Article

Tensor-FLAMINGO unravels the complexity of single-cell spatial architectures of genomes at high-resolution

Hao Wang[1], Jiaxin Yang[1], Xinrui Yu [1], Yu Zhang [2] ✉, Jianliang Qian [1,3] ✉ & Jianrong Wang [1] ✉

The dynamic three-dimensional spatial conformations of chromosomes demonstrate complex structural variations across single cells, which plays pivotal roles in modulating single-cell specific transcription and epigenetics landscapes. The high rates of missing contacts in single-cell chromatin contact maps impose significant challenges to reconstruct high-resolution spatial chromatin configurations. We develop a data-driven algorithm, Tensor-FLA-MINGO, based on a low-rank tensor completion strategy. Implemented on a diverse panel of single-cell chromatin datasets, Tensor-FLAMINGO generates 10kb- and 30kb-resolution spatial chromosomal architectures across individual cells. Tensor-FLAMINGO achieves superior accuracy in reconstructing 3D chromatin structures, recovering missing contacts, and delineating cell clusters. The unprecedented high-resolution characterization of single-cell genome folding enables expanded identification of single-cell specific long-range chromatin interactions, multi-way spatial hubs, and the mechanisms of disease-associated GWAS variants. Beyond the sparse 2D contact maps, the complete 3D chromatin conformations promote an avenue to understand the dynamics of spatially coordinated molecular processes across different cells.

Because it is the structural foundation of diverse molecular processes, the spatial conformation of chromosomes in the three-dimensional (3D) space plays important roles in modulating gene regulation[1], epigenetic plasticity[2], genome stability[3] and cell differentiation[4]. Comprehensive profiling of the 3D chromosomal structures across different cell types and species[5–8] using genome-wide 3C-based technologies, such as Hi-C[6], Capture-C[9,10] and ChIA-PET[11], has shed light on the patterns of genome folding, including chromatin compartments[12,13], topologically associated domains (TADs)[14] and chromatin loops[15]. These multi-scale conformation units and their associated signatures[16–18], such as CTCF and the cohesin complex[15,19], have led to specific biological models of the underlying principles of the 3D genome organization, including the loop extrusion[15,20] and phase separation[21,22] models. Quantitative analyses of the 2D matrices

of chromatin contact maps have improved the insights into the basic patterns of chromatin configuration and loops (e.g. SnapHiC, Fit-Hi-C and DeepLoop)[23–26]. Moreover, computational reconstruction of the chromosomal architecture in 3D space (such as the pioneering landmark algorithms MCMC5C and BACH that defined this topic)[27–34], based on pairwise spatial distances converted from chromatin contact frequencies between genomic regions[12], has further advanced the understandings of spatially orchestrated regulatory interactions, including the long-range ones (>100 kb), linking specific genes, enhancers, transcription factors, open chromatin loci, and distal genetic variants into 3D spatial neighborhood[35].

Despite these biological discoveries enabled by Hi-C experiments, the bulk level chromatin contact frequency maps represent averaged signals across large numbers of cells, and most existing computational

[1]Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI 48824, USA. [2]Department of Microbiology, Genetics, and Immunology, Michigan State University, East Lansing, MI 48824, USA. [3]Department of Mathematics, Michigan State University, East Lansing, MI 48824, USA. ✉e-mail: yzhang22@msu.edu; jqian@msu.edu; wangj164@msu.edu

algorithms are designed to predict the consensus structures[27–32]. Hence, these experiments and computational tools are unable to capture the cell-to-cell structural variabilities of single-cell chromatin. Although there are algorithms proposed to predict structure ensembles (i.e. populations of chromatin structures) using either machine learning or simulation approaches[36–38], significant difficulties exist in justifying the validity of predicted ensembles, distinguishing noisy fluctuations from biological variations, verifying the sensitive influence of prior assumptions and parameters, and interpreting individual structures within the inferred ensembles. Although ensemble-based methods can predict groups of structures, the interpretations and insights are limited to variance analyses of the overall distribution landscape, usually at low-resolution and for relatively homogeneous pools of cells. The predicted structure ensemble cannot be explicitly assigned to individual cells, restricting in-depth comparative 3D genome analysis between specific cells. The lack of explicit single-cell specific structures of ensemble methods also limits the functional genomics research for single-cells, such as annotations of single-cell specific long-range regulatory interactions, gene regulation, and long-range genetic associations. Due to these restrictions, large-scale quantitative modeling methods of the single-cell spatial genome configurations are currently under-developed.

Recent years have seen a breakthrough in single-cell 3C-based technology of profiling genome-wide chromatin contacts from individual cells (Fig. 1a). These experimental approaches include single-cell Hi-C (scHi-C)[39–41], single-cell combinatorial indexed Hi-C (sciHi-C)[42,43], single-nucleosome Hi-C (snHi-C)[44], diploid chromatin conformation capture (Dip-C)[45,46], single-nucleus methyl-3C sequencing (snm3C)[47], and Methyl-HiC[48]. Applied on a mixed pool of cells, these techniques can generate single-cell specific chromatin contact maps for a large number of cells, which may belong to heterogeneous cell types within the sample. Analyses of the single-cell specific chromatin contact maps profiled by these techniques revealed large-scale cell-to-cell structural variabilities of TADs and chromatin loops and also demonstrated chromatin reorganization through the processes of cell cycles or cell differentiation[39–48]. These valuable resources have also helped the delineation of cell types and cell-type specific chromatin contacts from the mixed pool of cells[42–47], which further facilitates the investigation of spatially coordinated gene expression, regulatory elements (e.g. enhancers), and epigenetic signals (e.g. DNA methylation)[45,47]. However, because of the low sequencing depth allocated for each individual cell, the single-cell chromatin contact maps are much sparser than bulk-level experiments and contain numerous entries with missing data[39–48]. The sparsity issue is even more challenging at higher resolutions. Taking the single-cell Dip-C dataset of GM12878 as an example[45], even at the low 300kb-resolution, 94.3% entries of the single-cell level chromatin contact maps are missing data and do not have any observed signals based on the experiment. Moreover, at the high 10kb-resolution, the missing rate is >99.9%. Thus, the high missing rates of single-cell chromatin contacts, especially at higher resolution, raise the unique demand for advanced computational models to efficiently extract the information of chromatin conformation from limited observations and to generate accurate 3D structure predictions for individual cells.

Impeded by the challenge of sparsity, the available computational infrastructure of predicting single-cell 3D chromosomal structures is lacking and is mostly confined to only low-resolution reconstructions, which are not capable of characterizing the detailed transcriptional regulation and QTL associations in 3D coordination. As data-driven approaches, RPR[30] has been developed to predict single-cell 3D chromosomal structures using the method of recurrent plots, and ShRec3D[27], which was originally developed for bulk-level predictions, has also been suggested to handle single-cell Hi-C data using an MDS-based approach. Relying solely on the observed data of chromatin structures, these methods alleviate the dependence on specific prior

assumptions of genome structures. However, due to the significant fraction of missing data in single-cell chromatin contact maps, their reconstruction accuracy degrades significantly for high-resolution predictions (e.g. 10kb- or 30kb-resolution). To improve the performance at higher resolutions, specific prior assumptions about chromosomal structures and potential mechanisms are explored to complement the highly sparse input data, which lead to polymer simulation-based approaches, including isdHi-C[49], Si-C[50] and NucDynamics[40]. With the observed chromatin contacts of different cells serving as the optimization constraints, these three algorithms simulate single-cell 3D chromosomal structures based on a variety of pre-assumed biophysical properties[40,49,50], which are themselves susceptible to more investigations for validation and may not hold across different datasets. In addition, the strong pre-assumed biophysical properties, along with the pre-defined parameters, join into the models as invariants and dominate the inference procedure, which effectively smoothen out the cell-to-cell variabilities, resulting in decreased ability of capturing single-cell specific structural signatures.

Despite the computational difficulties, high-resolution 3D reconstruction of single-cell chromosomal structures is highly needed and is crucial to promote the understanding of spatial organizations of gene regulation to a greater extent from the single-cell perspective[39–42], as also demonstrated in this study. Compared to traditional analyses of 2D matrices of pairwise chromatin contacts, 3D characterization of spatial configurations yields crucial information about the genome architectures. Advantages of 3D genome modeling against 1D or 2D analyses include, but are not limited to, 1) identifying multi-way chromatin interactions beyond pairwise contacts and 2) producing completed chromatin contact maps and imputing numerous missing contacts based on the global structural information of whole chromosomes instead of the local contexts. Both advantages will facilitate the annotation of the TAD dynamics across single cells, enable the discovery of detailed higher-order chromatin organization signatures (e.g. spatial hubs), and boost the delineation of single cell clusters. In addition, compared to low-resolution modeling, high-resolution reconstruction is essential to generate specific biological understandings, such as pinpointing the interdependent long-range cis-regulatory interactions of genes and interpreting the functionally relevant genetic variants of complex diseases, especially distal non-coding variants.

In this study, to expedite biological insights into the dynamic single-cell 3D genomes, a low-rank tensor completion based model, Tensor-FLAMINGO, is developed to reconstruct high-resolution 3D chromosomal structures based on single-cell chromatin contact maps (Fig. 1a). The key feature of Tensor-FLAMINGO is to systematically utilize the low-rank property of the observed single-cell chromatin contact maps, which can efficiently address the critical challenge of high missing rates in input data. Moreover, the low-rank property holds its validity for diverse single-cell chromatin contact experiments and datasets because of two fundamental facts: 1) for a given sample, single cells belonging to each cell cluster are biologically related or similar to each other; and 2) the chromatin contacts in each cell are generated from the underlying low-rank 3D structure. By quantitatively exploring these two types of low-rank properties, the low-rank tensor completion technique is guaranteed to recover the variable single-cell specific 3D chromosomal structures at high-resolution from a small set of observed pairwise chromatin contacts. Moreover, the sophisticated tensor completion framework employed by Tensor-FLAMINGO simultaneously delineates the latent sub-clusters of cells, which demonstrate distinct chromosome structures across sub-clusters. Therefore, Tensor-FLAMINGO can effectively handle the structural heterogeneity among the mixed pool of different cells, which is also justified in this study (see Results). Analogous to video compression and reconstruction, for which the low-rank tensor completion methods were originally designed[51], consider every single cell's
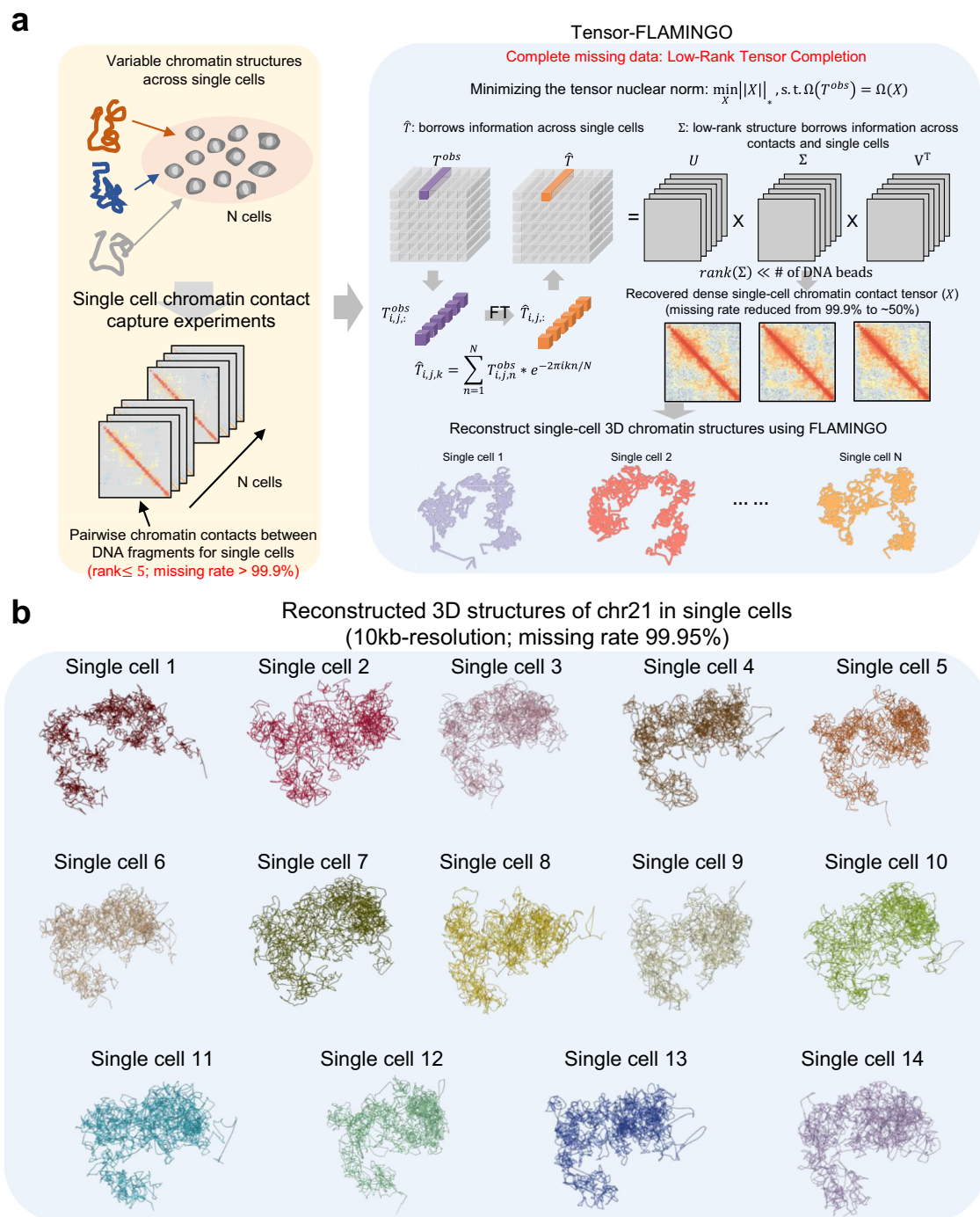
Fig. 1 | Tensor-FLAMINGO reconstructs high-resolution 3D chromatin structures from single-cell Hi-C experiments. a Schematic figure of the Tensor-FLAMINGO algorithm. Biologically, scHi-C experiments generate highly sparse contact maps for N cells. For every single cell, the chromatin contact map derived from the scHi-C experiment is a low-rank matrix (rank ≤ 5). In addition, single cells belonging to similar cell types are biologically similar. Thus, the tensor organizing the chromatin contact matrices of N cells is a low-rank tensor and the missing values can be completed using the low-rank tensor completion method based on the small subset of observed values. To accurately reconstruct the 3D chromatin structures of single cells at high resolution, Tensor-FLAMINGO utilizes the tubewise Fourier Transformation to borrow information across single-cells, and employs tSVD to borrow information across contacts within the same cells. Based on the completed dense tensor of single-cell chromatin contact maps, Tensor-FLAMINGO leverages FLAMINGO to reconstruct the 3D chromatin structure for every individual cell. **b** Reconstruction of the 3D chromatin structures for 14 single cells at 10kb-resolution by Tensor-FLAMINGO based on the GM12878 Dip-C data (missing rate = 99.95% at 10kb-resolution). Source data are provided as a Source Data file.

chromatin contact map as a frame and the whole dataset as a video, the low-rank tensor completion strategy leverages the relationship across and within frames and facilitates information-sharing to infer missing pixels and to capture the underlying latent structures of each frame. To note, the low-rank tensor completion technique based on the t-SVD

framework (see Methods) enables the unique capability of Tensor-FLAMINGO to address the large sparsity of single-cell chromatin contact datasets, beyond bulk datasets, which is the fundamental advantage of Tensor-FLAMINGO compared to our prior model of FLAMINGO.

To demonstrate its superior performance, Tensor-FLAMINGO (https://github.com/wangjr03/Tensor-FLAMINGO) is implemented on a variety of single-cell chromatin conformation datasets generated by different types of experiments, including scHi-C, snHi-C, Dip-C and snm3C, and reconstructed the 3D chromosomal structures for individual cells at both 10kb- and 30kb-resolution, the highest resolutions for single-cell level predictions to date. The accuracy of predicted structures is systematically evaluated and compared to other methods based on both simulations and extensive experimental datasets as orthogonal gold-standards. Tensor-FLAMINGO robustly outperforms existing methods across all comparisons. Based on the reconstructed 3D structures, completed chromatin contact maps are also produced simultaneously for each cell, where the missing data of single-cell chromatin contacts are explicitly imputed at high-resolution. The completed single-cell chromatin contact maps substantially enhance the clustering of different groups of cells, boosting the deconvolution of mixed cell types or subtypes. The inferred single-cell specific 3D genome configurations capture the cell-to-cell structural variabilities. They facilitate the analysis of chromatin dynamics with respect to the formation of chromatin compartments and TADs in single cells and their cross-cell variances coupled with the landscape of transcription and epigenetic states. Moreover, Tensor-FLAMINGO enables the identification of high-resolution significant chromatin interactions in single cells, the majority of which cannot be observed in the original sparse contact maps. These dynamic single-cell chromatin interactions also uncover long-range links mediating specific distal cis-regulatory elements to 3D spatial proximity for single-cell gene regulation, allowing mechanistic interpretations of eQTLs and disease-associated non-coding genetic variants at the single-cell level. The discoveries of single-cell multi-way chromatin interactions and the spatial hubs of cell-type specific genes further reveal the interplay between higher-order genome folding and gene expression, leading to biological insights into the spatial organization of transcriptional regulation and its dynamics.

Results

Tensor-FLAMINGO reconstructs high-resolution 3D chromosome structures for single cells

For a given dataset of single-cell 3 C experiment[39–48], the observed chromatin contacts from tens to hundreds of cells are summarized into a sparse 3-mode tensor as input, where each frontal slice of the tensor represents the chromatin contact map of a single cell and the majority of entries in the tensor are un-observed missing data due to limited sequencing depth per cell (Fig. 1a left). The genome of each cell has its own 3D spatial conformation featured by single-cell specific structural variabilities. Every chromosome in a single cell is modeled as a chain of 'beads', based on the 'beads on a string' polymer model, where each 'bead' represents a genomic locus under the specified resolution. The observed chromatin contact map of a single cell is biologically induced from the underlying 3D coordinates of different genomic loci. To reconstruct single-cell 3D spatial structures of each chromosome, Tensor-FLAMINGO employs a two-step strategy based on a low tubal-rank tensor completion framework (see Methods). In the first step, the objective function aims to recover a dense tensor with the minimum tubal rank, while optimally maintaining the consistency with the set of observed single-cell chromatin contacts from the input (Fig. 1a right). An Alternating Direction Method of Multipliers (ADMM) based tensor tubal-rank-minimization algorithm[51] is implemented. In each iteration, the low-rank latent structure is updated using the tensor-SVD (t-SVD) approach[51], which first transforms the input tensor to the Fourier domain and then carries out SVD analysis (see Methods). This strategy enables simultaneous modeling of single cells jointly, instead of separate analyses as done by previous methods[27], by exploring the low-complexity relationship between cell groups. Through optimization, the t-SVD approach in each iteration

enables the model to use information both across different cells and across different chromatin contacts within a cell (Fig. 1a right). The output generated by the first step is a dense low tubal-rank tensor, with explicitly imputed signals of high-resolution chromatin contacts for large portions of missing data in each cell. As a highlight, the t-SVD approach employed by Tensor-FLAMINGO to generate the dense low tubal-rank tensor is the critical algorithmic innovation to overcome the fundamental challenge of highly sparse single-cell chromatin contact data. The second step of Tensor-FLAMINGO takes the recovered dense tensor as input and employs our recently developed algorithm FLA-MINGO to further reconstruct single-cell 3D spatial structures, based on an extended low-rank matrix completion technique[35] (Fig. 1a right). With its performance already justified before[35], FLAMINGO is applied on each frontal slice of the recovered dense tensor from the first step to predict the final output of high-resolution 3D chromosomal structures of each single cell, along with the completed single-cell chromatin contact maps derived from the predicted 3D structures (see Methods). To note, FLAMINGO by itself cannot be directly applied on the original single-cell chromatin contact dataset due to the high sparsity at the single-cell level. The dense low tubal-rank tensor generated by the first step of t-SVD tensor completion coherently integrates with FLAMINGO, leading to the combined performance advantage of Tensor-FLAMINGO.

The key design of Tensor-FLAMINGO is to quantitatively exploit the low-complexity structures of the sparse single-cell chromatin contact tensor. Two sources of pervasive inter-dependencies among observed signals, both of which are fundamental facts holding their validity for all different types of single-cell 3 C techniques, guarantee the underlying low-rank signatures. First, within a mixed pool of single cells, the cells can be grouped into multiple clusters corresponding to different cell types or subtypes, and hence, the cells in each cluster are biologically related to each other and share similar cluster-specific consensus structures. For a specific pairwise chromatin contact between two genomic loci across different cells, i.e. a specific 'tube' of the input tensor, its missing value in one cell can be inferred from other cells belonging to the same cluster (Fig. 1a right). Because of this tube-wise dependency across single cells, the first step of Tensor-FLAMINGO jointly models the cells together. Algorithmically, in each iteration, the t-SVD approach conducts tube-wise Fourier transformation on the observed tensor, with subsequent steps carried out in the Fourier transformed domain, allowing information sharing across different cells. Second, because the numerous pairwise distances between genomic loci are induced from the underlying 3D spatial coordinates of each locus, the matrix rank of each single cell's chromatin contact map is at most five, as we have previously justified in the design of FLAMINGO[35]. Therefore, the intrinsic degrees of freedom of single cell chromatin contact maps are very small compared to the size of the matrix, suggesting that the pairwise chromatin contact information can be significantly compressed and can be efficiently reconstructed based on just a small set of observed entries (Fig. 1a right). Algorithmically, in the first step, the soft-thresholding SVD analysis of each transformed frontal slice in the Fourier domain explicitly explores the latent low-rank dependency to recover a dense tensor with the minimum tubal rank (see Methods). Additionally, in the second step, Tensor-FLAMINGO implements FLAMINGO to further leverage the dependency among chromatin contacts within each cell, based on a low-rank matrix completion strategy (see Methods), leading to the final prediction of high-resolution 3D spatial coordinates of chromosomes in each cell. The superior and robust performance of FLAMINGO against high rates of missing data is especially desired for high-resolution reconstructions since the read-depth allocated in each high-resolution genomic locus is much lower and sparser. Overall, by delineating the two sources of low-complexity latent structures corresponding to the low tubal-rank of the 3-mode tensor and the low rank of the chromatin contact map, Tensor-FLAMINGO thoroughly addresses the core

challenge of sparsity of single-cell 3C-based experiments and significantly promotes single-cell 3D genome modeling, without demanding increased experimental burden of sequencing depths.

Tensor-FLAMINGO is systematically implemented on four single-cell 3C-based datasets profiled by different techniques, including scHi-C[40], snHi-C[44], Dip-C[45] and snm3C[47], to predict single-cell 3D chromosomal structures at both 10kb- and 30kb-resolution. These predictions represent the largest cohort of spatial conformations of 3D genomes for single cells at the highest resolution to date (Fig. 1b, Supplementary Figs. 1–3). As an example, the single-cell Dip-C experiment of GM12878[45], after quality control, profiled the chromatin contact maps for 14 individual cells. As a representative example, based on this Dip-C dataset, Tensor-FLAMINGO successfully reconstructed the 10kb-resolution 3D structures of chromosome 21 for all 14 cells (Fig. 1b), with more than two-fold of boost in accuracy compared to other methods (average Spearman correlation to the benchmark >0.42, (Fig. 4a). The overall foldings of the reconstructed 10kb-resolution single-cell specific structures highly align with the consensus structure inferred from bulk Hi-C data of GM12878 (Supplementary Fig. 16c). In addition to the overall consensus of chromosome folding, all 14 predicted structures show

single-cell specific structural variabilities (Fig. 1b), which are further quantified with respect to distinct TAD formation, specific long-range chromatin interactions involved in gene regulation, and multi-way interaction hubs across different cells (Fig. 8a–i). To note, these detailed evaluations of single-cell structural variabilities are made feasible only by the high-resolution reconstructions, while low-resolution predictions are not sufficient to depict the specific variations in 3D genomes. Remarkably, Tensor-FLAMINGO can build the high-resolution spatial structures based on the highly sparse Dip-C data, which has a missing rate >99.95% at 10kb-resolution (Fig. 1b).

Benchmark performance based on simulations

The performance of Tensor-FLAMINGO is first benchmarked by a series of simulated pools of single-cell structures, for which the gold-standard structures are known (Fig. 2a, Supplementary Fig. 4a). In each pool of simulations, multiple cell types are mixed together and each cell type has its cell-type specific consensus chromosome structure, in order to account for the underlying heterogeneity of single-cell samples. Based on the consensus structure of each cell type, a number of single-cell specific chromosome structures and their corresponding
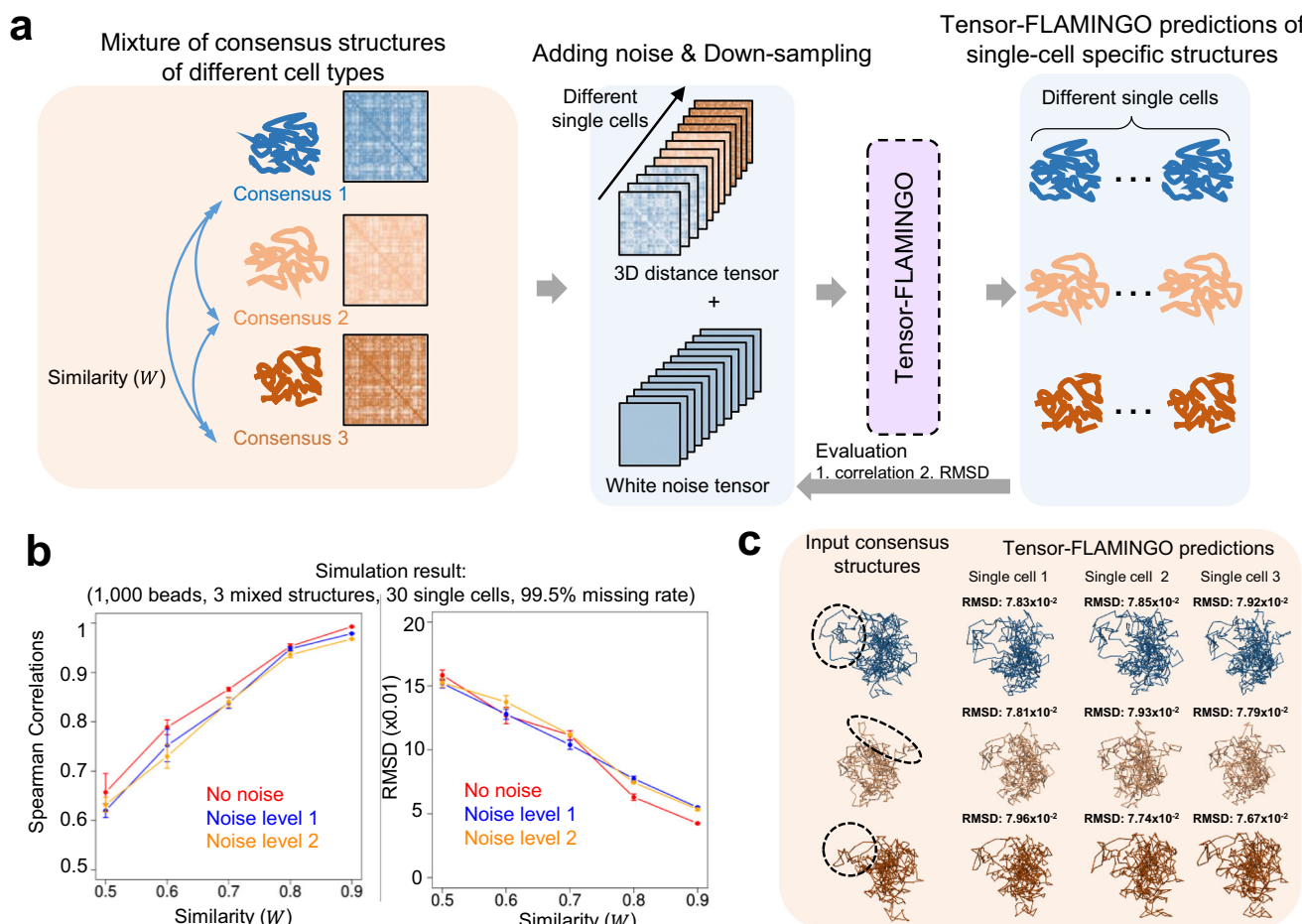

Fig. 2 | Performance evaluation of Tensor-FLAMINGO based on a diverse panel of simulations. a Schematic figure of the simulation procedure. Three consensus structures are included with parameter W controlling the similarity between consensus structures. Each consensus structure is repeated ten times and mixed with different levels of noise (no noise, noise level 1, and noise level 2) to generate single-cell specific benchmark structures. The corresponding single-cell distance matrices are subsequently down-sampled to simulate a sparse dataset of 30 single cells. The highly noisy and incomplete sparse tensor is used as the input for Tensor-FLAMINGO to reconstruct the single-cell specific structures. The accuracy is

evaluated by calculating the correlations and RMSE, compared to the original simulated single-cell benchmark structures. **b** Performance of Tensor-FLAMINGO under different values of similarity (W), which quantifies the heterogeneity across the underlying mixed cell types, and different noise levels (1000 beads, and 0.5% down-sampling rates = 99.5% missing rates). The error bars represent the standard deviations across frontal slices (i.e. single cells). **c** Examples of the consensus benchmark structures and Tensor-FLAMINGO's single-cell predictions for the three mixed cell types ($W = 0.6$), where the cell-type specific 3D structures are reconstructed correctly (correlations = 0.783, RMSD = 0.133).

chromatin contact maps are generated by randomly down-sampling of chromatin contacts and by adding white noise to represent both experimental missing data rates and single-cell specific structural variabilities (Fig. 2a), see Methods. A combination of simulation parameters is systematically varied to generate different pools of simulated structures and chromatin contact maps, which are used to thoroughly test the performance of Tensor-FLAMINGO under a diverse panel of conditions (see Methods). Prediction accuracy is quantified by two metrics, i.e. Spearman correlations and RMSD values, with respect to the benchmark single-cell structures (see Methods), both of which have been suggested by previous studies as quantitative performance indicators[27,49,50].

Remarkably, Tensor-FLAMINGO robustly achieves high accuracy of single-cell 3D structure reconstructions under a wide range of structural heterogeneities among the mixed cell types, with respect to performance metrics of both Spearman correlations (>0.6) and RMSD (<0.16) to the real benchmark structures (Fig. 2b). Lower structural similarity among mixed cell types corresponds to higher heterogeneity within the pool of single cells and causes more stringent challenges to recover the structures of individual cells belonging to different cell types. In addition, the accuracy is robust to the different levels of noise added to each single cell (Fig. 2b), which represent different degrees of cell-to-cell variabilities within each cell type. As an example of Tensor-FLAMINGO's performance, (Fig. 2c) shows the representative predictions when it is applied to a pool of simulated heterogeneous single-cell chromatin contact maps where three underlying cell types are mixed. Additional predictions are shown in Supplementary Fig. 4b. The consensus structures of the three mixed cell types contain distinct cell-type specific 3D structure signatures (Fig. 2c,) and ten variable single-cell structures are simulated per cell type based on the consensus structures. Although challenged by the two levels of structural variabilities (i.e. cell-type level and single-cell level), Tensor-FLAMINGO successfully reconstructed the single-cell 3D structures with high accuracy (RMSD < 0.08) and recovered the cell-type specific chromatin folding signatures (Fig. 2c and Supplementary Fig. 4b). The predicted single-cell structures also form three clear clusters, which are found to be consistent with each cell's membership to the three underlying mixed cell types (Supplementary Fig. 4c). These striking results strongly support Tensor-FLAMINGO's capability of delineating distinctive 3D structures for individual cells from heterogeneous datasets.

Furthermore, tested on different pools of simulated structures with a wide range of down-sampling rates (≥0.5%), Tensor-FLAMINGO consistently demonstrates high reconstruction accuracy Spearman correlations>0.97, RMSD < 0.0018, (Supplementary Fig. 5). Even with a down-sampling rate of 0.1% (that is 99.9% missing data), Tensor-FLAMINGO can still achieve high reconstruction accuracy Spearman correlations>0.6, RMSD < 0.23, (Supplementary Fig. 6). These results systematically suggest its superior performance against high rates of missing data, which is the fundamental challenge of sparse single-cell chromatin contact datasets. In addition, the robustness of Tensor-FLAMINGO's performance is further extensively justified across diverse simulated datasets with different chromosome sizes (Supplementary Fig. 7a), different numbers of single cells for each mixed cell type (Supplementary Fig. 7b), different number of cell types mixed in the sample (Supplementary Fig. 7c), and different order of organization of frontal slices in the chromatin contact tensor (Supplementary Fig. 7d). Furthermore, we also tested the low-rank property of the simulated pool of single-cell contact map tensor with high-level of noise added to the inputs. The singular values of the noisy input tensor can be robustly estimated where only the few top-ranking singular values are non-zero, suggesting that the low-rank property can be effectively captured by Tensor-FLAMINGO (Supplementary Fig. 7e).

Superior performance of reconstructing single-cell 3D structures

To directly compare the performance on real experimental data, the image-based K562 single-cell structures of a 2 Mb region in human chromosome 21 (chr21:29,372,390-31,322,257) from the STORM dataset[52] are collected as the benchmark testing structures. These image-based experimental data provide 30kb-resolution structures across a large number of cells. Tensor-FLAMINGO is applied on the snHi-C dataset of K562 cells[44], which contains single-cell chromatin contact maps of 16 cells, to reconstruct the 30kb-resolution 3D spatial configurations of chromosome 21. As comparison, the state-of-the-art algorithms, including Si-C[50], isdHiC[49] and RPR[30], are also applied on the same snHi-C dataset based on their suggested settings (see Methods). Since the STORM dataset is profiled from a different pool of K562 cells than the snHi-C dataset, the predicted 3D structures by each algorithm are first evaluated with respect to the number of highly correlated (Spearman correlation>0.8) STORM single-cell structures, which indicates the degree of supports from a de-coupled pool of single cells belonging to the same cell type. Examples of Tensor-FLAMINGO's single-cell predictions and the supporting structures from the STORM dataset are shown in Fig. 3a, where Tensor-FLAMINGO characterizes the complicated chromatin folding of this specific 2 Mb region consistent with the observations from the image-based STORM dataset. Strikingly, for each of the 16 cells in the snHi-C dataset, Tensor-FLAMINGO's predictions are supported by an average of 73 image-based single-cell structures from the STORM dataset, while the state-of-the-art methods are supported by much fewer structures (Fig. 3b).

Moreover, the image-based STORM dataset is further filtered to identify the subset of top-matching single-cell structures that are highly aligned to each of the 16 experimental snHi-C input chromatin contact maps as the benchmark structures for specific cells (see Methods). Based on these unbiased benchmark structures of each cell, Tensor-FLAMINGO demonstrates the highest accuracy across all tested methods (Fig. 3c). The predicted single-cell chromatin structures by Tensor-FLAMINGO are strongly correlated to the benchmark STORM structures for each cell (median Spearman correlation=0.56), while the other state-of-the-art algorithms show much lower correlations to the benchmarks (0.25<median Spearman correlations<0.5). As a representative example, (Fig. 3d) shows the aligned image-based 3D structure and its associated pairwise distance matrix for a specific cell from the snHi-C dataset. The predicted 3D conformation by Tensor-FLAMINGO accurately recapitulates this structure Spearman correlation=0.85, (Fig. 3d), while other algorithms lead to large deviations from the benchmark and only achieve much lower accuracy (Spearman correlations<0.4). Overall, these performance comparisons benchmarked by the image-based single-cell structures directly support the drastic improvement in 3D reconstruction accuracy by Tensor-FLAMINGO at high resolution.

Robust accuracy across diverse experimental contexts

The performance of Tensor-FLAMINGO, when it is applied on a variety of input datasets profiled by different single-cell experimental techniques (i.e. scHi-C, snHi-C, Dip-C, and snm3C), is further evaluated by the corresponding orthogonal benchmark evidence from matched cell types, including Hi-C[6,53], 3D ATAC-PALM[54], and GAM[55] data (see Methods). Based on extensive evaluations across different cell types, including GM12878, K562, and mESC, Tensor-FLAMINGO consistently achieves superior accuracy compared to the state-of-the-art algorithms in reconstructing 3D chromosome structures at both 10kb-resolution (Fig. 4a) and 30kb-resolution (Fig. 4b). At 10kb-resolution, the highest resolution for single-cell 3D reconstruction to date, Tensor-FLAMINGO's predictions robustly demonstrate high correlations to the benchmark structures across all tests (Fig. 4a) top, Spearman correlation>0.4). In comparison, other models show much lower accuracy, with correlations to the benchmark structures largely below
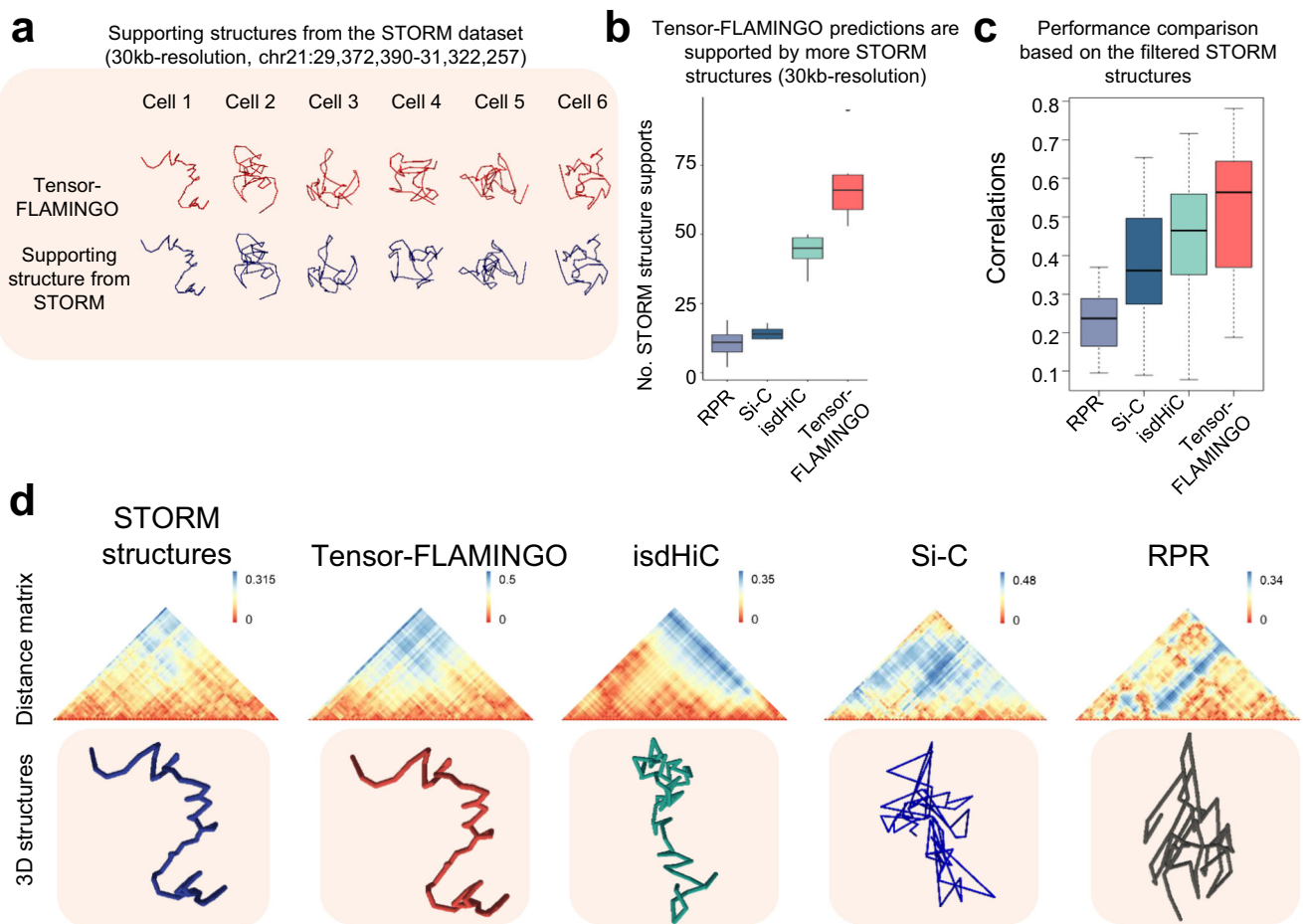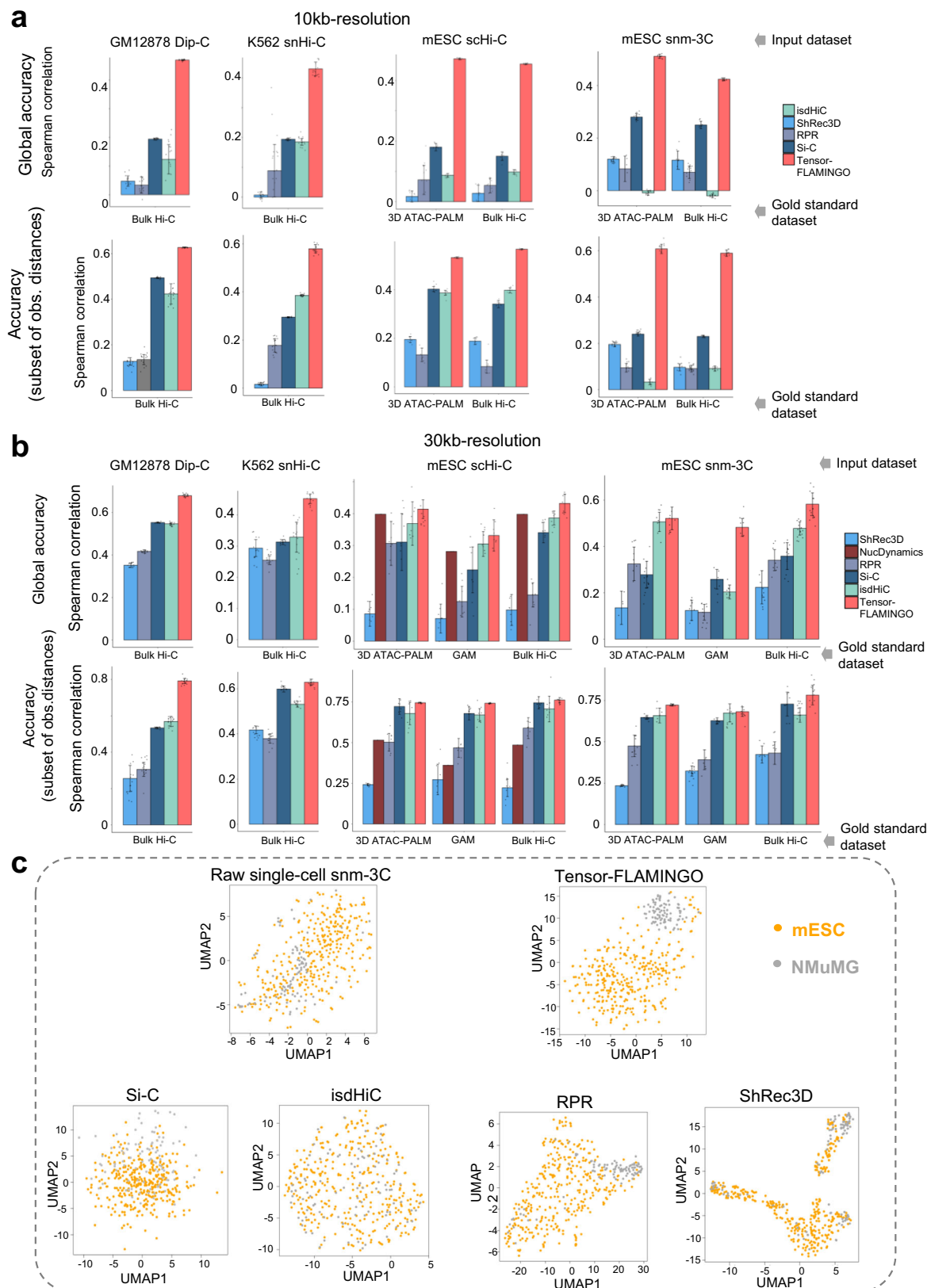
Fig. 3 | Superior accuracy of Tensor-FLAMINGO benchmarked by multiplexed STORM image-based single-cell 3D structures. Tensor-FLAMINGO, RPR, Si-C and isdHi-C are applied on the K562 snHi-C data of 16 single-cells at 30kb-resolution to reconstruct the single-cell 3D chromatin structures. **a** Examples of predicted single-cell 3D chromatin structures using Tensor-FLAMINGO and the supporting structures from the STORM image-based dataset. **b** The number of image-based structures from the STORM dataset supporting the predicted single-cell structures from each algorithm. The STORM single-cell structures with correlations>0.8 to the predicted structures ($n = 16$) are considered as the supporting structures. The center lines of boxplots show the median; the upper and lower box limits show the 25th and 75th percentiles, respectively. The whiskers extend up to 1.5 times the interquartile range away from the limits of the boxes. **c, d** Tensor-FLAMINGO accurately reconstructs the single-cell 3D structures from snHi-C data. For each cell

in the snHi-C dataset, the correlations between the original input snHi-C distance matrix and the STORM distance matrices are calculated and the top 20 STORM structures with the highest correlations, top-matching STORM structures, are filtered to represent the true underlying structures of the specific cell. **c** Tensor-FLAMINGO predictions show the highest correlations with the top-matching STORM structures of each cell. The center lines of boxplots show the median; the upper and lower box limits show the 25th and 75th percentiles, respectively. The whiskers extend up to 1.5 times the interquartile range away from the limits of the boxes. **d** Examples of the predicted 3D chromatin structure based on snHi-C for single cell 1 and the top-matching STORM structure. Tensor-FLAMINGO shows the highest correlation with the top-matching STORM structure (correlation=0.676). Source data are provided as a Source Data file.

0.2. Notably, the 2-fold improvements in these global accuracy evaluations suggest that Tensor-FLAMINGO not only recapitulates the observed chromatin contacts in single cells but also recovers the spatial distances for un-observed missing values in the original single-cell chromatin contact maps. Restricted to the subset of observed single-cell chromatin contacts in the original input datasets, Tensor-FLAMINGO also consistently achieves the highest accuracy (Fig. 4a) bottom, Spearman correlation>0.55, see Methods compared to other methods (Spearman correlations<0.4). The relative contributions of the t-SVD-based tensor completion step and the FLAMINGO-based matrix completion step to the overall performance of Tensor-FLAMINGO are also evaluated respectively, both of which are found to be important (Supplementary Fig. 7f and g). Similar advantages in accuracy are observed for Tensor-FLAMINGO across all comparisons for predictions at 30kb-resolution (Fig. 4b), see Methods, although the relative accuracy improvements over other models are less drastic than those for comparisons of 10kb-resolution predictions. This

extensive panel of tests provides quantitative supports of the significant boost in performance by Tensor-FLAMINGO, especially at higher resolutions, which is due to its algorithmic design of exploring the ground truth of the low-rank structures embedded in all different types of single-cell chromatin contact experiments, instead of relying on specific pre-assumed polymer models.

Furthermore, the predicted 3D chromosome structures by different algorithms are evaluated with respect to characterizing cell-type specific structural variations and delineating different cell types mixed within a sample, which are two of the core biological goals of single-cell experiments. The snm3C dataset[47] is used for this evaluation since it contains 351 mESC cells and 96 NMumG cells. Although these cells are mixed together, the identity of each single cell is known. Because a large majority of single-cell chromatin contacts (>99.9%) are missing data in the original dataset at 30kb-resolution, the snm3C data cannot directly distinguish the two cell types when the original single-cell chromatin contact maps are

a 10kb-resolution

GM12878 Dip-C | K562 snHi-C | mESC scHi-C | mESC snm-3C

Legend: isdHiC, ShRec3D, RPR, Si-C, Tensor-FLAMINGO

Global accuracy (Spearman correlation); Accuracy (subset of obs. distances) Spearman correlation

Datasets: Bulk Hi-C, 3D ATAC-PALM, Gold standard dataset, Input dataset

b 30kb-resolution

GM12878 Dip-C | K562 snHi-C | mESC scHi-C | mESC snm-3C

Legend: ShRec3D, NucDynamics, RPR, Si-C, isdHiC, Tensor-FLAMINGO

Global accuracy (Spearman correlation); Accuracy (subset of obs. distances) Spearman correlation

Datasets: Bulk Hi-C, 3D ATAC-PALM, GAM, Gold standard dataset, Input dataset

c

Raw single-cell snm-3C | Tensor-FLAMINGO | Si-C | isdHiC | RPR | ShRec3D

Legend: mESC (orange), NMuMG (gray)

projected onto the two-dimensional space by UMAP (Fig. 4c). In Fig. 4c, the two cell types (mESC and NMumG) merge together as one single cloud of cells. Similar problems remain when the completed single-cell chromatin contact matrices are used based on the predicted structures by other algorithms, including Si-C, isdHiC, RPR and ShRec3D, where the missing values are completed by the corresponding predicted spatial distances (Fig. 4c). The two cell types are still not separable based on the predictions from these state-of-the-art models. In contrast, the complete single-cell chromatin contact matrices from Tensor-FLAMINGO's predicted 3D structures clearly group the single cells into two separable clusters, which are highly consistent with the real cell identity of the two cell

Fig. 4 | Extensive performance comparisons at different resolutions based on diverse input datasets and gold-standard benchmarks support Tensor-FLAMINGO's superior performance in reconstructing single-cell 3D chromatin structures. a Performance comparison at 10kb-resolution across four different input single-cell datasets (GM12878 Dip-C dataset $n = 14$, K562 snHi-C dataset $n = 16$, mESC scHi-C dataset $n = 8$, and mESC snm3C dataset $n = 351$). The chromatin contact maps (Hi-C and 3D ATAC-PALM) profiled from the matching cell types are used as gold standards. Spearman correlations are used as the performance metric, including both global accuracy for all pairwise distances in the gold-standard datasets and the accuracy for the subset of pairwise distances also observed in the input datasets. The error bars represent the standard deviations across single cells. **b** Performance comparison at 30kb-resolution across four different input single-cell datasets (GM12878 Dip-C dataset $n = 14$, K562 snHi-C dataset $n = 16$, mESC scHi-

C dataset $n = 8$, and mESC snm3C dataset $n = 351$). The predictions of NucDynamics provided by Si-C are directly used. The chromatin contact maps (Hi-C, GAM and 3D ATAC-PALM) profiled from the matching cell types are used as gold standards. Spearman correlations are used as the performance metric, including both global accuracy for all pairwise distances in the gold-standard datasets and the accuracy for the subset of pairwise distances also observed in the input datasets. The error bars represent the standard deviations across single cells. **c** UMAP visualization of the reconstructed single-cell structures from the snm-3C dataset, which contains two cell types: mESC and NMuMG. Tensor-FLAMINGO's predictions capture the cell-type specific structural signatures and lead to clear separations between the two cell types, while other methods result in highly mixed cells of the two cell types. Source data are provided as a Source Data file.

---

types (Fig. 4c). This remarkable result further suggests the boosted accuracy of the predicted 3D structures and the completed single-cell chromatin contacts for missing data at high-resolution. It also underscores the advantage of Tensor-FLAMINGO, by jointly modeling single cells simultaneously, in capturing the cell-type specific structural variations and deconvolving heterogeneous types of cells.

Enhanced high-resolution imputation of single-cell chromatin contacts

The identification of high-resolution chromatin contacts in single cells is significantly impeded by the sparsity of single-cell datasets, hindering detailed analysis of chromatin dynamics and cis-regulatory interactions across different cells. The predicted single-cell 3D structures provide the information of all pairwise spatial distances along chromosomes, which can further generate the complete chromatin contact maps (see Methods). Therefore, Tensor-FLAMINGO can also serve as an efficient imputation algorithm for single-cell chromatin contacts at high-resolution. Unlike the state-of-the-art imputation method Higashi[56], Tensor-FLAMINGO leverages the global information of the whole chromosomal 3D structure to complete the missing values, instead of relying on local chromatin contexts.

To evaluate the performance of high-resolution imputation, both Tensor-FLAMINGO and Higashi are implemented on the snm3C dataset of 351 mESC cells[47]. The accuracy of imputed single-cell chromatin contact maps at 30kb-resolution from the two methods is benchmarked, respectively, by their correlations to orthogonal datasets of mESC[53–55] (i.e. 3D ATAC-PALM, GAM, and Hi-C). Tensor-FLAMINGO achieves substantially higher correlations (Spearman correlations>0.44) than Higashi across all comparisons using different benchmark datasets (Fig. 5a). As a representative example of imputed chromatin contacts for a single cell in the dataset, Tensor-FLAMINGO recovers clear TAD structures along the chromosome (chr19) in the completed single-cell chromatin contact map, while the imputed signals by Higashi do not demonstrate chromatin organizing patterns (Fig. 5b, Supplementary Fig. 8d). Furthermore, the recovered TAD boundaries by Tensor-FLAMINGO are largely supported by the ChIP-seq signals of CTCF and Rad21 binding clusters (Fig. 5b).

The imputed single-cell chromatin contacts are further evaluated with respect to their capability of capturing cell-type specificity, by applying Tensor-FLAMINGO and Higashi on the mixed snm3C dataset with 351 mESC cells and 96 NMuMG cells. Imputations are carried out and compared at different resolutions, including 1Mb-, 250kb- and 30kb-resolutions, and cell clusters are identified based on the imputed single-cell chromatin contact maps using k-means clustering. Adjusted Rand Index (ARI) is used as the performance metric to quantify the consistency between predicted cell-cluster memberships and the real cell-type identities of all cells (see Methods). Across all the three tested resolutions, Tensor-FLAMINGO consistently demonstrates higher accuracy of cell-type identification (Fig. 5c). Notably, at 30kb-resolution, the imputed signals by Tensor-FLAMINGO can still classify the single cells into their corresponding cell types with high accuracy

(ARI = 0.53), which is a significant improvement over Higashi (ARI = 0.13). As shown in Fig. 5d, based on the imputed single-cell chromatin contacts at low resolutions (1Mb- and 250kb-resolution), both Tensor-FLAMINGO and Higashi can classify the two mixed cell types into distinct clusters. However, at higher resolution (30kb-resolution), the imputed signals by Higashi cannot distinguish the two cell types, while Tensor-FLAMINGO can group them into clearly separable clusters. These results not only suggest the vigorous performance of Tensor-FLAMINGO in high-resolution imputation of single-cell chromatin contacts, but also the strong capability in capturing cell-type specific signatures of spatial genome architectures.

Single-cell specific spatial chromatin organization patterns revealed by Tensor-FLAMINGO

The predicted high-resolution spatial chromosomal conformations in single cells enable systematic characterizations of 3D genome folding patterns and the cell-to-cell variations (see Methods), which cannot be analyzed based on the highly sparse input data. Chromatin compartments, TADs and TAD boundaries across individual cells are identified based on the single-cell 3D structures. Their overall landscapes along the chromosome largely align with the patterns from bulk samples (Fig. 6a–d, Supplementary Fig. 8a–c), while each cell demonstrates cell-specific variations. The cross-cell structural variabilities are significantly associated with different chromatin compartments and gene expression specificities (Fig. 6e and f). Based on the joint profiled single-cell chromatin contacts and DNA methylation signals (Fig. 7a), the cell-specific hypomethylated genomic loci are found to be spatially proximal in the reconstructed single-cell 3D structures (Fig. 7b), suggesting spatially coordinated 1D epigenetic activities. Sub-clusters of cells are delineated based on each cell's 3D structures (Fig. 7c, Supplementary Fig. 9). Interestingly, the differentially hypomethylated genes are more densely organized on the cluster-specific chromosomal structures (Fig. 7d, e, Supplementary Fig. 10), suggesting the coordinated interplay between gene regulation and the spatial genome architectures. Detailed analyses of these findings can be found in the Supplementary Notes.

Dynamic single-cell long-range chromatin interactions identified by Tensor-FLAMINGO

As a major contribution to single-cell chromatin analysis by Tensor-FLAMINGO, the completed high-resolution chromatin contact maps facilitate the identification of significant long-range chromatin interactions in single cells, revealing not only the dynamics of chromatin interactions but also the functional impacts of distal genetic variants and cis-regulatory elements. Applied on the Dip-C dataset of 15 GM12878 cells, Tensor-FLAMINGO generated the 30kb-resolution 3D chromosomal structures, along with the completed single-cell chromatin contact maps. Significant long-range chromatin interactions at 30kb-resolution (i.e. scCI) are called for each cell using FitHi-C[57] (see Methods) based on the completed contact maps. ~1000 significant single-cell long-range interactions are identified per cell at 30kb-
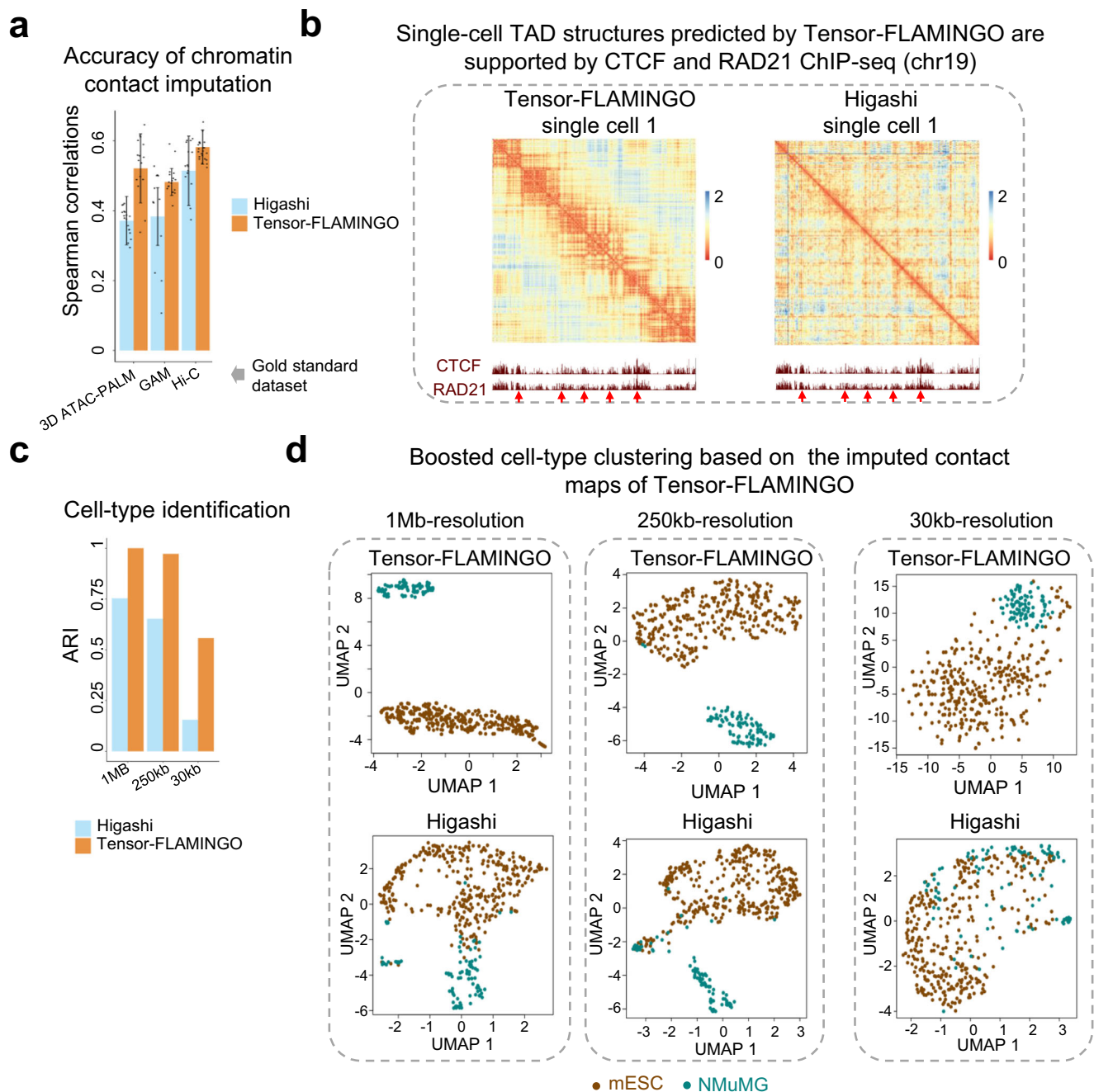
Fig. 5 | Tensor-FLAMINGO outperforms Higashi in imputing single-cell contact maps at high resolution. a Evaluation of the accuracy of imputed single-cell contact maps for mESC cells based on the snm3C input dataset at 30kb-resolution. Chromatin contact maps (3D ATAC-PALM, GAM and Hi-C) profiled by orthogonal experiments are used as gold standards. The error bar represents the standard deviations across 351 single cells. The centers of the bars represent the mean values. **b** Imputed single-cell contact maps of chr19 by Tensor-FLAMINGO and Higashi. Clear TAD patterns are observed in Tensor-FLAMINGO's imputed contact map, which overall align with the CTCF and Rad31 ChIP-seq peaks. **c, d** Performance comparison in identifying cell types based on the imputed single-cell contact maps across different resolutions. **c** The quantitative accuracy of cell-type identification is evaluated by the Adjusted Rand Index (ARI). **d** UMAP of distance matrices predicted by Tensor-FLAMINGO and Higashi at different resolutions of imputation. Dots represent single cells and are colored by the cell types (mESC and NMuMG). Source data are provided as a Source Data file.

resolution (*p-value* cutoff = 0.05). To note, although cell-type level interactions can be inferred (e.g. SnapHiC)[25], it is not easy to identify single-cell specific long-range interactions based on the original Dip-C contact maps due to their high rates of missing data (99.9%) at high-resolution. As a representative example, (Fig. 8a) shows the significant single-cell chromatin interactions in a 5.5 Mb genomic region on chr21 (chr21:15,500,000-21,000,000), which demonstrate substantial variabilities across 15 different cells at 30kb-resolution. Overall, the single-

cell long-range chromatin interactions show high consistency with bulk-level interactions profiled by the Capture-C experiment[10], with 73% single-cell interactions supported by the Capture-C data (Fig. 8b). In addition, the single-cell interactions that are also captured by Capture-C tend to be stronger interactions at the bulk-level (Fig. 8c), *p-value* = 3.47 × 10^{-3}, Mann-Whitney U test, suggesting they are more stable across single cells. On the other hand, the Capture-C data missed 27% single-cell interactions, emphasizing the limitation of bulk-level
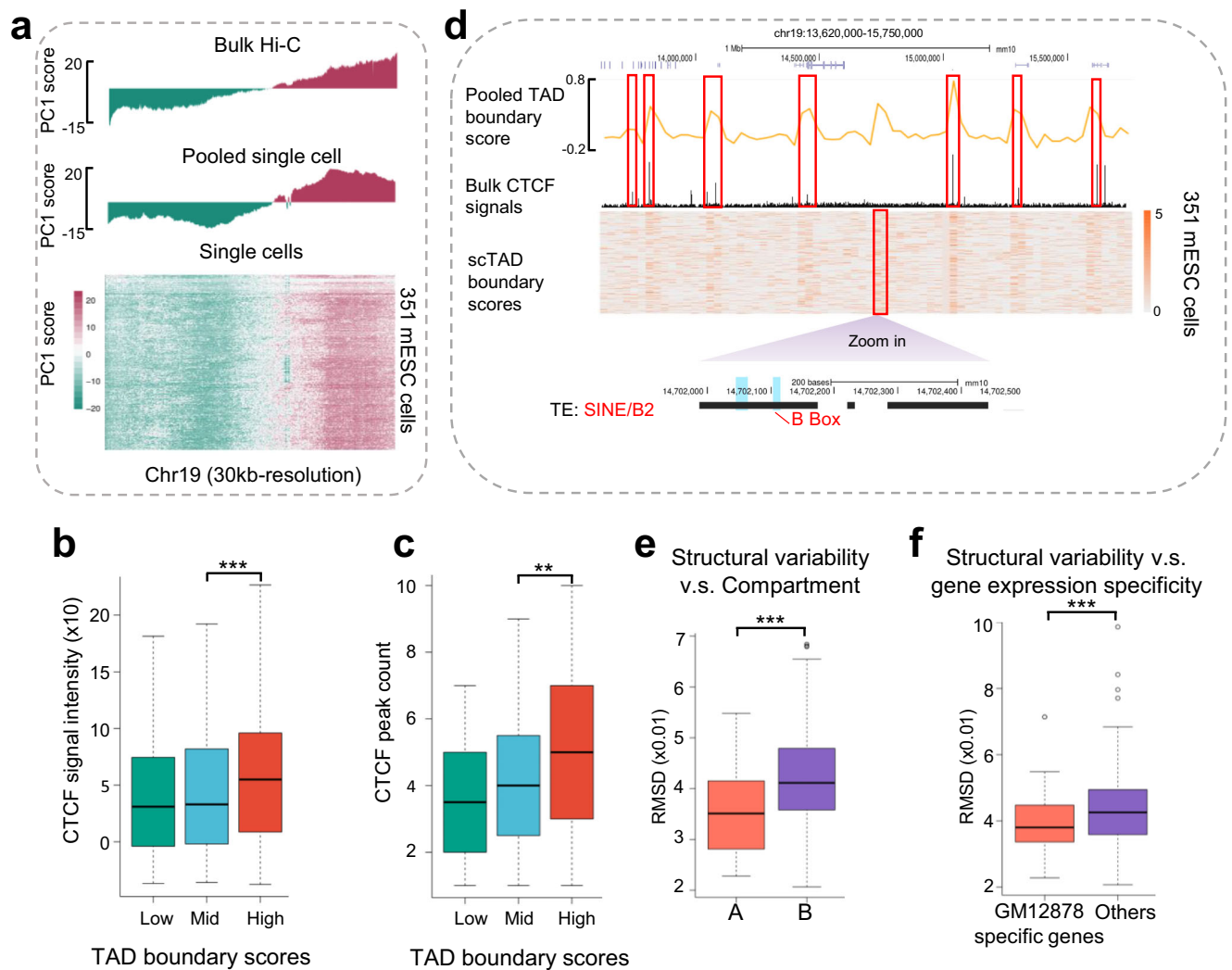
Fig. 6 | Identification of A/B compartments and TADs based on the completed single-cell contact maps. a Compartment identification in chromosome 19 for 351 mESC single cells at 30kb-resolution. PC1 scores based on the bulk Hi-C contact maps, the pooled single-cell contact maps, and all 351 single cells are shown separately. (**b, c**) Regions with higher TAD boundary scores tend to have **b** higher CTCF binding ChIP-seq signal strength (the number of called TAD boundaries $n = 789$, one-sided Mann-Whitney U test *p-value* = 2.6×10^{-4}) and **c** higher number of CTCF peaks (High: <20% quantile; Mid: 20%–80%; Low: >80%; the number of called TAD boundaries $n = 789$, one-sided Mann-Whitney U test *p-value* = 2.7×10^{-2}). For **b, c**, the center lines of boxplots show the median; the upper and lower box limits show the 25th and 75th percentiles, respectively. The whiskers extend up to 1.5 times the interquartile range away from the limits of the boxes. **d** Example of TAD boundary identification at chr19:13,620,000-15,750,000. Seven out of eight predicted TAD boundaries are supported by the bulk CTCF ChIP-seq peaks. The TAD boundary without CTCF peak contains a SINE B2 transposable element with a B-box for TFIIIC. **e, f** Tensor-FLA-MINGO's single-cell predictions reveal the patterns of structural variabilities across different genomic regions. The RMSD is calculated between the 3D chromatin structures of every single cell and the averaged structures of all cells. Lower RMSD indicates less structural variability and the specific genomic region is more structurally stable across single cells. **e** Compartment A regions show lower RMSD compared with compartment B regions (the number of called chromatin compartments $n = 121$, one-sided Mann-Whitney U test *p-value*:1.8×10^{-3}). **f** Genes with GM12878-specific expressions show lower RMSD compared with other genes (the total number of genes $n = 19,003$, one-sided Mann-Whitney U test *p-value*:0.037). For **e, f**, the center lines of boxplots show the median; the upper and lower box limits show the 25th and 75th percentiles, respectively. The whiskers extend up to 1.5 times the interquartile range away from the limits of the boxes. Source data are provided as a Source Data file.

experiments in identifying the variable chromatin interactions that occur only in small subsets of cells.

Furthermore, the landscapes of chromatin interactions across different cells signify the variable TAD organizations at the single-cell level that differ from the emergent TAD patterns observed at the bulk-level. As shown by the example in Fig. 8a, the bulk-level Hi-C chromatin contact map suggests three TADs in this specific genomic region. However, the single-cell level chromatin interaction landscapes drastically change across cells so that distinct TAD organizations are found in different subsets of cells (Fig. 8a). Detailed information of single-cell chromatin interactions and genome folding for this region can be found in Supplementary Fig. 11. In the first subset of cells, the chromatin interactions are densely organized into three domains, resulting

in three TADs that are consistent with the bulk-level observations. For these cells, their completed chromatin contact maps are also highly similar to the bulk-level contact map. The reconstructed 3D spatial conformation of these cells demonstrates closely folded loops along this genomic region (Fig. 8a), top-right. Interestingly, in the second subset of cells, the landscape of chromatin interactions is densely organized into two TAD domains, which is also shown in the completed single-cell chromatin contact maps and the reconstructed 3D conformation (Fig. 8a), middle-right. Compared to the three-TAD configuration observed at the bulk-level, two of the TADs on the 3' end merge together into one larger TAD. Moreover, in the third subset of cells, the landscape of chromatin interactions only shows one highly interacting domain, either on the 5' or the 3' end, depending on the
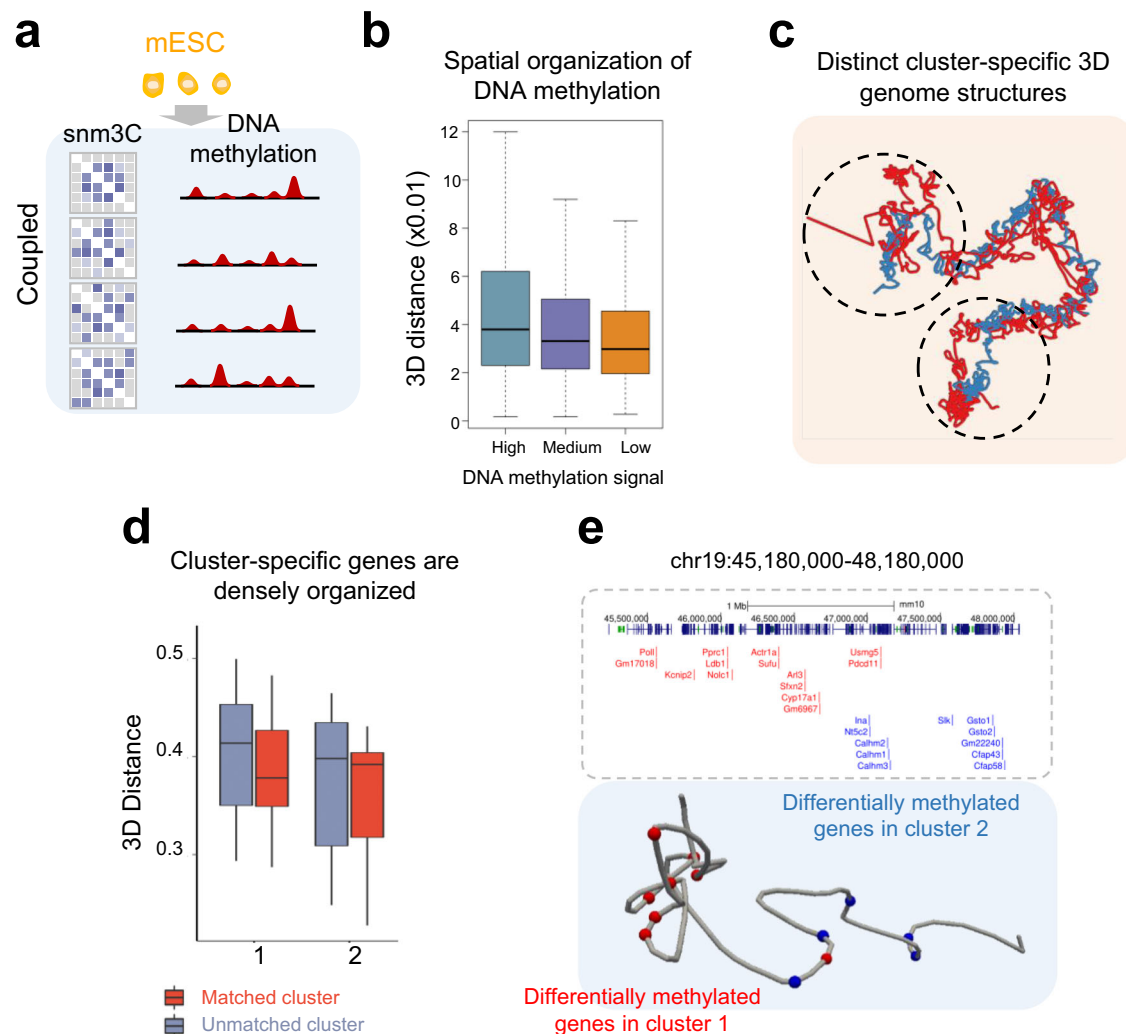
Fig. 7 | Dynamic 3D chromatin structures across single cells demonstrate distinct spatial organizations of gene expression. a The schematic figure of the jointly profiled single-cell chromatin contacts and the coupled DNA methylation in the snm3C dataset. **b** Hypomethylated gene promoters are spatially closer in 3D space across single cells. For every single cell, genes are divided into three groups based on their DNA methylation scores around promoters (Low: <30% quantile, $n = 105$; Medium: 30%–70% quantile, $n = 140$; High: >70% quantile, $n = 105$). The pairwise 3D distances between two genes within the same methylation group are calculated across all single cells. The center lines of boxplots show the median; the upper and lower box limits show the 25th and 75th percentiles, respectively. The whiskers extend up to 1.5 times the interquar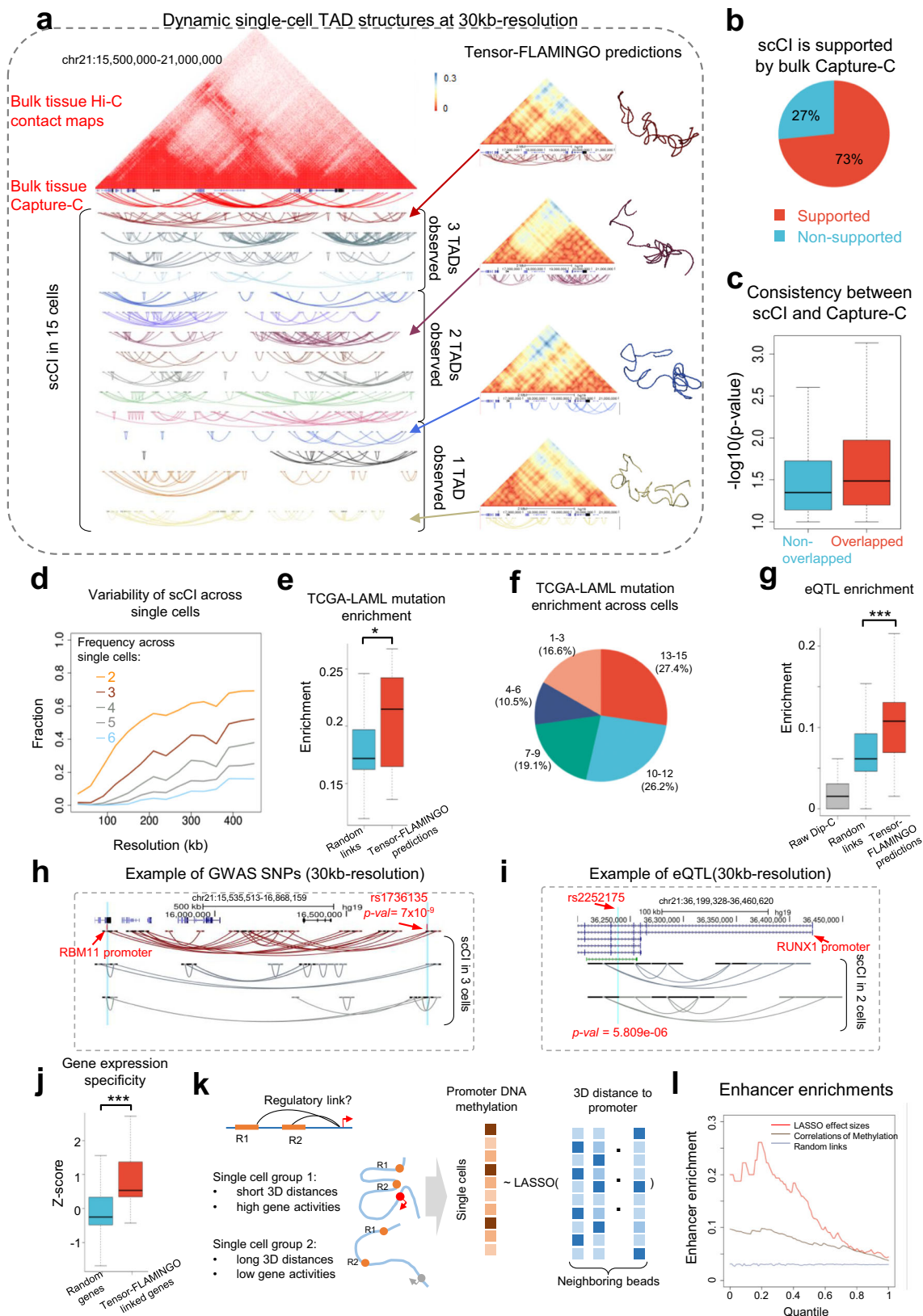tile range away from the limits of the boxes. **c** 351 single cells are divided into two clusters based on the similarity of their predicted 3D chromatin structures. Tensor-FLAMINGO delineates the cells into two distinct clusters of chromatin structures (red vs. blue). **d** Distribution of the spatial distances between the cluster-specific differentially methylated genes across cell clusters. Cluster-specific hypomethylated genes show shorter 3D distances in cells belonging to the matched cluster ($n = 77$) versus the unmatched cluster ($n = 49$). The center lines of boxplots show the median; the upper and lower box limits show the 25th and 75th percentiles, respectively. The whiskers extend up to 1.5 times the interquartile range away from the limits of the boxes. **e** Example of the cluster-specific 3D chromatin structure spatially organizing cluster-specific differentially methylated genes into the dense 3D neighborhood. Source data are provided as a Source Data file.

specific cells (Fig. 8a), bottom-right. For these cells, the chromatin interactions in some of the bulk-level domains disappear, resulting in the mutually-exclusive vanishment of TADs at the single-cell level, which is also shown by the completed single-cell chromatin contact maps and the reconstructed spatial conformation. These remarkable observations underscore the sophisticated dynamics of chromatin interactions in single cells that is not distinguishable based on the collective patterns profiled at the bulk-level experiments.

Spectrum of variable chromatin interactions in single cells across different resolutions

The variability of long-range chromatin interactions in single cells is systematically evaluated across a diverse range of resolutions. Based on the completed high-resolution single-cell chromatin contact maps of the GM12878 Dip-C dataset[45], contact maps at lower resolutions are also generated, which are used to call the significant long-range

interactions[57] at corresponding resolutions. As expected, at lower resolutions, the long-range chromatin interactions are more likely to be shared by multiple cells. For example, at 250kb-resolution, over 50% interactions are shared by at least two cells and around 40% are shared by at least three cells (Fig. 8d), suggesting the aggregated bulk-level chromatin interactions can be reasonable approximations to single-cell signals at low resolutions. In contrast, at higher resolutions, the frequencies of long-range chromatin interactions shared by more cells dramatically decrease. For instance, at 50kb-resolution, only around 10% interactions are shared by more than two cells and less than 5% are shared by more than three cells, with the vast majority of chromatin interactions being specific to one single cell (Fig. 8d). Collectively, the spectrum of variabilities indicates the relatively stable backbone of genome folding at low resolutions, coupled with dynamic chromatin architecture and epigenetic plasticity across different cells at high resolutions.

a Dynamic single-cell TAD structures at 30kb-resolution

b scCI is supported by bulk Capture-C

c Consistency between scCI and Capture-C

d Variability of scCI across single cells

e TCGA-LAML mutation enrichment

f TCGA-LAML mutation enrichment across cells

g eQTL enrichment

h Example of GWAS SNPs (30kb-resolution)

i Example of eQTL (30kb-resolution)

j Gene expression specificity

k Regulatory link?

l Enhancer enrichments

Functional interpretation of genetic variants based on single-cell chromatin interactions

A unique advantage of high-resolution reconstruction of 3D chromatin, compared to low-resolution predictions, is the capability of interpreting the functional effects of specific genetic variants. The majority of significant genetic variants discovered in GWAS and eQTL studies are non-coding and located distal to genes[58]. A prevailing mechanistic assumption for these genetic variants is that they may disrupt long-range chromatin interactions involved in gene regulation[59]. Consistent with this hypothesis, the acute myeloid leukemia (LAML) associated somatic mutations profiled from the TCGA dataset[60] are found to be significantly enriched in genomic loci that participate in long-range

Fig. 8 | Prediction of the long-range high-resolution single-cell chromatin interactions based on the imputed contact maps by Tensor-FLAMINGO.
a Predicted single-cell long-range chromatin interactions across 15 GM12878 cells at 30kb-resolution, along with the bulk Hi-C contact maps and Capture-C interactions. Across single cells, different 3D chromatin structures and the landscape of long-range chromatin interactions are observed, revealing the structural variability across cells. Only statistically significant chromatin interactions in each cell are shown (FitHi-C p-value < 5×10^{-5}). **b** The fraction of predicted single-cell chromatin interactions supported by Capture-C interactions. **c** Bulk Capture-C interactions overlapping with predicted single-cell chromatin interactions (scCI) have lower p-values (p-value = 3.47×10^{-3}, one-sided Mann-Whitney U test, non-overlapped n = 11,967, overlapped n = 2663), suggesting strong consistency between Capture-C and predicted scCI. The center lines of boxplots show the median; the upper and lower box limits show the 25th and 75th percentiles, respectively. The whiskers extend up to 1.5 times the interquartile range away from the limits of the boxes. **d** Variability of single-cell chromatin interactions across 15 cells at different resolutions. At a specific resolution, fractions of chromatin interactions that are observed in different numbers of cells are shown. **e** Single-cell chromatin interactions are enriched with TCGA-LAML somatic mutations compared with distance-controlled random interactions (one-sided Mann-Whitney U test p-value = 0.007, the count of somatic mutations n = 419). The center lines of boxplots show the median; the upper and lower box limits show the 25th and 75th percentiles, respectively. The whiskers extend up to 1.5 times the interquartile range away from the limits of the boxes. **f** Frequency of the TCGA-LAML somatic mutations being captured by the predicted single-cell chromatin interactions that are shared by different numbers of cells, including rare interactions shared by only 1-3 cells and conserved interactions shared by 13-15 cells. **g** Single-cell chromatin interactions are enriched with eQTLs compared with distance-controlled random interactions

and the contacts in the raw Dip-C data (one-sided Mann-Whitney U test p-value = 0.0022, the count of eQTLs n = 65). An eQTL is considered to be captured by the raw Dip-C data if the corresponding read count is greater than 1 (i.e. not missing data) in each single cell. The center lines of boxplots show the median; the upper and lower box limits show the 25th and 75th percentiles, respectively. The whiskers extend up to 1.5 times the interquartile range away from the limits of the boxes. **h** Example of GWAS SNP overlapping with predicted single-cell chromatin interactions. rs1736135 (GWAS test p-value = 7×10^{-9}) is linked to the gene RBM11 through predicted long-range chromatin interactions in three single cells. **i** Example of eQTL (rs2252175-RUNX2 pair, eQTL test p-value = 5.8×10^{-6}) overlapping with predicted single-cell chromatin interactions in two single cells. **j** Genes linked by the predicted single-cell chromatin interactions have higher expression specificity in GM12878 (one-sided Mann-Whitney U test p-value = 0.0032, the total number of genes n = 18,324). The center lines of boxplots show the median; the upper and lower box limits show the 25th and 75th percentiles, respectively. The whiskers extend up to 1.5 times the interquartile range away from the limits of the boxes. **k** Schematic figure showing the prediction of functional regulatory links based on Tensor-FLAMINGO's predicted spatial distances and the coupled DNA methylation scores of genes in mESC from the snm3C data. LASSO regression is used to prioritize the DNA fragments whose 3D distances to the gene promoter can best predict the promoter DNA methylation scores of the target gene. The longest distances between DNA fragments and target gene promoters are limited to 3 Mb. **l** Enrichment of enhancers in top-ranking DNA fragments prioritized by the LASSO effect sizes. As comparison, the correlations of DNA methylation between the DNA fragments and target gene promoters across single cells are used as the predictive score to prioritize DNA fragments for enhancer enrichment analysis. The result of distance-controlled random chromatin interactions is also shown. Source data are provided as a Source Data file.

single-cell chromatin interactions, which are predicted by Tensor-FLAMINGO based on the GM12878 Dip-C dataset (Fig. 8e). The enrichment is statistically significant, compared to the stringent background of genomic distance-controlled random interactions (p-value = 7×10^{-3}, Mann-Whitney test), suggesting the potential mechanisms of LAML-association for these somatic mutations are likely mediated by long-range chromatin interactions. Strikingly, among these LAML-associated mutations involved in single-cell chromatin interactions, 46% of which overlap with interactions that are only observed in less than ten cells (totally 15 cells in the GM12878 Dip-C dataset) and 16.6% of which overlap with interactions that are only observed in less than three cells (Fig. 8f). In addition, the SNP-gene pairs of eQTLs profiled from blood tissues are significantly enriched in the predicted single-cell specific long-range chromatin interactions, compared to the background of genomic distance-controlled random interactions p-value = 0.0013, Mann-Whitney test, (Fig. 8g). To note, here, the efficient interpretability of eQTLs is achieved by the completed single-cell 3D chromatin structures from Tensor-FLAMINGO. Using the original Dip-C dataset, the single-cell chromatin interactions can only interpret a much smaller fraction of eQTLs overlapping fraction = 0.017, (Fig. 8g), because of the very high rates of missing contacts in the original dataset. These results imply the importance of single-cell specific chromatin interactions in mediating the functional effects of genetic variants, beyond common interactions.

As a representative example of GWAS SNPs, the Crohn's disease associated SNP rs1736135 is predicted to have single-cell specific long-range chromatin interactions to the promoter of RBM11 (Fig. 8h), which has been previously documented to be functionally important and related with the survival rates of patients[61]. Based on Tensor-FLAMINGO's predictions, significant long-range chromatin interactions are discovered for three single cells, highlighting the necessity of delineating the heterogeneity of chromatin across cell populations. As another example for the SNP-gene pairs of eQTLs, the SNP rs2252175 is statistically linked to the gene RUNX1 based on genetic association studies[58]. Interestingly, in two single cells, significant long-range chromatin interactions are identified by Tensor-FLAMINGO, which suggest that the SNP rs2252175 is spatially proximal to the promoter of

RUNX1 gene in a single-cell specific manner (Fig. 8i). These observations further highlight the potential limitation of bulk-level interactions in interpreting disease-associated genetic variants or eQTLs, since the dynamic or transient interactions in subsets of cells may not be captured. Notably, Tensor-FLAMINGO's predictions substantially improve the single-cell level interpretations of genetic variants and reveal the underlying mechanistic basis of genetic associations.

Predicting functional gene regulatory links in single cells
The single-cell specific 3D spatial distances between genomic loci predicted by Tensor-FLAMINGO lead to avenues of decoding the long-range transcriptional regulatory programs in individual cells, which may not be observable based on bulk-level Hi-C due to the transient and dynamic nature of regulatory activities. As shown in Fig. 8j, genes linked by predicted single-cell chromatin interactions have higher expression specificity in GM12878 (p-value = 0.0032, Mann-Whitney test), suggesting precise transcriptional regulation is spatially facilitated in 3D space. Based on the snm3C dataset, which jointly profiled the coupled single-cell information of chromatin contacts and DNA methylation signals from the same set of cells, a LASSO model is built for each gene across cells to identify the single-cell specific regulatory elements involved in its transcriptional regulation (Fig. 8k), see Methods. The predicted single-cell 3D spatial distances between a genomic loci and the target gene's promoter across individual cells are incorporated as features. The single-cell DNA methylation signals of the target gene's promoter, from the same set of single cells, are used to quantify its expression activities, based on the anti-correlation between the promoter's DNA methylation signals and the gene's expression level[62]. Based on the LASSO models, specific genomic loci, whose single-cell spatial distances to the target gene's promoters are strongly associated with the promoter's DNA methylations across cells, are prioritized as the linked candidate regulatory elements for each gene. The inferred effect sizes from LASSO for each candidate regulatory element represent the quantitative regulatory strengths and are used to rank the linked elements. Remarkably, the top-ranking regulatory elements linked to each gene are highly enriched with previously annotated enhancers in mESC[63] (Fig. 8l), suggesting the

accuracy and efficiency of Tensor-FLAMINGO's results in delineating transcriptional regulatory networks. Additionally, the enrichment is consistently higher than the results obtained using the activity correlations of DNA methylations between candidate loci and target gene's promoters across single cells (Fig. 8l). The strikingly elevated enrichment indicates the importance of the dynamic high-resolution chromatin contacts in modulating gene expressions across single cells, beyond the information from 1D epigenetic landscapes.

Single-cell specific multi-way interactions revealed by Tensor-FLAMINGO

Beyond the pairwise two-way chromatin contacts annotated in chromatin contact maps, the spatial architecture of chromosomes contains numerous multi-way interactions. Multi-way chromatin interactions establish spatial hubs of interacting loci and facilitate coordinated molecular processes. The predicted high-resolution 3D chromosomal structures by Tensor-FLAMINGO enable the probing of spatial distances among multiple genomic loci so as to identify multi-way interactions directly in the 3D space. Taking three-way interactions as examples, the averaged 3D chromosomal structures based on Tensor-FLAMINGO's predictions across 351 mESC cells are first used to call the stable three-way interactions, based on the spatial compactness and

the associated statistical significance see Methods, (Fig. 9a). Totally 973 statistically significant three-way interactions (p-value < 0.05) were discovered, compared against 1D genomic distance controlled backgrounds. Using the single-cell SPRITE (scSPRITE) data as the orthogonal gold-standard for multi-way interactions[64], the predicted three-way interactions by Tensor-FLAMINGO demonstrate significantly higher enrichment of overlapping with scSPRITE multi-way clusters (Fig. 9b), strongly supporting the accuracy. To note, this is a significant improvement since the original single-cell chromatin contact data is too sparse to identify three-way interactions. Figure 9c shows an example of predicted three-way interactions, where three interacting anchors are brought to the same 3D neighborhood based on the predicted high-resolution chromatin folding. Interestingly, the predicted three-way interactions are consistently observed across single cells in the scSPRITE dataset and appear as a strong scSPRITE three-way cluster.

To further delineate the structural heterogeneity of multi-way interactions across single cells, the 15 GM12878 cells are divided into three groups based on their predicted chromosomal structures. For each cell group, cell-group specific three-way interactions are then identified (see Methods). On average, each cell group contains around 1000 significant group-specific three-way interactions. As one
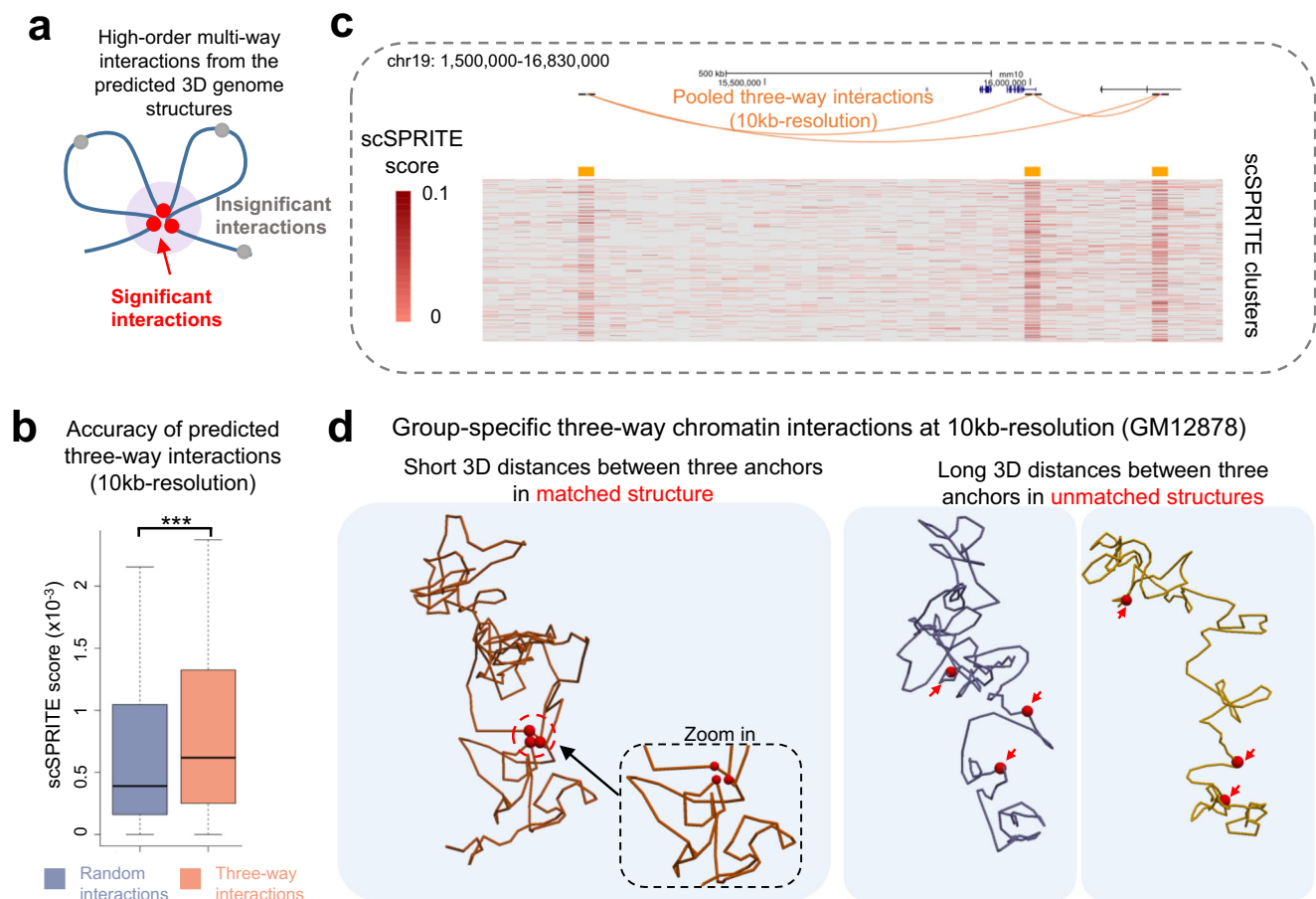

Fig. 9 | Prediction of three-way chromatin interactions based on Tensor-FLAMINGO's predicted 3D chromatin structures. a Three-way chromatin interactions are predicted from the 3D chromatin structures by evaluating the set of pairwise 3D distances. **b** Predicted three-way chromatin interactions at 10kb-resolution are supported by single-cell SPRITE (scSPRITE) data (the number of predicted three-way interactions $n = 3408$, p-value $= 1.4 \times 10^{-5}$, one-sided Mann-Whitney U test). Higher scSPRITE scores suggest that the predicted three-way interactions are observed in more single cells in the scSPRITE dataset. The center lines of boxplots show the median; the upper and lower box limits show the 25th and 75th percentiles, respectively. The whiskers extend up to 1.5 times the interquartile range away from the limits of the boxes. **c** Example of the predicted three-way interactions supported by the scSPRITE cluster at the same genomic location. The anchors of the predicted three-way interaction align with the hotspots from the scSPRITE data. **d** Example of the group-specific single-cell 3D chromatin structures and group-specific single-cell three-way interactions. Source data are provided as a Source Data file.

interesting example shown in Fig. 9d, the significant three-way interaction brings the genomic loci into the proximal 3D neighborhood, which is only specific in one cell group. As comparison, in the other two cell groups, the three genomic loci are far away from each other in 3D space, due to the altered chromatin folding (Fig. 9d). These observations not only suggest the complicated variability of higher-order chromatin organizations across single cells, but also indicate that the predicted high-resolution 3D structures by Tensor-FLAMINGO open up additional ways of understanding the spatial cooperation among multiple genomic regions, beyond the conventional analyses of two-way contacts.

Discussion

In this work, we developed Tensor-FLAMINGO to reconstruct single-cell specific 3D chromosomal structures at high-resolution from the sparse single-cell chromatin contact maps. Uniquely equipped with the low-rank tensor completion technique, Tensor-FLAMINGO mitigates the very high rates of missing contacts in single-cell chromatin contact maps. The application of Tensor-FLAMINGO across four different single-cell chromatin conformation capture datasets provides a rich resource of single-cell 3D chromosome structures at 10kb- and 30kb-resolution. Based on the extensive performance evaluations, Tensor-FLAMINGO achieves superior accuracy over the state-of-art methods in: (1) reconstructing single-cell 3D chromosomal structures, (2) imputing single-cell chromatin contact maps, (3) capturing the cell-type specific structural variations, and (4) delineating different cell types or clusters. The high agreement between the experimental super-resolution imaging data and the Tensor-FLAMINGO's predictions further supports the accuracy of Tensor-FLAMINGO in predicting the dynamic and heterogeneous spatial chromatin configurations across single cells. Biologically, Tensor-FLAMINGO enables the systematic evaluation of the cell-to-cell structural variations of chromatin compartments and TADs, along with the relationship to transcription. Coupled with the single-cell DNA methylation landscape, Tensor-FLAMINGO unveils single-cell specific interplays between the dynamic gene regulation and spatial chromosomal structures. The in-depth characterizations of high-resolution 3D structures facilitate the identification of single-cell specific long-range chromatin interactions involved in gene regulation and provide mechanistic interpretations of GWAS SNPs and long-range eQTLs. Beyond conventional analysis of 2D chromatin contact maps, the reconstructed chromatin conformation in the 3D space further reveals complex multi-way chromatin interactions, leading to insights into the higher-order spatial organizations.

As a data-driven model, Tensor-FLAMINGO solely relies on the input single-cell chromatin contact maps and the low-rank property, which is intrinsic due to the nature of 3D spatial coordinates and the dependency among individual cells. It does not rely on any specific biophysical assumptions, which may not be true for different genomic regions or different datasets and may introduce unexpected bias into the model. This is crucial for reconstructing high-resolution single-cell 3D chromosomal structures, considering the critical challenge imposed by the highly sparse data of single-cell chromatin. On the other hand, the simulation-based constrained polymer models rely on both pre-assumed biophysical properties of the DNA sequences and the single-cell chromatin contact maps to predict 3D structures. This strategy can reconstruct low-resolution chromosomal structures (i.e. 1MB-resolution) when most of the DNA fragments have regular constraints in the relatively dense low-resolution single-cell chromatin contact maps. However, the predictive accuracy at high-resolution (i.e. 10 kb) usually decreases drastically for two reasons. Because the high-resolution chromatin contact maps are incredibly sparse, the simulation-based prediction is essentially dominated by the pre-assumed biophysical properties, which are invariant across single cells and cell types and may contradict the observed chromatin contact maps. In this case, the model cannot find an optimal structure to satisfy both constraints, thus deviating from the observed chromatin contact signals and diluting the cell-to-cell structural variabilities. Furthermore, because the underlying chromatin loop formation mechanisms are still being actively investigated, the pre-assumed biophysical properties may not hold for the specific dataset. It is also possible that different mechanisms establish chromatin loops at different genomic regions or in different species. Thus, since no prior assumptions of chromatin folding are used, Tensor-FLAMINGO achieves robust performance in reconstructing high-resolution 3D structures across different single cells and distinct cell types. Considering the complementary advantages of the data-driven Tensor-FLAMINGO model and the biophysical-assumption based methods, combining these two types of approaches for 3D genome modeling across diverse panels of single-cell chromatin contact datasets is a promising strategy for deriving mechanistic insights into the dynamics of chromatin folding.

Three directions are envisioned for future developments of Tensor-FLAMINGO. First, due to the large-scale tensor of high-resolution chromatin contacts across cells, the computational scalability needs to be further improved. Currently, Tensor-FLAMINGO requires >36 hours to complete the 10kb-resolution single-cell reconstructions for 15 cells, based on 20 GPUs and 100 G memory. Given the fast development of single-cell Hi-C techniques, the number of cells studied by one experiment is expected to increase in the near future, requiring a substantial boost in computational scalability and memory efficiency of 3D genome reconstruction. Advanced algorithm design and parallelized optimization are thus desired for future development. Second, the reconstructed single-cell specific 3D structures by Tensor-FLAMINGO simultaneously generate the completed single-cell specific chromatin contact maps. Therefore, Tensor-FLAMINGO can also be considered as a single-cell Hi-C imputation algorithm, based on the global spatial folding information across individual cells, providing more candidate single-cell chromatin interactions for statistical testing. On the other hand, to gain insights into gene regulation and epigenetics, it is important to identify statistically significant long-range chromatin interactions. The state-of-the-art method SnapHiC[25] can take single-cell Hi-C data and predict significant chromatin loops. Therefore, based on the complementary advantages of Tensor-FLAMINGO and SnapHiC, it is expected to have promising future developments to integrate the two types of models and expand systematical annotations of long-range cis-regulatory interactions at both cell-type and single-cell levels (Supplementary Fig. 12). Third, to enable precise interpretation of functional non-coding regulatory elements and genetic variants, the resolution of 3D chromosomal configuration needs to be improved. Currently, Tensor-FLAMINGO is able to achieve robust reconstructions at 10kb-resolution for single cells. Although 10kb-resolution is the highest for single-cell level reconstructions to date, future developments are demanded to further increase the resolution to ~1 kb, so that specific regulatory genetic variants can be pinpointed to study detailed gene regulation and disease mechanisms. Higher resolutions will also facilitate the analysis of small-scale structural variations across single cells.

Methods

Single-cell chromatin contact maps and data preprocessing

Different types of single-cell chromatin contact maps are collected from four studies, including the Dip-C experiment of GM12878 cells[45], the snHi-C experiment of K562 cells[44], the snm3C experiment of mESC cells[47], and the scHi-C experiment of mESC cells[40] (see details in the Data Availability Statement). Single-cell chromatin contact frequencies are normalized and scaled using band-wise log-linear regression models, where the bands of contact maps represent groups of chromatin interactions sharing similar 1D genomic distances as suggested by previous work[65] and our own comparisons of single-cell and bulk chromatin contacts at different scales (Supplementary Fig. 13). For

each pairwise chromatin contact, the band-specific log-linear regression model systematically controls the confounding effects of genomic distances, missing rates, single-cell specific sequencing depths, and loci-specific interaction frequency backgrounds Supplementary Figs. 14–15. Regression parameters are estimated for each band separately. The scaled and normalized single-cell chromatin contact frequency maps are then organized into sparse tensors as inputs for Tensor-FLAMINGO. To note, as suggested by the previous studies[66,67], we also evaluated the potential bias induced by GC contents and found that our normalized signals do not have significant correlations with GC contents (Spearman correlation=0.0146, p-value = 0.35, randomly sampled signals n = 4107).

Bulk-tissue chromatin contact maps generated from the same cell types are collected from the GEO[68] and 4D Nucleome[69] databases (see more details in the Data Availability Statement), including bulk Hi-C experiments in GM12878 and K562 (GSE63525), GAM experiment in mESC (GSE64881), 3D ATAC-PALM experiment in mESC (GSE126112) and bulk Hi-C experiment in mESC (4DNFI5IAH9H1)[6,53–55]. The GAM dataset of mESC only provides chromatin contact maps at 30kb-resolution. All other bulk-tissue chromatin contact maps are used in analyses at both 10kb- and 30kb-resolution. Chromatin contact maps are normalized by the Knight-Ruiz normalization method[70]. The single-cell SPRITE data (scSPRITE) of multi-way chromatin interactions[64] is collected from the GEO database (GSE154353) and preprocessed according to the instructions from the original dataset.

Model framework of Tensor-FLAMINGO

Single-cell chromatin contact maps are summarized into a sparse 3-mode tensor (T^{obs}, Fig. 1a), with the frontal slices representing the chromatin contact maps of individual cells and each tube along the third mode of the 3-mode tensor representing a specific pairwise DNA fragment interaction across different cells. The dimension of T^{obs} is $M \times M \times N$, where M corresponds to the number of genomic loci under the specified resolution and N corresponds to the number of single cells.

Tensor-FLAMINGO (https://github.com/wangjr03/Tensor-FLAMINGO) employs the low tubal-rank tensor completion technique to reconstruct high-resolution single-cell 3D chromatin structures by taking advantage of the hidden low-complexity structures in single-cell chromatin contact maps from two fundamental perspectives: the first is the low tubal-rank structure of the 3-mode tensor corresponding to multiple single cells, and the second the low-rank structure of the chromatin contact maps corresponding to each single cell. First, Tensor-FLAMINGO uses information across cells and jointly models the chromatin contact signals from different single cells simultaneously. Biologically, it leverages the relationships among individual cells under consideration by viewing that chromatin maps from cell to cell are highly correlated and such correlation is captured by a convolution-type operation. The t-SVD is based on such an operation along the third mode using the discrete Fourier transform, so its application to the 3-mode tensor results in a much efficient representation, leading to efficiency in recovering the missing data in the 3-mode tensor. Moreover, the low-complexity structure induced by this correlation can be characterized by the low tubal-rank property of the 3-mode tensor and can be captured by minimizing the tensor nuclear norm which is closely related to minimizing the tubal rank of the 3-mode tensor. Second, Tensor-FLAMINGO explores the low-rank property of chromatin contacts within each cell. Biologically, it makes use of the fact that the numerous pairwise distances are induced by the underlying low-rank 3D spatial coordinates of genomic loci[35]. Computationally, an Alternating Direction Method of Multipliers (ADMM) based tensor tubal-rank-minimization algorithm[51] is first implemented to impute the missing values of single-cell chromatin contact maps and to recover a dense tensor, followed by our low-rank matrix completion based structure reconstruction algorithm, FLAMINGO[35], to infer the

high-resolution spatial chromosomal conformations for each single cell.

Jointly impute single-cell chromatin contact maps based on low tubal-rank tensor completion

Based on the sparse input tensor of observed single-cell chromatin contact maps (T^{obs}), Tensor-FLAMINGO first employs a low tubal-rank tensor completion strategy[51] to impute missing values and recover a dense tensor (Fig. 1a) from the set of limited observed entries (Ω). The recovered dense tensor (X) captures the low-complexity structure embedded in the input high-dimensional tensor, while maintaining the minimum error on the set of observed pairwise entries of single-cell chromatin contact frequencies. The overall low-complexity structures consisting of the low tubal-rank structure at the 3-mode-tensor level (the chromatin contact maps of multiple single cells) and the low-rank structure at the matrix level (the chromatin contact map of each single cell) enables efficient imputation of missing data of un-observed single-cell chromatin contacts by leveraging the biological relationship across different cells and by leveraging the inter-dependency among pairwise chromatin contacts within each cell. Therefore, the objective function for the low tubal-rank tensor reconstruction is:

$$\min ||X||_{TNN}, \ s.t. \Omega(T^{obs}) = \Omega(X), \tag{1}$$

where T^{obs} represents the sparse input tensor of single-cell chromatin contact maps, X represents the dense tensor to be recovered, Ω represents the set of observed chromatin contact entries in T^{obs}, and TNN represents the Tensor Nuclear Norm, which is closely related to the tubal rank of a 3-mode tensor.

Tensor-FLAMINGO employs an efficient Alternating Direction Method of Multipliers (ADMM) strategy to solve the optimization problem. For ADMM, the objective function can be re-written in the equivalent form as:

$$\min_{X, Z} ||X||_{TNN}, \ s.t. \ y = A \times Vec(Z), \ X = Z \tag{2}$$

where $y = A \times Vec(T^{obs})$. Here, $A \in R^{(ijk)(ijk)}$ is a binary matrix, which has the value of 1 on $\Omega_{T^{obs}}$ and 0 otherwise. Vec represents the vectorization process of a tensor. And the Lagrangian can be written as:

$$L(X, Z, \lambda_1, \lambda_2) = ||X||_* + \langle \lambda_1, y - A \times Vec(Z) \rangle + \langle \lambda_2, X - Z \rangle$$
$$+ \frac{\mu}{2} |y - A \times Vec(Z)|^2 + \frac{\mu}{2} |X - Z|^2 \tag{3}$$

Accordingly, X, Z and the Lagrangian multipliers λ_1, λ_2 can be updated through iterations as:

$$X^{l+1} = \text{argmin}_X ||X||_* + \frac{\mu}{2} ||X - Z^l + \lambda_2^l/\mu||_F$$
$$Z^{l+1} = (A^T A + I)^{-1}(-A^T \lambda_1^l/\mu + Vec(\lambda_2^l)/\mu + A^T y + Vec(X^{l+1}))$$
$$\lambda_1^{l+1} = \lambda_1^l + \mu(A \times Vec(Z^{l+1}) - y) \tag{4}$$
$$\lambda_2^{l+1} = \lambda_2^l + \mu(X^{l+1} - Z^{l+1})$$

In each iteration, the update of X involves an optimization, which is solved by the t-SVD method (see below)[51].

Through iterations, the updating scheme of X is equivalent to solving the optimization problem

$$\min ||X^{l+1}||_*, \ s.t. \ X = T_l^{obs}$$

where $T_l^{obs} = Z^l - \lambda_2^l/\mu$ represents the optimal approximation of the observed data in the lth ADMM iteration. This optimization problem can be solved using the t-SVD algorithm[51]. The t-SVD decomposes a 3-mode tensor (A) into the multiplication of three tensors: $A = U * \Sigma * V^T$, where * represents the circular convolution product (t-

product) of tensors, U and V are orthogonal in the t-product sense, and Σ is an f-diagonal tensor with all frontal slices being diagonal matrices. Based on the t-SVD, the tubal rank of tensor (A) is defined as the number of nonzero singular tubes of Σ[51]. Now we apply the t-SVD to decompose the tensor of observed data (T_l^{obs}) into the multiplication of three tensors: $T_l^{obs} = U * \Sigma * V^T$. According to the tensor-completion theory[51], the tensor completion problem can be solved by calculating the matrix SVD across all frontal slices of the tensor in the Fourier domain. Thus, the observed 3-mode tensor T_l^{obs} is transformed into the Fourier domain using a tube-wise Fourier transformation:

$$\hat{T}_{l;i,j,k}^{obs} = \sum_{n=1}^{N} T_{l;i,j,n}^{obs} e^{-\frac{2\pi i k n}{N}} \tag{5}$$

Intuitively, the contact between DNA fragment i and DNA fragment j for single cell k in the Fourier domain ($\hat{T}_{l;i,j,k}^{obs}$) is calculated from the same contact across all single cells ($T_{l;i,j,n}^{obs}$ for all n). Therefore, if any single cell chromatin contact map contains observed values for the contact (i,j), all values in the tube $\hat{T}_{i,j,:}^{obs}$ will be completed in the Fourier domain by aggregating the observations of all cells. Given the tensor $\hat{T}_l^{obs}$ in the Fourier domain, the soft-thresholding SVD is applied on every frontal slice of $\hat{T}_l^{obs}$ ($\hat{T}_{:,:,k}^{obs}$) to calculate the dense low tubal-rank tensor X^l: $X^l = U_k^{obs} * (\Sigma_k^{obs})_{1..m} * (V_k^{obs})^T$, where $(\Sigma_k^{obs})_{1..m}$ represents the m largest singular values in Σ_k^{obs}. The SVD procedure captures the low-rank structures of the frontal slices and uses information across all contacts within each cell. The recovered tensor is then transformed back into the original domain using the inverse Fourier transformation, and the resulting tensor can maximally approximate the input one.

The updated X facilitates the use of information across all different single cells and across different chromatin contacts within each cell. The updated Z guarantees that the predicted values are close to the observed signals on the observed set (Ω). To note, all model parameters are jointly inferred through iterations adaptively based on the input data and there are no free parameters requiring prior settings. Upon convergence, the first step of Tensor-FLAMINGO imputes much denser contact maps for every cell, which maximally align with the observed single-cell chromatin contact signals and will be used to construct the final single-cell specific 3D structures in subsequent steps (Supplementary. 16).

Reconstruct single-cell 3D chromatin structures based on low-rank matrix completion

Based on the imputed low-rank single-cell chromatin contact tensor, Tensor-FLAMINGO further implements our low-rank matrix completion based algorithm, FLAMINGO[35], to reconstruct the 3D chromosomal structures for each single cell (Fig. 1a). The imputed single-cell chromatin contact maps of interaction frequencies (IF) are converted to pairwise spatial 3D distance (PD) matrices using the exponential conversion function: $PD_{ij} = IF_{ij}^{-\alpha}$ ($\alpha = 0.25$), which has been estimated and justified by previous studies[12]. By explicitly exploring the low-rank dependence among the large amounts of pairwise spatial distances (matrix rank is at most five), FLAMINGO achieves (1) high-resolution reconstruction of 3D conformations, (2) strong robustness against high rates of missing values, and (3) orders of magnitudes boost in scalability[35]. The superior accuracy of predicted 3D structures by FLAMINGO at 5kb- and 1kb-resolution, along with their functional impacts on gene regulation and distal genetic variants, has been systematically benchmarked by both simulations and different orthogonal chromatin contact datasets[35]. Applied on the converted spatial distance matrices of single cells, FLAMINGO generates the high-resolution spatial coordinates of 3D structures for chromosomes of each individual cell.

Performance evaluation based on simulated chromatin structures

The performance of Tensor-FLAMINGO is extensively benchmarked by a series of simulations, with a number of parameters systematically varied. For each simulated dataset, multiple different consensus structures (m structures) with equal chromosome size l (i.e. l beads) are used to generate a mixed pool of single cell structures, in order to account for the heterogeneity of different cell types. The underlying consensus structures are generated with structural variations to each other, and the structural similarity across different consensus structures are controlled by the similarity parameter $W(0 < W < 1)$. The l by l pairwise spatial distance matrix of each consensus structure is used to generate N different distance matrices of single cells, based on random down-sampling of entries and adding random noise. Therefore, each simulated dataset is organized into a sparse tensor with mN frontal slices, which is used as the input for Tensor-FLAMINGO to reconstruct the 3D structures of mN single cells. To test the performance of Tensor-FLAMINGO under different conditions, a series of simulated datasets are generated based on combinatorial settings of the key parameters with different values, including: (1) the similarity parameter $W(0 < W < 1)$ to control the similarity across different cell types in the pool; (2) the down-sampling rate $\gamma(0 < \gamma < 1)$ to control the missing rate in single-cell distance matrices; (3) the noise level to control the random variations of single-cell specific observations; (4) the number of mixing cell-types, m, in the simulated pool of structures; (5) the number of single cells, N, belonging to each cell type; (6) the length of the chromosome structure l, and (7) the organizing order of frontal slices in the input tensor. The added white noise takes three different levels as suggested by previous studies[71], including level-zero: no noise; level-one: generated by the normal distribution $Normal(\delta, \delta)$; and level-two: generated by $Normal(2\delta, \delta)$; where δ represents the minimum value from the down-sampled pairwise distances[71]. For other parameters, a wide range of values of each parameter is used to create different sets of simulated structures to benchmark the robust performance systematically.

The reconstructed single-cell structures from Tensor-FLAMINGO are compared to the benchmark structures of simulations for performance evaluation based on two metrics: (1) Spearman correlations of the distance matrices between the predicted and the benchmark structures; and (2) the RMSD of 3D coordinates between the predicted (C^{pred}) and the benchmark structures ($C^{benchmark}$): $RMSD = \sqrt{\frac{1}{l} \sum_{i=1}^{l} ||C_i^{pred} - C_i^{benchmark}||^2}$. Both of these two metrics have been widely used in previous studies for performance quantifications[27,49,50].

Performance comparison based on STORM 3D genome imaging data of single cells

To quantitatively evaluate the performance of predicting single-cell specific 3D structures, the multiplexed single-cell STORM 3D genome imaging data[52] of the human chromosome 21 in K562 is collected as the orthogonal benchmark testing dataset. This dataset is used to compare the accuracy of different algorithms' predictions based on the K562 snHi-C chromatin contact data of 16 cells as inputs[44]. As data quality control, single-cell structures measured by the STORM dataset with missing rates greater than 0.5 are first excluded from the analysis. Because the STORM and snHi-C datasets used different populations of K562 cells, the single-cell structures from these two experiments are not directly matched to each other. Therefore, the cells of the STORM dataset are further filtered to identify the ones that are highly similar to the 16 cells of the snHi-C dataset, by calculating their correlations to the input snHi-C chromatin contact maps. For each single cell of the snHi-C dataset, the top 20 STORM structures with the highest correlations are considered as the matched structures and are included for

the downstream benchmark analysis. Hence, in total, 320 high-quality cells from the STORM dataset with matching structures are used as the benchmark structures for the performance evaluation at the single-cell level (Fig. 3c and d). Different algorithms, as summarized in the next section, are tested, and the predicted single-cell specific structures based on the same input snHi-C dataset are evaluated based on their Spearman correlations to the same benchmark STORM single-cell structures.

Performance comparison across diverse experimental contexts based on orthogonal chromatin contact data

The performance of Tensor-FLAMINGO is further evaluated across different cell types (i.e. human GM12878, human K562 and mouse ESC cells), different types of inputs of single-cell chromatin contacts (i.e. Dip-C, snHi-C, scHi-C and snm3C)[40,44,45,47], different levels of resolutions (i.e. 10kb- and 30kb-resolution), and different types of benchmark datasets as gold-standards (i.e. 3D ATAC-PALM, GAM, and bulk Hi-C). To compare with predicted structures, the chromatin contact frequencies of benchmark datasets are converted into spatial pairwise distances of genomic regions, based on the exponential transformation function as explained above[12]. Spearman correlations between the predicted and the observed benchmark spatial pairwise distances are used as the metric of accuracy as suggested by previous studies[27]. This metric, termed 'Accuracy (global)', evaluates the overall accuracy of predictions for both observed and imputed entries in single-cell contact maps. In addition, to specifically evaluate the accuracy of predicted spatial distances for the subset of observed entries (Ω) in single-cell contact maps, a second metric of accuracy based on Spearman correlations, termed 'Accuracy (measured subset)', is calculated in the same way as 'Accuracy (global)' but restricted to the subset of single-cell observed entries.

There are currently only six algorithms that have been developed or suggested to reconstruct single-cell chromosomal structures, while other methods[28,29,31,32] can only work on bulk-level data. Out of the six algorithms, MBO[72] is excluded from the analysis due to the unavailability of code. Therefore, Tensor-FLAMINGO is compared to five algorithms: ShRec3D[27], NucDynamics[40], RPR[30], isdHi-C[49] and Si-C[50]. Because the code of NucDynamics is unable to run, the available predictions of NucDynamics based on mESC scHi-C are directly downloaded for the corresponding comparisons. The other four algorithms are all applied on the same datasets using their suggested parameter settings to predict single-cell 3D chromatin structures at 10kb- and 30kb-resolution. The same performance evaluations using Spearman correlations to orthogonal gold-standard datasets as described above are implemented on the predicted structures from different algorithms.

In addition, the cell-type specificity of predicted single-cell 3D chromosomal structures are evaluated and compared across different algorithms. Tensor-FLAMINGO and the four algorithms are applied on the snm3C data[47], which contains 351 mESC cells and 96 NMuMG cells, to reconstruct the 3D chromatin structures for every cell. For each algorithm, UMAP plots based on the predicted distance matrices are generated to visualize the two types of single cells. An improved separation of cell-type clustering of single cells is expected if an algorithm achieves better performance of reconstructing 3D structures of single cells that belong to different cell types.

Performance comparison in imputing high-resolution single-cell chromatin contacts and classifying cell types

The reconstructed 3D structures from Tensor-FLAMINGO generate the completed matrices of single-cell chromatin contacts. The completed matrices provide high-resolution imputed values for the missing data of the sparse single-cell 3C-based experiments. The accuracy of the imputed high-resolution single-cell chromatin contacts by Tensor-FLAMINGO is evaluated by their correlations to the observed

chromatin contacts from bulk-tissue experiments. The usefulness of the imputed chromatin contacts in classifying single cells belonging to different cell types is quantified by the Adjusted Random Index (ARI) between the predicted cluster memberships (based on k-means clustering) and the real cell-type identities of single cells. The performance is benchmarked by applying Tensor-FLAMINGO on the snm3C dataset at different resolutions (i.e. 1Mb-, 250kb- and 30kb-resolution). In comparison, the Higashi algorithm[56], which imputes single-cell chromatin contacts, is applied on the same dataset and is evaluated by the same metrics.

Identify single-cell chromatin compartments, TAD boundaries and structural variabilities

The completed chromatin contact matrices of single cells are derived from the reconstructed 3D structures and are further normalized based on the standard protocol of previous studies[6]. Principal component analysis is implemented on the normalized chromatin contact matrix of each single cell and the PC1 scores are calculated to identify chromatin compartments A/B[5]. Topologically associated domains (TADs) and their boundaries are called based on the completed single-cell chromatin contact matrices by the TADCompare software[73].

To quantify the structural variation of specific genomic regions across different cells, the RMSD metrics between single-cell structures and the pooled average structure is calculated: $RMSD_i^k = |Coord_i^k - Coord_i^{pooled}|$, where $RMSD_i^k$ represents the RMSD in single cell k at genomic location i. The distribution of $RMSD_i$ values across all single cells demonstrates the overall degree of structural variability for the specific genomic location i.

Differentially methylated gene analysis across clusters of single-cell structures

The 351 mESC cells from the snm3C dataset[47] are analyzed to delineate the relationship between high-resolution 3D chromatin structures and gene regulation. Hierarchical clustering is applied on the reconstructed single-cell structures by Tensor-FLAMINGO for these 351 cells. Two cell clusters with distinct 3D structures are identified based on the highest average Silhouette score. The coupled single-cell DNA methylation signals from the same snm3C dataset at gene promoters are used to quantify the epigenetic environment and regulation of genes in different cells. DEseq2 package[74] with default settings is used to call differentially methylated genes across the two clusters of single cells.

Identify statistically significant single-cell chromatin interactions

To predict significant single-cell chromatin interactions of GM12878, the distance matrices induced by the predicted high-resolution 3D structures are converted to chromatin contact matrices using the same exponential conversion function as described above. FitHi-C[57] is then applied to identify the statistically significant chromatin interactions. To control the false positive rates, a stringent p-value threshold is selected as 1×10^{-20} [57]. As support, the identified single-cell chromatin interactions are compared to the GM12878 Capture-C dataset[10] and the fractions of overlapping with Capture-C interactions are calculated. For each significant chromatin interaction, its occurrence frequency among single cells is calculated. The overall variability across single cells is further analyzed by the distributions of occurrence frequencies, i.e. the spectrum, of significant chromatin interactions, across a diverse panel of resolutions.

Predict single-cell cis-regulatory links based on high-resolution 3D structures

Based on the snm-3C dataset with coupled single-cell chromatin contact and DNA methylation signals[47], Lasso regression models are built to identify single-cell cis-regulatory regions linked to the promoters of

specific target genes, using the predicted high-resolution 3D spatial distances as the only features. For each gene, the promoter DNA methylation signals across single cells are used to represent its single-cell activities. All flanking genomic bins within +/-3Mb from its promoter are considered in the model, and the predicted 3D spatial distances of each bin to the promoter are used as features. Based on the fitted Lasso regression model, the flanking genomic bins are ranked by their Lasso effect sizes, where larger Lasso effect sizes suggest stronger association between the single-cell gene activity and the spatial proximity of the specific genomic bins in 3D space. For each gene, the top-ranking genomic bins are prioritized as the candidate cis-regulatory elements, including long-range regulatory links from distal genomic regions.

Interpret disease-associated genetic variants and eQTLs

The genetic variants that are significantly associated with LAML are collected from the TCGA dataset[60]. The fractions of these genetic variants that overlap with genomic regions of predicted GM12878 chromatin interactions in different single cells are calculated. In comparison, the background set of the same number of random chromatin interactions in each cell is generated, with genomic distances strictly controlled to follow the same distribution as the foreground. The fractions of genetic variants overlapping with genomic regions of random single-cell chromatin interactions are also calculated. One-sided Mann-Whitney U test is used to evaluate the statistical significance of functional genetic variant enrichments in genomic regions with predicted single-cell chromatin interactions.

The SNP-gene pairs of cis-eQTLs are collected for immune-related cell types[58]. The fractions of these eQTL SNP-gene pairs that overlap with GM12878 chromatin interactions in different single cells are calculated. The background set of random chromatin interactions in each cell is generated with genomic distances controlled and the fractions of eQTL SNP-gene pairs overlapping with random single-cell chromatin interactions are calculated. One-sided Mann-Whitney U test is used to evaluate the enrichment of cis-eQTLs in predicted single-cell chromatin interactions. As comparison, the fractions of cis-eQTLs overlapping with single-cell chromatin interactions from the raw input GM12878 Dip-C contact matrices are also calculated.

Discover single-cell three-way chromatin interactions at high-resolution

Based on the predicted single-cell 3D chromatin structures, three-way chromatin interactions are defined as groups of three genomic regions that are spatially proximal to each other. For every combination of three genomic regions along a chromosome, the average pairwise spatial distances are calculated to quantify the spatial compactness. Specifically, for a set of genomic regions i, j, k ($i < j < k$), the average 3D pairwise distance is calculated as: $D_{i,j,k} = \frac{1}{3}(D_{ij} + D_{ik} + D_{jk})$, where D_{ij} represents the predicted 3D genomic distance between genomic regions i and j. As comparison, the average spatial distances of 1000 sets of genomic regions with the same 1D genomic distances are calculated: $D^{bg}_{m,n,p} = \frac{1}{3}(D_{mn} + D_{mp} + D_{np})$, where $n - m = j - i$ and $p - n = k - j$, in order to control the 1D genomic distances. The empirical p-values are calculated as $P_{ijk} = \frac{1}{1000}(1 + count\{D^{bg} < D_{ijk}\})$. As suggested by previous studies on the identification of two-way chromatin interactions[6], the adjacent three-way interactions are pruned and the most significant ones are selected: $SI_{i,j,k} = argmin_{m,n,p}\{P_{mnp} || m - i| < 5, |n - j| < 5, |p - k| < 5\}$.

Statistics and reproducibility

In performance comparison, all methods are independently applied to reconstruct the single-cell chromosomal 3D structures in the human genome. The sample sizes of different algorithms are determined by the number of completed reconstructions. In the analyses of bulk tissue chromatin contact capture datasets and the eQTL datasets, the sample sizes are determined by the original datasets and no data points are excluded from the analysis. No statistical method was used to predetermine the sample size. The preprocessing steps for the STORM dataset are described above (see the section Performance comparison based on STORM 3D genome imaging data of single cells). In the simulation analysis and the down-sampling analysis, the performance is evaluated based on randomly down-sampled datasets, and no data points are excluded from the analyses. To reproduce the analysis, the predicted 3D structures from four single-cell datasets at 10kb- and 30kb-resolution and the simulated datasets are provided (see Data Availability and Code Availability).

Reporting summary

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

Data availability

The data generated in this study, including the code of Tensor-FLA-MINGO, the simulated data, the predicted single-cell chromosome structures at 10 kb and 30 kb resolutions and the intermediate analysis results, are available in Tensor-FLAMINGO GitHub repository: https://github.com/wangjr03/Tensor-FLAMINGO. In addition, a Zenodo repository is also created to access the results and the code of Tensor-FLAMINGO: https://doi.org/10.5281/zenodo.13645233. Dip-C data in GM12878 used in this study are available under GSE117876. snHi-C data in K562 are available under GSE80006. scHi-C data in mESC are available under GSE80280. snm3C data in mESC and NMuMG are available under GSE124391. Bulk-tissue Hi-C data in GM12878 and K562, along with the annotations of chromatin compartments and TADs, are available under GSE63525 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi]. Bulk-tissue Hi-C data in mESC are available at the 4D Nucleome Data Portal under 4DNFI5IAH9H1. 3D ATAC-PALM data in mESC are available under GSE126112. Capture-C data in GM12878 are available under GSE86189. STORM data is downloaded from https://github.com/BogdanBintu/ChromatinImaging/tree/master/Data. Single-cell SPRITE data are available under GSE154353. ChIP-seq data are downloaded from the GEO database (CTCF:, RAD21:). Gene expression data are available at the Roadmap data consortia [https://egg2.wustl.edu/roadmap/web_portal/processed_data.html]. The eQTL dataset in K562 is collected from the MuTHER consortia (http://www.muther.ac.uk/Data.html). TCGA-LAML mutations are available at the GDC portal [https://portal.gdc.cancer.gov/projects/TCGA-LAML]. Source data are provided as a Source Data file. Source data are provided with this paper.

Code availability

The Tensor-FLAMINGO software, along with all the predicted 10kb- and 30kb-resolution single-cell chromosomal structures, is available at GitHub: https://github.com/wangjr03/Tensor-FLAMINGO. The repository also provides scripts and instructions for structure visualizations, sample codes, sample input data, intermediate results, and information of relevant datasets to reproduce the analyses. In addition, a Zenodo repository for Tensor-FLAMINGO[75] is also created: https://doi.org/10.5281/zenodo.13645233.

References

1. Sexton, T., Schober, H., Fraser, P. & Gasser, S. M. Gene regulation through nuclear organization. *Nat. Struct. Mol. Biol.* **14**, 1049–1055 (2007).
2. Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic plasticity and the hallmarks of cancer. *Science* **357**, eaal2380 (2017).
3. Dion, V. & Gasser, S. M. Chromatin movement in the maintenance of genome stability. *Cell* **152**, 1355–1364 (2013).

4. Liu, M. et al. Multiplexed imaging of nucleome architectures in single cells of mammalian tissue. *Nat. Commun.* **11**, 2907 (2020).
5. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
6. 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).
7. Dixon, J. R. et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
8. Schmitt, A. D. et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* **17**, 2042–2059 (2016).
9. Mifsud, B. et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet* **47**, 598–606 (2015).
10. Jung, I. et al. A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nat. Genet* **51**, 1442–1449 (2019).
11. Fullwood, M. J. et al. An oestrogen-receptor-α-bound human chromatin interactome. *Nature* **462**, 58–64 (2009).
12. Wang, S. et al. Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353**, 598–602 (2016).
13. Chubb, J. R., Boyle, S., Perry, P. & Bickmore, W. A. Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr. Biol.* **12**, 439–445 (2002).
14. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
15. Sanborn, A. L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl Acad. Sci.* **112**, E6456–E6465 (2015).
16. Misteli, T. The self-organizing genome: principles of genome architecture and function. *Cell* **183**, 28–45 (2020).
17. Bonev, B. et al. Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572.e24 (2017).
18. Wang, H., Huang, B. & Wang, J. Predict long-range enhancer regulation based on protein–protein interactions between transcription factors. *Nucleic Acids Res* **49**, 10347–10368 (2021).
19. Zuin, J. et al. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl Acad. Sci.* **111**, 996–1001 (2014).
20. Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
21. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A phase separation model for transcriptional control. *Cell* **169**, 13–23 (2017).
22. Sabari, B. R. et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018).
23. 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).
24. Zhang, Y. et al. Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus. *Nat. Commun.* **9**, 750 (2018).
25. Yu, M. et al. SnapHiC: a computational pipeline to identify chromatin loops from single-cell Hi-C data. *Nat. Methods* **18**, 1056–1059 (2021).
26. Kaul, A., Bhattacharyya, S. & Ay, F. Identifying statistically significant chromatin contacts from Hi-C data with FitHiC2. *Nat. Protoc.* **15**, 991–1012 (2020).
27. Lesne, A., Riposo, J., Roger, P., Cournac, A. & Mozziconacci, J. 3D genome reconstruction from chromosomal contacts. *Nat. Methods* **11**, 1141–1143 (2014).
28. Abbas, A. et al. Integrating Hi-C and FISH data for modeling of the 3D organization of chromosomes. *Nat. Commun.* **10**, 2049 (2019).
29. Trieu, T., Oluwadare, O. & Cheng, J. Hierarchical reconstruction of high-resolution 3D models of large chromosomes. *Sci. Rep.* **9**, 4971 (2019).
30. Hirata, Y., Oda, A., Ohta, K. & Aihara, K. Three-dimensional reconstruction of single-cell chromosome structure using recurrence plots. *Sci. Rep.* **6**, 34982 (2016).
31. Zhang, Y., Liu, W., Lin, Y., Ng, Y. K. & Li, S. Large-scale 3D chromatin reconstruction from chromosomal contacts. *BMC Genomics* **20**, 186 (2019).
32. Li, F.-Z. et al. Chromatin 3D structure reconstruction with consideration of adjacency relationship among genomic loci. *BMC Bioinforma.* **21**, 272 (2020).
33. Hu, M. et al. Bayesian inference of spatial organizations of chromosomes. *PLoS Comput Biol.* **9**, e1002893 (2013).
34. Rousseau, M., Fraser, J., Ferraiuolo, M. A., Dostie, J. & Blanchette, M. Three-dimensional modeling of chromatin structure from interaction frequency data using Markov chain Monte Carlo sampling. *BMC Bioinforma.* **12**, 414 (2011).
35. Wang, H., Yang, J., Zhang, Y., Qian, J. & Wang, J. Reconstruct high-resolution 3D genome structures for diverse cell-types using FLA-MINGO. *Nat. Commun.* **13**, 2645 (2022).
36. Tjong, H. et al. Population-based 3D genome structure analysis reveals driving forces in spatial genome organization. *Proc. Natl Acad. Sci.* **113**, E1663–E1672 (2016).
37. Boninsegna, L. et al. Integrative genome modeling platform reveals essentiality of rare contact events in 3D genome organizations. *Nat. Methods* **19**, 938–949 (2022).
38. Qi, Y. & Zhang, B. Predicting three-dimensional genome organization with chromatin states. *PLoS Comput Biol.* **15**, e1007024 (2019).
39. Nagano, T. et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).
40. Stevens, T. J. et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017).
41. Nagano, T. et al. Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61–67 (2017).
42. Ramani, V. et al. Massively multiplex single-cell Hi-C. *Nat. Methods* **14**, 263–266 (2017).
43. Kim, H.-J. et al. Capturing cell type-specific chromatin compartment patterns by applying topic modeling to single-cell Hi-C data. *PLoS Comput Biol.* **16**, e1008173 (2020).
44. Flyamer, I. M. et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **544**, 110–114 (2017).
45. 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).
46. 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).
47. Lee, D.-S. et al. Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. *Nat. Methods* **16**, 999–1006 (2019).
48. Li, G. et al. Joint profiling of DNA methylation and chromatin architecture in single cells. *Nat. Methods* **16**, 991–993 (2019).
49. Carstens, S., Nilges, M. & Habeck, M. Inferential structure determination of chromosomes from single-cell Hi-C data. *PLoS Comput Biol.* **12**, e1005292 (2016).
50. Meng, L., Wang, C., Shi, Y. & Luo, Q. Si-C is a method for inferring super-resolution intact genome structure from single-cell Hi-C data. *Nat. Commun.* **12**, 4369 (2021).

51. Lu, C., Feng, J., Lin, Z. & Yan, S. Exact Low Tubal Rank Tensor Recovery from Gaussian Measurements. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence* 2504–2510 (International Joint Conferences on Artificial Intelligence Organization, California, 2018). https://doi.org/10.24963/ijcai.2018/347.

52. Bintu, B. et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, eaau1783 (2018).

53. Du, Z. et al. Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* **547**, 232–235 (2017).

54. Xie, L. et al. 3D ATAC-PALM: super-resolution imaging of the accessible genome. *Nat. Methods* **17**, 430–436 (2020).

55. Beagrie, R. A. et al. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524 (2017).

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

57. Ay, F., Bailey, T. L. & Noble, W. S. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res* **24**, 999–1011 (2014).

58. Grundberg, E. et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat. Genet* **44**, 1084–1089 (2012).

59. Delaneau, O. et al. Chromatin three-dimensional interactions mediate genetic effects on gene expression. *Science* **364**, eaat8266 (2019).

60. Weinstein, J. N. et al. The cancer genome atlas pan-cancer analysis project. *Nat. Genet* **45**, 1113–1120 (2013).

61. Fu, C. et al. RNA-binding motif protein 11 (RBM11) serves as a prognostic biomarker and promotes ovarian cancer progression. *Dis. Markers* **2021**, 1–7 (2021).

62. Anastasiadi, D., Esteve-Codina, A. & Piferrer, F. Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics Chromatin* **11**, 37 (2018).

63. Gao, T. et al. EnhancerAtlas: a resource for enhancer annotation and analysis in 105 human cell/tissue types. *Bioinformatics* **32**, 3543–3551 (2016).

64. Arrastia, M. V. et al. Single-cell measurement of higher-order 3D genome organization with scSPRITE. *Nat. Biotechnol.* **40**, 64–73 (2022).

65. Zheng, Y., Shen, S. & Keleş, S. Normalization and de-noising of single-cell Hi-C data with BandNorm and scVI-3D. *Genome Biol.* **23**, 222 (2022).

66. Yaffe, E. & Tanay, A. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat. Genet* **43**, 1059–1065 (2011).

67. Hu, M. et al. HiCNorm: removing biases in Hi-C data via Poisson regression. *Bioinformatics* **28**, 3131–3133 (2012).

68. Edgar, R. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207–210 (2002).

69. Dekker, J. et al. The 4D nucleome project. *Nature* **549**, 219–226 (2017).

70. Knight, P. A. & Ruiz, D. A fast algorithm for matrix balancing. *IMA J. Numer. Anal.* **33**, 1029–1047 (2013).

71. Tasissa, A. & Lai, R. Exact reconstruction of euclidean distance geometry problem using low-rank matrix completion. *IEEE Trans. Inf. Theory* **65**, 3124–3144 (2019).

72. Paulsen, J., Gramstad, O. & Collas, P. Manifold based optimization for single-cell 3d genome reconstruction. *PLoS Comput Biol.* **11**, e1004396 (2015).

73. Cresswell, K. G. & Dozmorov, M. G. TADCompare: An R package for differential and temporal analysis of topologically associated domains. *Front Genet* **11**, 158 (2020).

74. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

75. Wang, H. et al. Tensor-FLAMINGO unravels the complexity of single-cell spatial architectures of genomes at high-resolution. Repository name: Tensor-FLAMINGO. *Zenodo* https://doi.org/10.5281/zenodo.13645233 (2024).

Acknowledgements

Author contributions

Y.Z., J.Q., and J.W. designed the study and supervised the overall research. H.W. developed the model, implemented the algorithm and performed data analyses. J.Y. and X.Y. participated in data analyses. Y.Z. participated in biological data analyses. H.W., Y.Z., J.Q., and J.W. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-58674-w.

Correspondence and requests for materials should be addressed to Yu Zhang, Jianliang Qian or Jianrong Wang.

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

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.