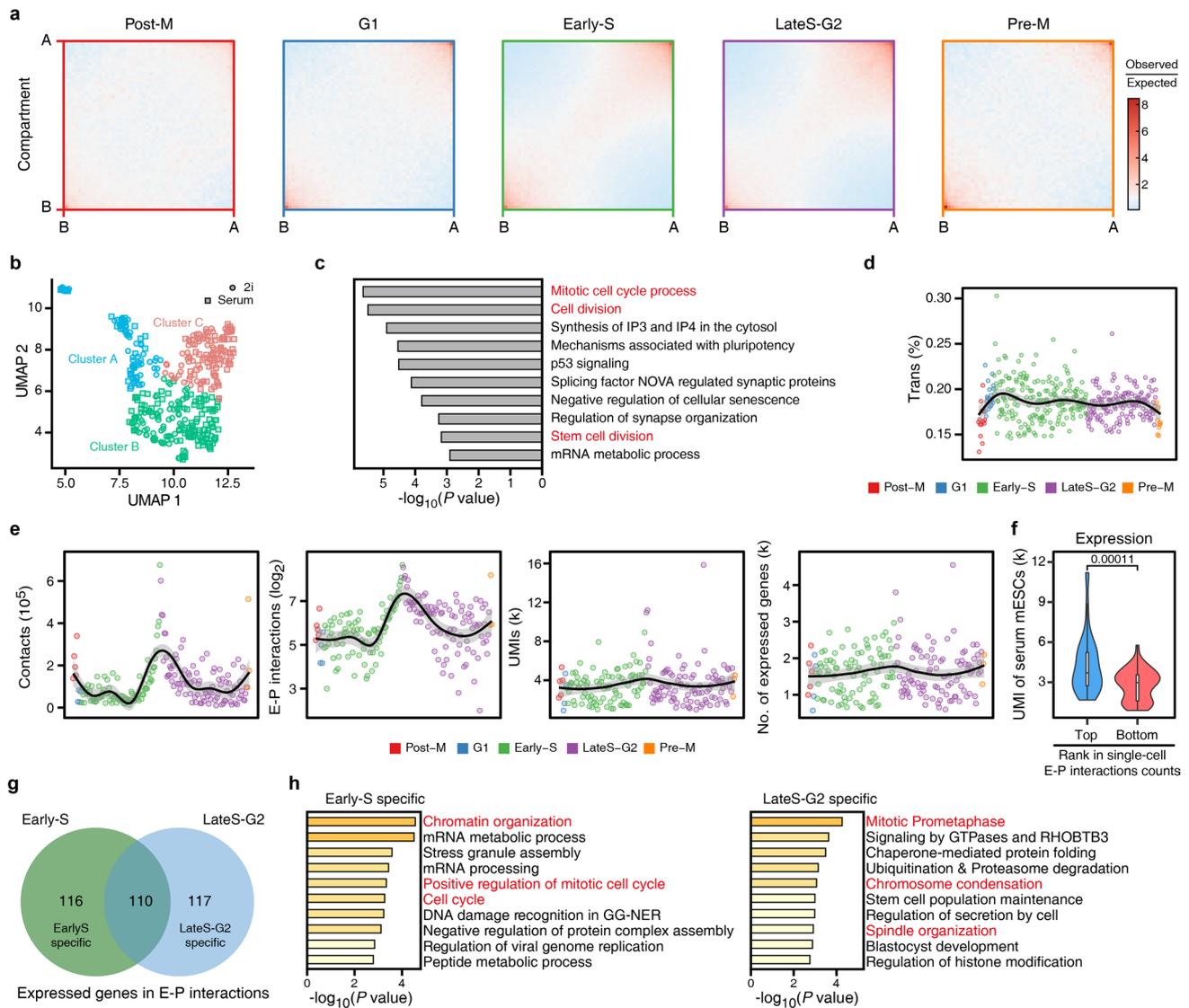
Extended Data Fig. 8 | The relationship between chromatin architecture and gene expression in the cell cycle. a, Saddle plots: average contact enrichment between pairs of 100 kb regions arranged by their compartment scores in the different cell cycle phases. **b**, Uniform manifold approximation and projection (UMAP) embedding showing the clustering of single cells from scCARE-seq 3D chromatin structural profiles. Each dot represents an individual cell and each color represents a cell cluster. **c**, GO enriched by marker genes of clusters in **b**. **d**, Percentage of inter-chromosomal contacts per single mESC in 2i and serum were ordered by cell-cycle phasing and each cell was annotated by cell type colored the same as in **a**. The black line represents mean trend. Shadow represents the confidence intervals of 0.95. **e**, Similar to **d**, Contacts, E-P interactions, UMIs and number of expressed genes per single serum mESCs from left to right. **f**, Comparison of UMI in single cells of top and bottom group in 2i mESCs. The top and bottom groups were selected based on the top and the bottom each 48 single cells (25%) ranked by number of E-P interactions from highest to lowest. The boxplots were drawn from lower quartile (Q1) to upper quartile (Q3), with the middle line denoting the median, whiskers with maximum 1.5 interquartile range (IQR) and outliers were not indicated. P value, one-sided Wilcoxon signed-rank test. **g**, Venn diagrams showing the expressed genes in E-P interactions of Early-S and LateS-G2 in **e** (right). **h**, Partial gene ontology (GO) terms, enriched by specific genes of Early-S and LateS-G2 in **g**, respectively.

Corresponding author(s): Junjun Ding

Last updated by author(s): Jul 9, 2023

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Illumina bcl2fastq (v2.20.0).
Data analysis	There are details in the online Methods. All parameters used to analyze data are described in methods section and custom scripts are available upon request. Trim Galore (v 0.6.6), HiC-Pro (v 3.0.0), GENOVA (v 1.0.0), hicexplorer (v 3.7.1), pyGenomeTracks (v 3.6), Juicebox (v 1.11.08), HOMER (v 4.11), TopDom (v0.0.2), Seurat (v 4.1.0), Higashi (https://github.com/ma-compbio/Higashi), sci-CAR analysis (https://github.com/JunyueC/sci-CAR_analysis). The custom scripts for the analysis are available from https://github.com/jsun9003/scCARE-seq.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequencing data have been deposited at the NCBI GEO with accession number GSE211395, using mm10 reference genome. Other public datasets used in this study were downloaded from NCBI GEO with accession numbers as follows: ChIP-seq (GSE90895; CTCF and H3K27ac), In situ Hi-C (GSE124342), 2013-Nagano (GSE48262), 2017-Flyamer (GSE80006), 2017-Nagano (GSE94489), 2017-Steven (GSE80280), 2019-Tan (GSE121791), 2021-Tan (GSE162511), sci-CAR (GSE117089), SNARE-seq (GSE126074), CoTECH (GSE158435), Paired-Tag (GSE152020) and Paired-seq (GSE130399). Source data are provided with this paper.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was determined according to previously published datasets with similar experiments (Xiong et al., Nature methods, 2021; Zhu et al., Nature methods, 2021; Cao et al., Science, 2018). Two biological replicates for our methods are sufficient for assessing the data quality.
Data exclusions	No data were excluded from the analysis.
Replication	Two biological replicates for our study libraries were used from independent experiments. All datasets from independent experiments have similar results.
Randomization	Allocation was random.
Blinding	The experiments were not blinded. Because the 2i mESCs and serum mESCs need to be used to compare and evaluate the sensitivities and specificities of scCARE-seq.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	The HEK293T cell line was a gift from Dr. Jianlong Wang and mESCs lines used for this study were R1 (a kind gift from Dr. Shaorong Gao, PMID: 33357405).
Authentication	Cells lines are confirmed by a typical round shape ESCs morphology with small and tightly packed cells, and a high nucleus-to-cytoplasm ratio.
Mycoplasma contamination	Cell lines are negative for Mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.