Ji *et al. Genome Biology* (2025) 26:340
https://doi.org/10.1186/s13059-025-03823-z

Genome Biology

METHODOLOGY

Open Access

Cisformer: a scalable cross-modality generation framework for decoding transcriptional regulation at single-cell resolution

Luzhang Ji[1,2,3†], Qihang Zou[1,2,3†], Ke Tang[1,2,3] and Chenfei Wang[1,2,3,4,5*]

†Luzhang Ji and Qihang Zou contributed equally to this work.

*Correspondence: 08chenfeiwang@tongji.edu.cn

[1] Key Laboratory of Spine and Spinal Cord Injury Repair and Regeneration of Ministry of Education, Department of Orthopedics, Tongji Hospital, School of Life Sciences and Technology, Tongji University, Shanghai 200092, China
[2] Sycamore Research Institute of Life Sciences, Shanghai 201210, China
[3] Frontier Science Center for Stem Cell Research, Tongji University, Shanghai 200092, China
[4] National Key Laboratory of Autonomous Intelligent Unmanned Systems, Tongji University, Shanghai 201210, China
[5] Frontier Science Center for Intelligent Autonomous Systems, Tongji University, Shanghai 201210, China

Abstract

Single-cell multiomic technologies enable the joint analysis of different modalities, but face challenges due to experimental complexity. Current computational methods for single-cell cross-modality translation lack biological interpretability. Here, we present Cisformer, a cross-attention-based generative model tailored for cross-modality generation between gene expression and chromatin accessibility at single-cell resolution. Systematic benchmarking demonstrates the superior accuracy and generalization of Cisformer against existing methods. Cisformer leverages its inherent interpretability to precisely link *cis*-regulatory elements to target genes, facilitating the identification of functional transcription factors associated with tumorigenesis and aging. Overall, Cisformer is a powerful tool for single-cell multiomic data analysis.

Keywords: Single-cell multiomics, Cross-modality generation, Transcriptional regulation, *Cis*-regulatory element, Transcription factor, Transformer, Model interpretability

Background

Single-cell technologies have been widely used to characterize complex cellular processes, measuring diverse modalities such as gene expression, chromatin accessibility, DNA methylation, genome organization, and protein abundance [1]. The emergence of single-cell multiomics methods enables the simultaneous profiling of multiple modalities within the same cell. For instance, SNARE-seq [2] and SHARE-seq [3] measure gene expression and chromatin accessibility, CITE-seq [4] and REAP-seq [5] quantify transcriptome and protein markers jointly, and HiRES [6] and GAGE-seq [7] combine higher-order chromatin organization with RNA abundance. These paired single-cell methods provide valuable insights into the regulatory mechanisms across different layers within a cell. However, single-cell multiomics methods face several limitations.

Careful specimen preparation and preservation are required to effectively capture target molecules, which increases experimental complexity. Furthermore, the high costs of multiomics profiling protocols limit their scalability. Lastly, multiomic profiling data tend to be more prone to noise and sparsity compared to standard single-cell data. Consequently, it can be time-consuming and often impractical to jointly profile single cells in many instances, prompting the need to explore information generation for one modality when another is available.

Several computational approaches have been developed to enable cross-modality generation at the single-cell level. For example, BABEL achieves cross-omics generation using two autoencoders (AE) [8]. Polarbear predicts missing modalities through a semi-supervised variational autoencoder (VAE) framework trained on both single-assay and co-assay data [9]. More recently, scButterfly employs a dual-aligned variational autoencoder for single-cell cross-modality prediction [10]. Despite these advances, existing methods still face several limitations. First, generation accuracy remains suboptimal, primarily due to constraints imposed by both model architecture and the inherent sparsity of single-cell data. Second, model generalizability requires further improvement, particularly for challenging cross-tissue-type prediction tasks that are crucial for practical applications. Most importantly, model interpretability in uncovering the interplay between modalities is often overlooked. Most existing single-cell cross-modality generation models rely on AE or VAE frameworks, making it difficult to extract meaningful biological insights from the models.

Understanding the transcriptional regulation of genes by *cis*-regulatory elements (CREs), such as promoters and enhancers, is fundamental to deciphering gene regulatory programs underlying cellular functions [11]. Single-cell multiome technologies that simultaneously profile gene expression and chromatin accessibility within individual cells offer an unprecedented opportunity to study these dynamics at cellular resolution. Current computational approaches for linking CREs to genes using single-cell multiome data are primarily correlation-based or linear models. For instance, ArchR leverages correlations between chromatin accessibility at CREs and gene expression levels to identify gene-associated CREs [12]. While this method offers a straightforward implementation, it is highly susceptible to confounding factors. In contrast, SCARlink employs regularized Poisson regression to model gene-level regulatory effects [13]. This approach provides good interpretability but struggles to capture the complex and non-linear relationships between CREs and their target genes. These limitations highlight the need for more robust and interpretable methods that can effectively integrate molecular information from both the transcriptome and epigenome.

In this study, we present Cisformer, a cross-attention-based generative model for single-cell cross-modality generation. We focus on modality translation between gene-wide single-cell RNA sequencing (scRNA-seq) data and genome-wide single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq) data, as existing datasets are more comprehensive and abundant. More importantly, our primary interest lies in understanding the interaction of transcriptional programs with epigenomic accessibility. Cisformer employs a decoder-only architecture with a cross-attention mechanism, striking a balance between model complexity and biological interpretability. A key innovation of Cisformer is its feature duplication and index encoding strategy, specifically

designed to address the challenge of processing ultra-long sequences from chromatin accessibility data. Through comprehensive benchmarking, we demonstrate the superior performance of Cisformer in the single-cell RNA-ATAC translation task across various biological systems and species. By leveraging the cross-attention mechanism, Cisformer effectively captures the intricate interactions between gene expression and chromatin accessibility. We further apply Cisformer in cancer and aging-related scenarios to identify potential functional CREs and transcription factors (TFs), thereby advancing our understanding of the molecular mechanisms underlying these critical biological processes.

Results

Overview of the Cisformer framework

Cisformer utilizes a Transformer-based architecture to facilitate cross-modality generation at single-cell resolution. To achieve a balance between model complexity and biological interpretability, we implemented a decoder-only structure featuring cross-attention mechanisms (Fig. 1a, Additional file 1: Fig. S1). Unlike AE or VAE-based approaches, our model does not apply dimensionality reduction to input genes or chromatin peaks.

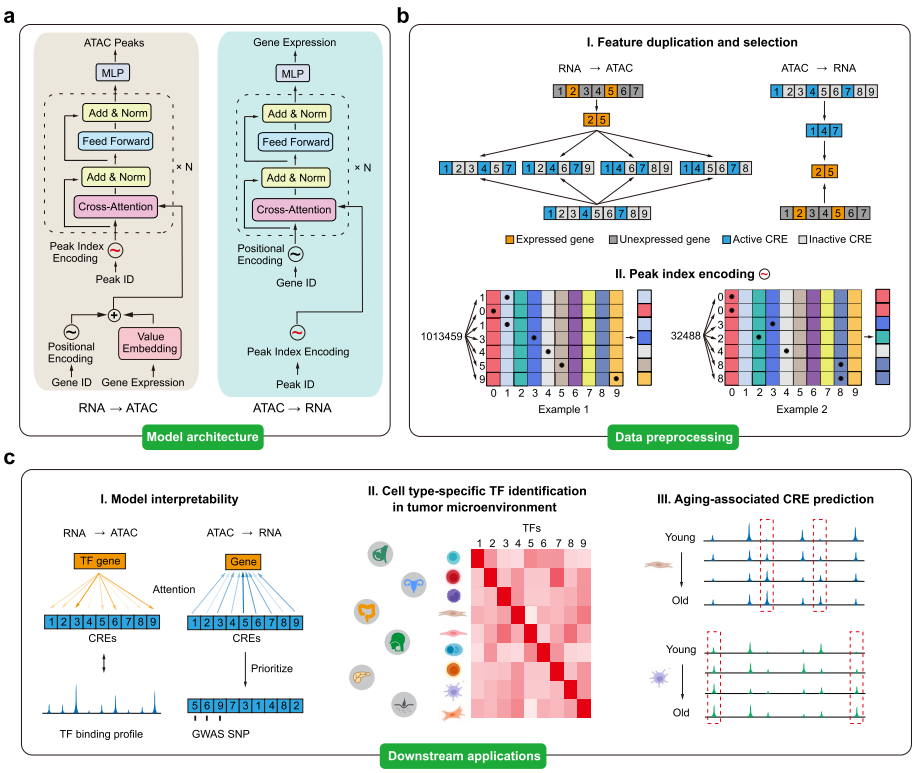

Fig. 1 Overview of the Cisformer model. **a** The model architecture of Cisformer for single-cell RNA-to-ATAC (left) and ATAC-to-RNA (right) generation. Gene and peak features are encoded respectively, integrated through Transformer blocks with cross-attention mechanisms, and subsequently transformed into final outputs by an MLP. **b** The genes and chromatin peaks are selected or duplicated to generate gene-peak pairs as model input (top). Chromatin peak indices are encoded through a digit decomposition strategy, with 1,013,459 and 32,488 as representative examples (bottom). **c** The downstream applications of Cisformer, including validating biological interpretability (left), identifying cell-type-specific TFs in pan-cancer datasets (middle), and predicting aging-related CREs (right)

Recognizing the challenges that attention-based models encounter when processing long sequences, particularly for chromatin peaks, we propose an innovative feature duplication and selection strategy. In the RNA-to-ATAC generation, we focus on expressed genes (non-zero expression), as this subset is more biologically significant. For chromatin peaks, we first select the active CREs after binarization (value of 1) and then balance the sequences by incorporating an equal number of inactive CREs (value of 0) (Fig. 1b, top left). This process generates multiple pseudo-cells from a single original cell, serving as a form of data augmentation. These gene-peak pairs are then used as input for model training. In the inference stage, Cisformer takes expressed genes from scRNA-seq data as input and predicts chromatin accessibility profiles covering all CREs. In the ATAC-to-RNA direction, we construct gene-peak pairs after filtering out inactive genes or peaks, adhering to biological principles (active CREs drive target gene expression) and enhancing computational efficiency (Fig. 1b, top right). Consequently, during the inference stage, Cisformer relies on prior knowledge of expressed genes (e.g., derived from single-cell multiome data) to generate a full transcriptome from chromatin accessibility profiles. This improves the accuracy of linking CREs to their target genes, albeit at the cost of reduced generative capacity for unexpressed genes. Another key innovation is the development of a novel indexing method for processing millions of chromatin peaks. As illustrated in the examples, each digit of the peak index is individually extracted and embedded, with the resulting representations subsequently combined (Fig. 1b, bottom). This peak index encoding strategy is more effective than directly embedding the index as a whole. Model training is guided by a categorical cross-entropy (CCE) loss to quantify the discrepancy between generated and measured RNA profiles, whereas a binary cross-entropy (BCE) loss assesses the accuracy of inferred ATAC values. Upon achieving high generation accuracy, Cisformer's generative capacity and intrinsic interpretability via attention mechanism enable the identification of cell type-specific TFs and functional CREs within the context of the intricate biological processes, including cancer development and aging (Fig. 1c).

Cisformer performs cross-omics generation with high accuracy and robust generalization

To evaluate Cisformer's performance in single-cell multi-omics generation, we systematically benchmarked our model against published state-of-the-art methods. For RNA-to-ATAC generation, we designed four evaluation scenarios: cell-level train-test splitting within a dataset (intra-dataset 1), cell-type-level train-test splitting within a dataset (intra-dataset 2), training on one tissue and testing on a similar tissue (inter-dataset 1), and training on one tissue and testing on a distinct tissue (inter-dataset 2) (Additional file 1: Fig. S2a). Inferred chromatin accessibility was assessed using cell clustering metrics, including adjusted mutual information (AMI), normalized mutual information (NMI), adjusted Rand index (ARI), and homogeneity score (HOM). In the "intra-dataset 1" scenario, we randomly split the peripheral blood mononuclear cell (PBMC) multiome dataset from 10X Genomics (9964 cells) into 80% for training and 20% for testing. In the "intra-dataset 2" scenario, we designated naïve $CD4^+$ T cells (CD4 naïve), natural killer cells (NK), regulatory T cells (Treg), type-2 conventional dendritic cells (cDC2), and plasmacytoid dendritic cells (pDC) as the testing set (1816 cells), and used the remaining cells for training. This cell-type-level splitting strategy ensures the testing set comprises

cell types as distinct as possible from those in the training set while retaining sufficient cells in the training set for effective model training. Cisformer demonstrated marginally superior performance compared to two existing methods, BABEL [8] and scButterfly [10], in both intra-dataset scenarios (Fig. 2a, left). From a practical perspective, inter-dataset prediction is more valuable than intra-dataset generation. In the inter-dataset scenarios, we trained Cisformer using PBMC multiome data again and then tested it on datasets from bone marrow mononuclear cells (BMMC; inter-dataset 1) and brain tissue (inter-dataset 2). Cisformer substantially outperformed BABEL and scButterfly in both inter-dataset scenarios (Fig. 2a, right). In the "inter-dataset 2" scenario, cells clustering based on brain chromatin accessibility profiles inferred by Cisformer closely resembled that of the measured ATAC data, while profiles generated by BABEL or scButterfly only distinguished microglia from other cell types (Fig. 2b). We further investigated the chromatin accessibility landscape surrounding the *MSI2* (an RNA-binding protein gene) locus. A putative intronic enhancer within *MSI2* exhibited astrocyte-specific openness, consistent with the elevated expression level of *MSI2* in astrocytes. Among the three methods, only Cisformer accurately recapitulated the cell-type-specific chromatin accessibility of this regulatory element in the brain (Fig. 2c). To further evaluate the accuracy of the generated chromatin accessibility profiles, we employed more direct metrics at the peak levels. Using precision, recall, and F1 score to assess the overlap between predicted and ground truth peaks, Cisformer exhibited comparable precision, substantially enhanced recall (~ 50% increase), and higher F1 scores at the cell level relative to BABEL and scButterfly across all evaluation scenarios (Additional file 1: Fig. S2b). A quantitative analysis based on Pearson correlation coefficients revealed that Cisformer's predicted ATAC signals showed stronger agreement with experimental data at the cell-type level compared to both BABEL and scButterfly (~ 15% increase) (Additional file 1: Fig. S2c). These results highlight the superior predictive accuracy and generalization capacity of Cisformer in translating the transcriptome to the epigenome at single-cell resolution.

Given our aim to link CREs and their target genes rather than reconstruct complete gene expression profiles, we focused on intra-dataset prediction in the ATAC-to-RNA generation task. For a fair comparison, we benchmarked the generation performance of Cisformer against two representative methods, ArchR [12] and SCARlink [13], all of

(See figure on next page.)

Fig. 2 Benchmarking model performance of Cisformer in cross-modality generation. **a** Barplots presenting four cell clustering metrics (AMI, NMI, ARI, and HOM) for predicted chromatin accessibility profiles by three models (BABEL, scButterfly, and Cisformer) in four different scenarios. **b** UMAP visualization of cells in the brain dataset using raw ATAC profiles and generated profiles from different methods. IT, intratelencephalic neuron; OPC, oligodendrocyte precursor cell; PVALB, PVALB$^+$ neuron; SST, SST$^+$ neuron; VIP, VIP$^+$ neuron. **c** Genomic tracks for *MSI2* from the brain dataset displaying aggregated chromatin accessibility signals from raw data and model predictions, along with gene expression distribution at the single-cell level. All track signals are normalized to a uniform data range (0–40). **d** Barplots of the mean Pearson (left) and Spearman (right) correlations between raw and model-predicted gene expression profiles (ArchR, SCARlink, and Cisformer) in three different single-cell multiome datasets (SHARE-seq K562, PBMC, and BCL). **e** Barplots showing cell clustering metrics of predicted gene expression profiles from different models in the PBMC and BCL datasets. **f** UMAP plots comparing cell clustering among raw and model-predicted RNA profiles in the BCL testing dataset. A mixed subset of cells in the UMAP plot for ArchR-predicted RNA values is marked by a black arrow. **g** Violin plots displaying the gene expression distribution of astrocyte-specific gene *BANK1* in raw and model-generated RNA profiles

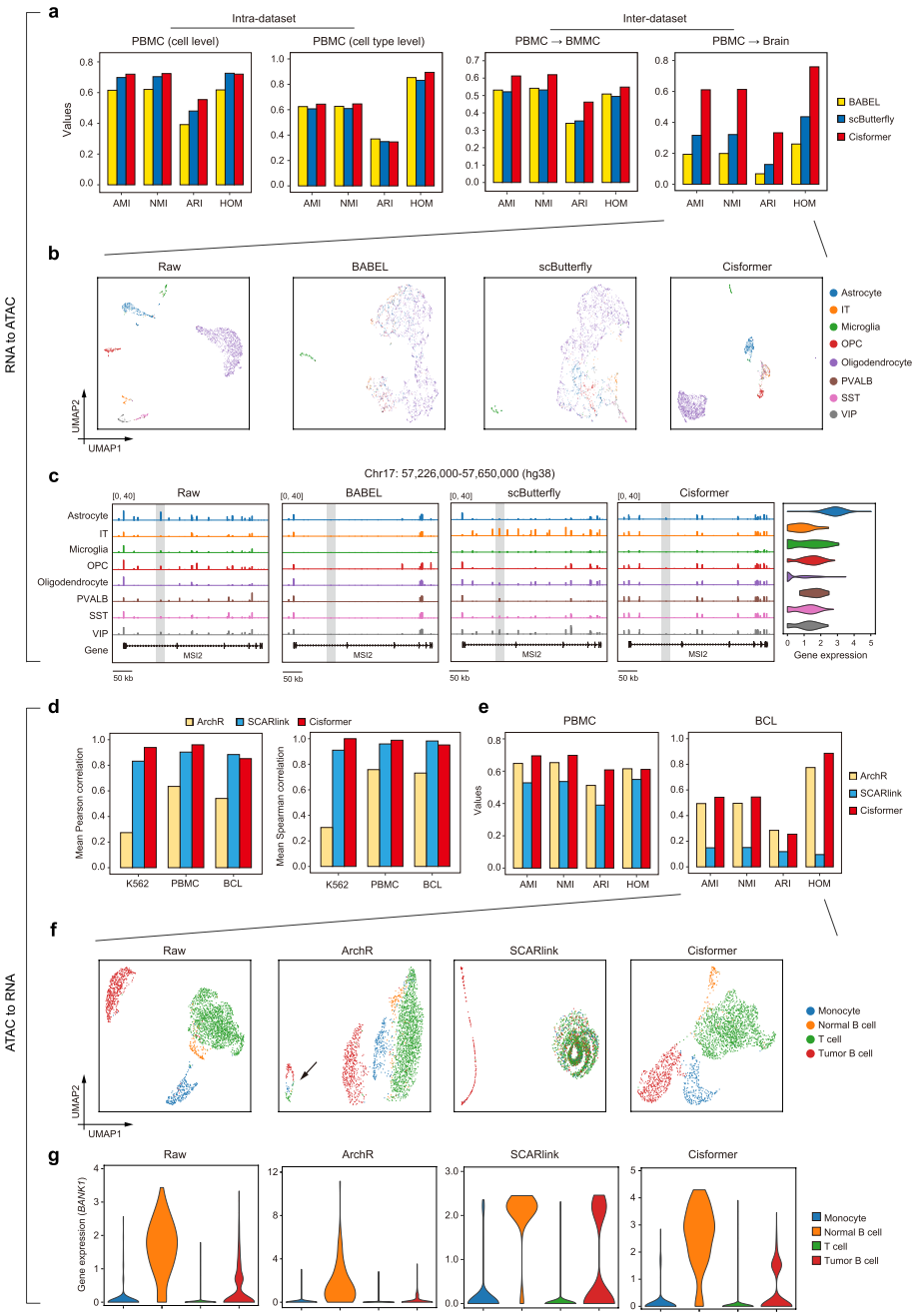

Fig. 2 (See legend on previous page.)

which infer gene expression from chromatin accessibility to establish CRE-gene relationships. Using the PBMC multiome dataset again, along with SHARE-seq data from K562 cells (1413 cells) and B-cell lymphomas (BCL) multiome dataset (14,566 cells), Cisformer achieved superior or comparable performance in gene expression prediction, as assessed by mean gene-wise Pearson and Spearman correlation metrics between predicted and observed values (Fig. 2d). We also clustered cells using the inferred gene expression profiles, observing that Cisformer consistently outperformed ArchR and SCARlink in both the PBMC and BCL datasets (Fig. 2e). For the BCL dataset, Cisformer-predicted RNA

values enabled clear separation of four cell types (monocytes, normal B cells, T cells, and tumor B cells), while SCARlink struggled, possibly due to its gene filtering step. ArchR showed overall correct clustering but exhibited some mixing of a small subset of cells within a single group (Fig. 2f). A similar observation was noted in the PBMC dataset (Additional file 1: Fig. S2d). *BANK1*, a potential tumor suppressor gene predominantly expressed in B cells [14], exhibited the highest concordance between the Cisformer-predicted and measured RNA profiles compared to those generated by ArchR and SCARlink (Fig. 2g). This trend was also observed for the monocyte-specific gene *CD14* (Additional file 1: Fig. S2e). Taken together, our analyses demonstrate that Cisformer provides accurate inference of gene expression from chromatin accessibility profiles.

Cisformer effectively captures the regulatory effects between transcriptome and epigenome

To investigate whether Cisformer possesses biologically relevant model interpretability, we extracted the attention maps between transcriptomic profiles and chromatin accessibility landscapes. Focusing on RNA-to-ATAC direction, we computed gene-wise attention scores by aggregating the attention values of associated gene-peak pairs using the PBMC dataset. We found that chromatin remodeler (CR) genes exhibited the highest attention scores, followed by transcription factor genes, while the remaining genes showed the lowest scores across all 11 cell types (Fig. 3a, Additional file 1: Fig. S3a). This observation aligns with the genome-wide impact of chromatin remodelers on chromatin accessibility, as well as the more site-specific regulatory roles of TFs. Furthermore, we examined the concordance between attention scores of specific genes from Cisformer and chromatin binding patterns of their corresponding proteins. We collected 12 TF ChIP-seq datasets in $CD4^+$ T cells, and calculated enrichment scores representing the relative overlap between attention score-derived peaks and the true peaks in each subtype (Fig. 3b). Cisformer predicted the chromatin binding sites of BCL6 and BRD4 relatively well. To further validate this observation, we categorized all CREs into two groups based on the presence or absence of BCL6 or BRD4 binding sites. As expected, the attention scores of CREs with BCL6 or BRD4 binding were significantly higher than

(See figure on next page.)

Fig. 3 Cisformer captures the interaction between gene expression and chromatin accessibility. **a** Boxplots showing normalized attention scores of CR genes, TF genes, and other genes in five cell types from the PBMC dataset. *P* values are calculated by one-sided *t*-test. CR, chromatin remodeler. **b** Heatmap displaying the enrichment scores of TF binding sites derived from the intersection of peaks from $CD4^+$ T cell ChIP-seq data and Cisformer-inferred attention scores in four $CD4^+$ T cell subtypes. **c** Boxplots comparing normalized attention scores between chromatin regions with and without BCL6 binding sites in $CD4^+$ T cell subtypes. *P* values are calculated using one-sided t-test. **d** Boxplots showing comparison of Pearson correlation between gene expression values and peak signals at the cell-type level for top (100, 200, 500, and 1000) Cisformer-predicted CRE-gene pairs versus randomly selected pairs in the BCL dataset. P values are computed by one-sided t-test. **e** Barplots of enrichment scores for causal GWAS variants in top CREs with different numbers from ArchR, SCARlink, and Cisformer models. **f** Heatmap showing the enrichment of GWAS variants associated with various immune-related traits in top 100 CREs predicted by Cisformer in T cells, monocytes, normal B cells, and tumor B cells. **g** Heatmap displaying normalized attention scores for cell-type differentially expressed genes in the BCL dataset. **h** Genomic tracks of aggregated ATAC profiles and CRE attention scores in four cell types at the *CD19* gene locus, with expression level distributions shown to the right. The highlighted region indicates a tumor B cell-specific enhancer linked to *CD19*. **i** Boxplots showing expression levels of genes linked to varying number of CREs in T cells, monocytes, normal B cells, and tumor B cells

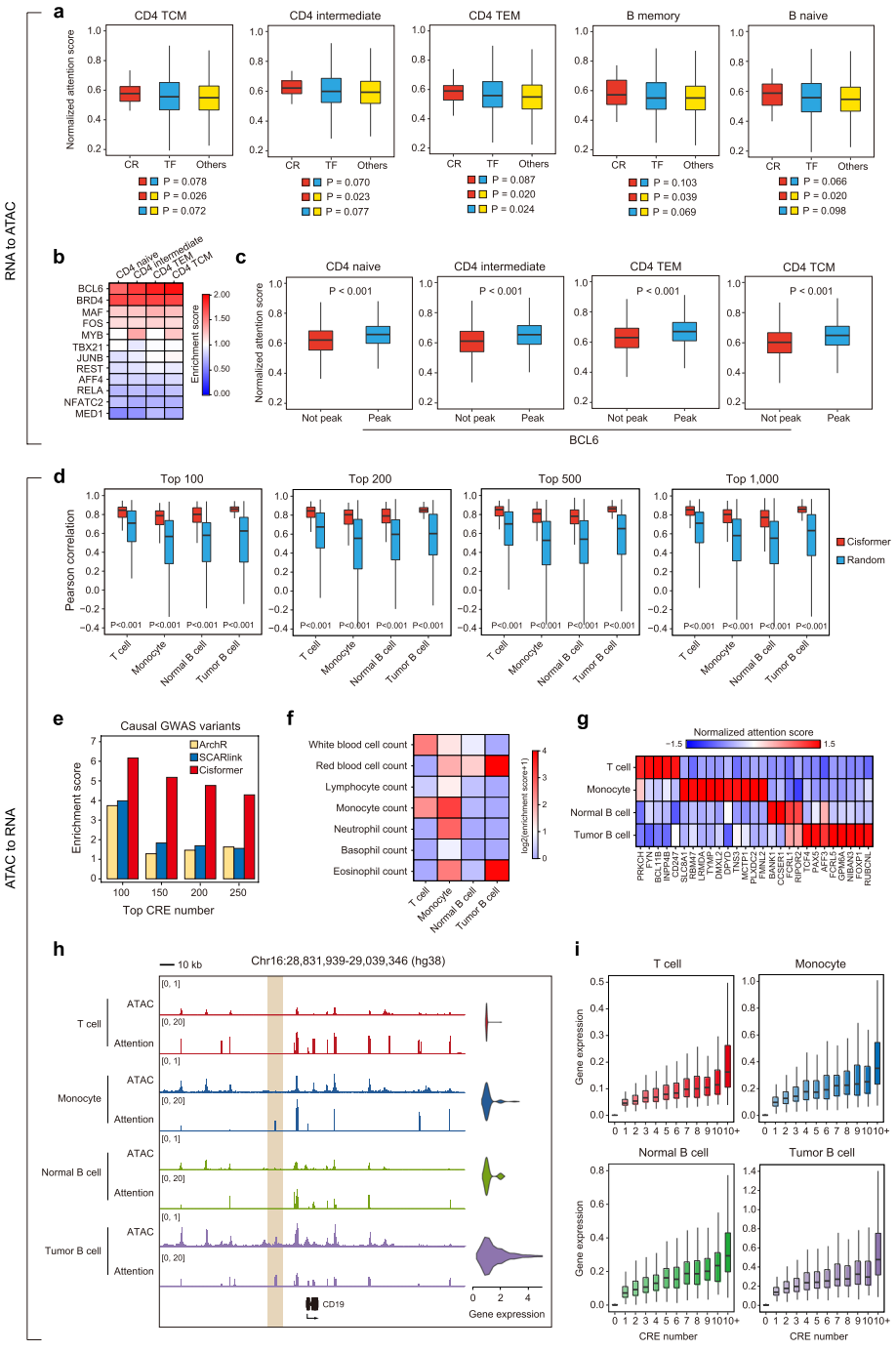

Fig. 3 (See legend on previous page.)

those without binding sites (Fig. 3c, Additional file 1: Fig. S3b). These two factors play important roles in the differentiation and function of $CD4^+$ T cells [15, 16]. In B cells and monocytes, we also validated Cisformer's ability to predict cell-type-specific TF binding sites based on the attention scores (Additional file 1: Fig. S3c–f). Notably, not all TF binding profiles could be accurately inferred by Cisformer, possibly due to the absence of corresponding ground truth (TF ChIP-seq data in relevant cell types). These

findings revealed that Cisformer effectively captures the relationship between genes and CREs when translating transcriptome to chromatin accessibility at the single-cell level.

Next, we evaluated the plausibility of the regulatory patterns learned by Cisformer in modeling the directional relationship from the epigenome to transcriptome. Firstly, we compared the correlation between peak intensity and gene expression of the top high-confidence CRE-gene pairs ranked by Cisformer's attention score against randomly selected pairs of the same size. In the BCL dataset, the high-confidence CRE-gene pairs exhibited significantly stronger peak-gene expression correlations than the random background across all tested thresholds (Fig. 3d). This trend was also consistently observed in the PBMC dataset (Additional file 1: Fig. S3g). Using the BCL multiome dataset again, we prioritized CREs based on their attention scores toward nearby genes located within 250 kb. Compared to ArchR and SCARlink, top enhancers ranked by Cisformer at different cutoffs were consistently more enriched with causal GWAS (genome-wide association study) variants, indicating that Cisformer can reliably identify CRE-gene associations (Fig. 3e). Furthermore, we calculated the enrichment scores for the top 100 CREs concerning immune cell-related variants in each cell type (Fig. 3f). As expected, the top-ranked CREs in monocytes were the most enriched with monocyte count-associated GWAS variants. Interestingly, genetic variants related to red blood cell and eosinophil count were preferentially enriched in the active CREs of tumor B cells, suggesting that tumor cells may hijack the regulatory programs of these two cell types to promote tumorigenesis. At the gene level, our analysis revealed that the aggregated attention scores of cell type-specific genes were highly specific to their respective cell types (Fig. 3g, Additional file 1: Fig. S3h). For example, *CD19* exhibited relatively elevated expression in tumor B cells compared to the normal B cells. Consistently, an upstream enhancer element showed a stronger linkage to *CD19*, as indicated by the higher attention score in tumor B cells versus normal B cells, suggesting its cell-type-specific regulatory role (Fig. 3h). To further investigate Cisformer-inferred CRE-gene associations, we stratified genes into groups based on the number of linked CREs (ranging from 0 to 10 or more). We observed that the gene expression levels increased with CRE count, and this pattern was conserved across all four cell types (Fig. 3i). Collectively, these results highlight that Cisformer is capable of capturing the cell-type-specific regulatory patterns of functional CREs in gene expression.

Cisformer uncovers transcriptional regulatory heterogeneity in the tumor microenvironment

To explore the scalability of Cisformer on large-scale single-cell multiome data, we applied it to a pan-cancer dataset encompassing over 1 million cells [17]. A total of 144,409 cells with paired RNA-ATAC information, representing five major non-tumor cell types (macrophages, T cells, fibroblasts, B cells, and endothelial cells), were extracted from the original dataset (Fig. 4a, Additional file 1: Fig. S4a). Using OV (ovarian cancer) samples as the test set and samples from other cancer types as the training set, we first evaluated the performance of Cisformer in RNA-to-ATAC generation. Cisformer outperformed BABEL and scButterfly across cell clustering metrics (AMI, NMI, ARI, and HOM), although the magnitude of improvement was modest, likely owing to the limited diversity of cell types (Additional file 1: Fig. S4b). Moreover, Cisformer demonstrated the

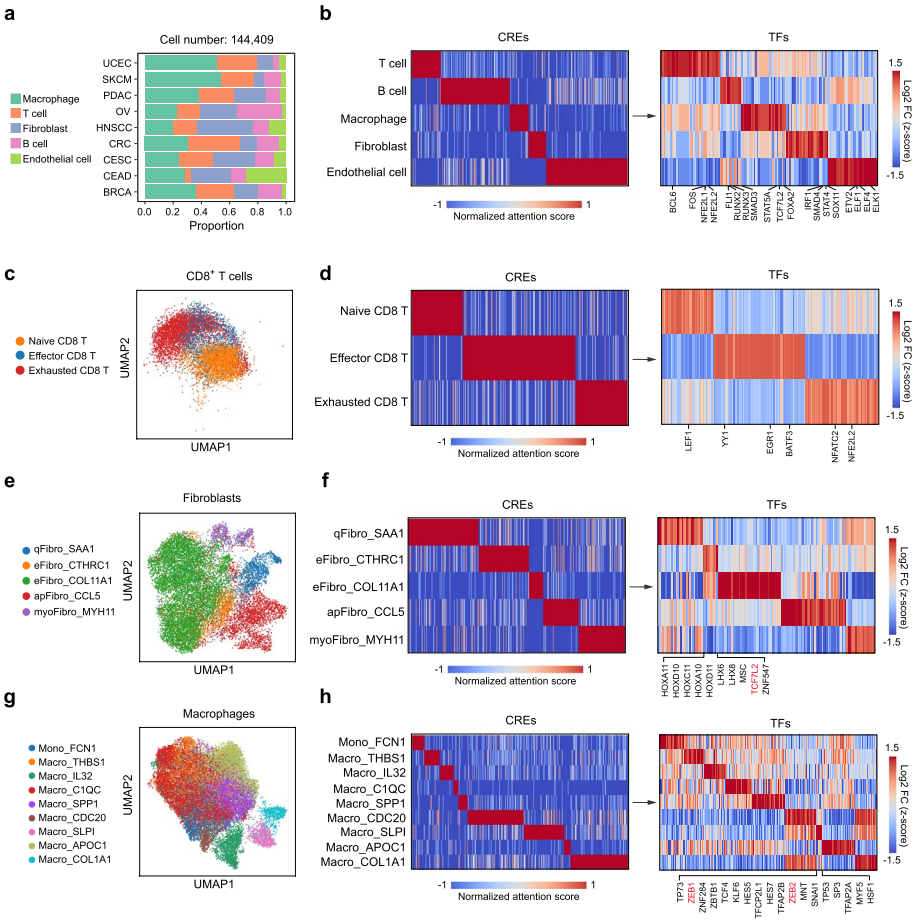

Fig. 4 Cisformer uncovers cell-type-specific regulatory programs in the tumor microenvironment. **a** Barplots showing the relative cell proportion in nine different cancer types. Total cell number is listed at the top. UCEC, uterine corpus endometrial carcinoma; SKCM, skin cutaneous melanoma; PDAC, pancreatic ductal adenocarcinoma; OV, ovarian cancer; HNSCC, head and neck squamous cell carcinoma; CRC, colorectal cancer; CESC, cervical squamous cell carcinoma; CEAD, cervical adenocarcinoma; BRCA, breast cancer. **b** Heatmaps illustrating the normalized attention score of CREs (left) and motif enrichment of TFs (right) in five major cell types (T cells, B cells, macrophages, fibroblasts, and endothelial cells). TFs associated with the regulation of respective cell functions are marked below the heatmap. **c** UMAP visualization of $CD8^+$ T cell subtypes ($CD8^+$ naïve T cell, $CD8^+$ effector T cell, and $CD8^+$ exhausted T cell). **d** Heatmaps showing the CRE normalized attention score (left) and TF motif enrichment (right) in $CD8^+$ T cell subtypes. TFs regulating each cell subtype's function are labeled under the heatmap. **e** UMAP visualization of five fibroblast subtypes. **f** Heatmaps displaying cell-type-specific CREs (left) and TFs (right) in fibroblast subtypes. Inferred TFs that regulate cellular function of $CTHRC1^+$ ECM-remodeling fibroblasts (eFibro_CTHRC1) are listed below the heatmap. **g** UMAP visualization of nine macrophage subtypes. h Heatmaps showing cell-type-specific CREs (left) and TFs (right) in macrophage subtypes. Predicted TFs involved in the regulation of $SLPI^+$ macrophage (Macro_SLPI) are listed below the heatmap

ability to mitigate batch effects (Additional file 1: Fig. S4c). We performed an additional evaluation using metrics of precision, recall, F1 score, and Pearson correlation, and demonstrated that Cisformer achieved much better alignment with ground truth at both peak identification (Additional file 1: Fig. S4d) and signal correlations (Additional file 1: Fig. S4e) for each cell type compared to BABEL and scButterfly. These results indicate that Cisformer enables accurate generation of chromatin accessibility using gene expression profiles at single-cell resolution across large heterogeneous datasets.

Epigenetic dysregulation involving TFs and CREs in both malignant cells and associated non-tumor cells plays crucial roles in tumorigenesis and cancer progression [18–20]. Next, we focused on Cisformer's capability to link CREs with their target genes via ATAC-to-RNA generation. Predicted RNA values from chromatin accessibility by Cisformer exhibited a strong correlation with the measured gene expression profiles, indicating its robust performance in this highly diverse dataset (Additional file 1: Fig. S4f). We identified CREs for each of the five major cell types based on their attention scores toward target genes, followed by TF motif enrichment analysis on these cell-type-specific CREs to uncover regulatory factors associated with cell lineages (Fig. 4b). We observed differential enrichment of FLI1, RUNX2, and RUNX3 binding motifs in B cell-specific CREs. FLI1, a member of the ETS transcription factor family, modulates B cell development and impacts the immune response [21]. RUNX2 regulates the proliferation and survival of B cells, with its dysregulation implicated in B-cell lymphomas and other hematological malignancies [22]. RUNX3, another member of the RUNX family, is essential for the proliferation of human B cells and may function as a tumor suppressor [23]. Within the tumor microenvironment, we identified 33 transcription factor candidates with motifs enriched in the endothelial cell-specific CREs (Fig. 4b, right). Among these factors, four ETS family members (ETV2, ELF1, ELF4, and ELK1) have well-established roles in endothelial biology [24–27]. Similarly, multiple TFs previously implicated in T cells, macrophages, and fibroblasts were recapitulated in our analysis (Fig. 4b, right). These results validate Cisformer's capacity to reveal transcriptional regulation specific to major cell types in the tumor microenvironment.

We then employed Cisformer to uncover differential regulatory programs at the cell subtype level. As a proof of concept, we investigated the regulators of T cell exhaustion, a process characterized by widespread epigenetic remodeling [28]. $CD8^+$ T cells were classified into three subtypes, including naïve, effector, and exhausted cells, based on the expression of relevant marker genes (Fig. 4c, Additional file 1: Fig. S4g). Pseudotime analysis by partition-based graph abstraction (PAGA) [29] further validated the cell annotations (Additional file 1: Fig. S4h). Using the paired RNA-ATAC profiles of these $CD8^+$ T cells, we applied the trained Cisformer model to compute attention scores for each CRE and identified distinct subsets for each of the three cell subtypes (Fig. 4d, left). Motif enrichment analysis subsequently implicated 38, 66, and 53 TFs as putative regulators in naïve, effector, and exhausted $CD8^+$ T cells, respectively (Fig. 4d, right). LEF1, a member of the high-mobility group (HMG) family proteins, maintains $CD8^+$ T cells in a naïve state by repressing cytotoxic gene expression [30]. In effector $CD8^+$ T cells, YY1 drives fate commitment [31], EGR1 controls cell expansion following acute lymphocytic choriomeningitis virus (LCMV) infection [32], and BATF3 deficiency impairs cell differentiation [33]. For exhausted $CD8^+$ T cells, NFATC2 [34] and NFE2L2 [35] have been identified as functional regulators of the exhaustion program. These results demonstrate that Cisformer is effective even under challenging conditions with highly similar cell populations.

We further applied Cisformer to investigate cell subtype-specific transcriptional regulatory programs in fibroblasts and macrophages within the tumor microenvironment. Fibroblasts were clustered and annotated into five distinct subtypes: $SAA1^+$ quiescent fibroblasts (qFibro_SAA1), $CTHRC1^+$ ECM-remodeling fibroblasts (eFibro_CTHRC1),

COL11A1$^+$ ECM-remodeling fibroblasts (eFibro_COL11A1), CCL5$^+$ antigen-presenting associated fibroblasts (apFibro_CCL5), and MYH11$^+$ myofibroblasts (myoFibro_MYH11) (Fig. 4e, Additional file 1: Fig. S4i). We then identified differentially enriched TF binding motifs in subtype-specific CREs (Fig. 4f). Our analysis revealed 10 candidate TFs potentially regulating CTHRC1$^+$ fibroblasts, a population implicated in immune suppression and tumor progression [36]. Among these factors, TCF7L2, a member of the T-cell factor/lymphoid enhancer factor (TCF/LEF) family, is highly expressed in fibroblasts and is considered a key regulator of fibroblast-to-myofibroblast differentiation [37], suggesting potential transition states for CTHRC1$^+$ fibroblasts. Interestingly, binding motifs of several HOX family TFs were enriched in CTHRC1$^+$ fibroblast-specific CREs. Although HOX proteins are well recognized for their crucial roles during embryogenesis, features of the embryonic HOX code appear to persist in adult fibroblasts [38]. Similarly, we clustered macrophages into nine subtypes (Fig. 4g, Additional file 1: Fig. S4j). Given that SLPI is upregulated in various cancers and contributes to metastasis formation, we focused on the SLPI$^+$ macrophages subset (Macro_SLPI) [39] and uncovered 18 putative regulatory TFs (Fig. 4h). ZEB1, a critical mediator of tumor-promoting activity in tumor-associated macrophages (TAMs) [40], and its homolog ZEB2, a master regulator of the TAM program [41], were among the candidate factors. Taken together, these results demonstrate that Cisformer effectively delineates transcriptional regulatory heterogeneity across both major cell types and subpopulations within the tumor microenvironment.

Cisformer enables characterization of aging-associated TFs in the mouse kidney

Finally, we applied Cisformer to a mouse kidney scRNA-seq dataset from the Tabula Muris Senis [42], encompassing 21,647 cells with annotated age information. To generate corresponding chromatin accessibility profiles at the single-cell level, we trained Cisformer using a mouse kidney multiome dataset from 10X Genomics. Using the well-trained model, we generated scATAC-seq profiles for nine cell types, each comprising over 500 cells from the aging kidney dataset. Again, Cisformer outperformed BABEL and scButterfly based on cell clustering metrics (Fig. 5a). We classified these cells into four major clusters: immune cells (T cells, macrophages, and B cells), fenestrated cells, kidney proximal epithelial cells (epithelial cells of proximal tubule and kidney proximal convoluted tubule epithelial cells), and kidney distal epithelial cells (kidney distal convoluted tubule epithelial cells, kidney loop of Henle thick ascending limb epithelial cells, and kidney collecting duct principal cells). The ATAC profiles predicted by Cisformer preserved similar cell heterogeneity to the raw RNA profiles (Fig. 5b).

To further validate the intrinsic consistency between the measured transcriptome and the generated chromatin accessibility profiles, we compared the inferred ages derived from both datasets. The expression level of *Cdkn1a*, a canonical marker of aging, showed a gradual increase with advancing age in kidney epithelial cells (Fig. 5c, left). Here, we grouped kidney proximal and distal epithelial cells into a single group due to their shared epithelial characteristics. The 30-month-old mice did not exhibit an upward trend, likely due to the slower senescence rate in long-living animals [42]. For the generated ATAC profiles, we used EpiTrace [43] to infer single-cell age. The EpiTrace-estimated age also increased progressively with individual age, except for the 30-month-old group in kidney

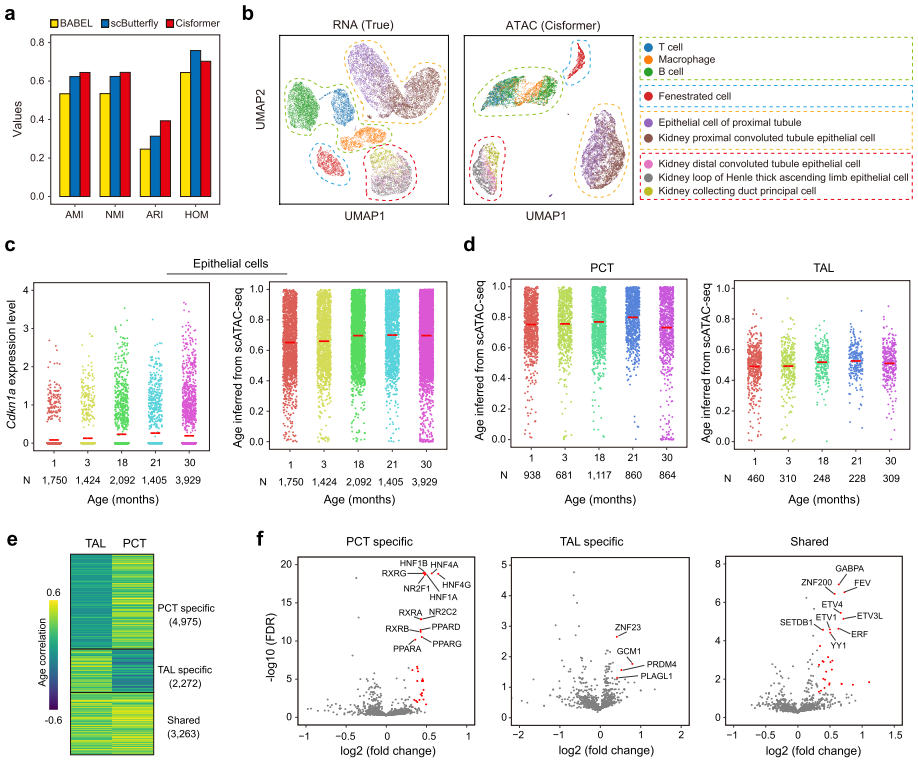

Fig. 5 Cisformer identifies aging-associated TFs in the mouse kidney. **a** Barplots comparing cell clustering metrics of ATAC profiles predicted by BABEL, scButterfly, and Cisformer. **b** UMAP visualization of mouse kidney cells with raw gene expression values (left) and Cisformer-predicted chromatin accessibility profiles (right). **c** Dotplots showing *Cdkn1a* expression level (left) and inferred cell age (right) from Cisformer-predicted chromatin accessibility profiles in mouse epithelial cells across five age groups (1 month, 3 months, 18 months, 21 months, and 30 months). Red bars denote the mean values. Cell numbers are listed below the plots. **d** Dotplots comparing inferred cell age from Cisformer-predicted ATAC profiles in kidney proximal convoluted tubule epithelial cells (PCTs) (left) and kidney loop of Henle thick ascending limb epithelial cells (TAL) (right). **e** Heatmaps showing the correlation between chromatin accessibility and cell age for PCT-specific, TAL-specific, and shared aging-associated CREs. **f** Volcano plots showing the enrichment of TF binding motifs in PCT-specific (left), TAL-specific (middle), and shared (right) CREs. Highly enriched TFs are labeled in the plots

epithelial cells, which was consistent with the RNA-inferred age (Fig. 5c, right). Immune cells and fenestrated cells also showed a general consistency in single-cell ages between gene expression and chromatin accessibility, with ATAC-inferred ages being relatively stable across different groups (Additional file 1: Fig. S5a, b). Focusing on each cell subtype, we found that ages inferred from ATAC profiles were consistent with those measured by *Cdkn1a* expression only in kidney proximal convoluted tubule epithelial cells (PCT) and kidney loop of Henle thick ascending limb epithelial cells (TAL), suggesting that the epigenome of these cell types alters with aging (Fig. 5d, Additional file 1: Fig. S5c).

To investigate the epigenetic mechanisms underlying the aging process, we evaluated the association between CRE accessibility and cell age in PCT and TAL. Among CREs with a relatively high correlation with cell age, 4975 were PCT-specific, 2272 were specific to TAL, and 3263 were shared between both cell types (Fig. 5e). We then performed motif enrichment analysis on these different sets of CREs to reveal putative TFs. DNA

recognition motifs of the hepatocyte nuclear factor (HNF) (HNF1A, HNF1B, HNF4A, and HNF4G), retinoid X receptor (RXR) (RXRA, RXRB, and RXRG), and peroxisome proliferator-activated receptor (PPAR) (PPARA, PPARD, and PPARG) family of TFs were enriched in the PCT-specific CREs (Fig. 5f, left). Among these factors, PPARG can bind to the promoter of *CDKN2A* and induce its expression in human fibroblasts, accelerating cellular senescence [44]. In the TAL-specific CREs, DNA binding motifs of ZNF23, GCM1, PRDM4, and PLAGL1 were enriched. PRDM4 and PLAGL1 have been reported to induce cell cycle arrest in the G1 phase, suggesting a role in promoting aging [45, 46]. The shared CREs were preferentially enriched with ETS family motifs, including GABPA, FEV, ETV1, ETV4, ETV3L, and ERF (Fig. 5f, right). Emerging evidence highlights ETS TFs as conserved regulators of lifespan in organisms ranging from Drosophila to humans [47, 48]. Notably, the aging-related regulatory functions of these TFs in the mouse kidney system or its relevant cell types require further experimental validation. Collectively, these results demonstrate that Cisformer can not only accurately translate transcriptomic profiles into chromatin accessibility landscapes but also facilitate the identification of aging-associated regulators in a cell-type-specific manner.

Discussion

In this study, we develop Cisformer, a Transformer-based model specifically tailored for single-cell RNA-ATAC translation. Cisformer accurately predicts single-cell chromatin accessibility profiles from gene expression in both intra- and inter-dataset contexts, particularly in cross-tissue generations, and surpasses current state-of-the-art methods. For inferring transcriptomic profiles from chromatin accessibility data, Cisformer establishes precise links between CREs and their target genes by leveraging cross-attention mechanisms in a cell-type-specific manner. It facilitates the identification of functional CREs and TFs, providing valuable insights into complex biological processes such as cancer and aging. The generative and interpretable capabilities of Cisformer enhance our understanding of the interplay between the transcriptome and epigenome, paving the way for dissecting the complex regulatory mechanisms underlying cell-type-specific gene expression in both physiological and pathological contexts. The deepened comprehension of epigenetic regulation during tumor development and progression offers significant translational potential for advancing precision oncology and optimizing clinical outcomes [49, 50].

In principle, the model architecture of Cisformer is inherently flexible and can be adapted to model relationships between any pair of omics modalities, including combinations like gene expression with DNA methylation or gene expression with TF binding data. For instance, in multiomics datasets that simultaneously profile gene expression and TF binding, Cisformer could process TF binding sites analogously to chromatin accessibility peaks with minimal architectural adjustments. The cross-attention mechanism naturally captures long-range dependencies between distal TF-binding events and their target genes, similar to its current application in RNA + ATAC data. However, two potential challenges require consideration. First, the typically sparser coverage of experimentally measured TF binding sites compared to ATAC peaks may reduce prediction accuracy at the single-cell level. To mitigate this, we propose integrating DNA sequence tokenization or motif information to guide the model's attention toward genomic

regions with high binding potential. Another challenge involves the divergent regulatory roles of different TFs, where activating and repressive TFs may require specialized modeling approaches. We envision that strategies such as TF-type-specific attention mechanisms or the incorporation of prior knowledge about TF functions could help address this complexity.

It is worth noting that imperfect alignment between chromatin accessibility and gene expression states presents a fundamental challenge for cross-modality prediction. Indeed, existing methods—including our method, Cisformer, as well as BABEL and scButterfly—typically assume a consistent relationship between ATAC and RNA states. However, biological asynchrony can arise. For instance, chromatin accessibility often precedes transcriptional activation during lineage commitment [3, 51]. To address these challenges, we propose a potential solution: (1) obtaining time-resolved multiome data that captures synchronized chromatin and transcriptional states; (2) performing pseudotime analysis on each modality individually, then identifying cell subsets with better-aligned RNA-ATAC states through flow matching between the two pseudotime trajectories; (3) explicitly incorporating state asynchrony into the model framework to better reflect biological reality. A more comprehensive evaluation of RNA-ATAC asymmetry could be feasible with the availability of large-scale paired scMultiome data [17, 52], which would enable the systematic identification of biological systems and cell types exhibiting potential epigenetic priming effects. In these systems, early chromatin signals could serve as predictors of impending transcriptional shifts.

Several potential improvements could further enhance Cisformer. First, training Cisformer with a larger set of model parameters could be beneficial, particularly with the growing availability of single-cell multiome datasets, which would likely improve prediction accuracy and generalization. Second, incorporating DNA sequence features into the model presents a promising avenue for future work, as much important transcriptional regulatory information, such as TF binding motifs, is encoded within the sequence. Third, with the anticipated advancement of spatial multiome datasets, exploring Cisformer's application to spatially resolved datasets may provide valuable insights into the spatial organization of gene regulation. Lastly, the computational efficiency of Cisformer could be further optimized to allow for faster and more scalable analyses. Even in its current form, Cisformer already stands as a powerful method for single-cell cross-modality generation, and holds great potential for providing deeper insights into transcriptional regulatory mechanisms across various biological contexts.

Conclusions

Cisformer represents a cross-attention-based generative framework tailored for cross-modality generation between gene expression and chromatin accessibility at single-cell resolution. Through its superior performance and model interpretability, Cisformer effectively captures the intricate interactions between regulatory landscapes and transcriptional outputs, and empowers the identification of potential functional CREs and key TFs in critical biological processes such as tumorigenesis and organismal aging. Cisformer emerges as a powerful tool for transcriptional regulation analysis, advancing our systems-level understanding of the molecular mechanism underlying both physiological and pathological contexts.

Methods

Data preprocessing

Vocabulary construction and mapping

To enable cross-dataset usages, Cisformer adopts a fixed gene or peak vocabulary. All input paired RNA-ATAC data are first mapped onto this unified vocabulary. Specifically, we constructed the vocabulary by retrieving human genes from the Ensembl genome database (https://www.ensembl.org) and CREs from the ENCODE project (https://screen.encodeproject.org). After filtering, we retained 38,244 genes and 1,033,239 CREs located on human autosomes to form the final vocabulary. Of note, we use genes or CREs directly from the mouse kidney multiome dataset as gene or peak vocabulary for mouse-related cross-dataset usages. In the following sections, we use human gene or peak vocabulary for illustration.

Gene and peak mapping and filtering

For genes from the RNA modality, we retain the original expression values present in both the dataset and the predefined vocabulary. Genes present in the dataset but absent from the vocabulary are discarded. Conversely, genes included in the vocabulary but not detected in the dataset are assigned a value of 0. This allows each cell to be represented by a fixed-length gene expression vector of 38,244.

ATAC profiles are binarized at the vocabulary level. We assess the overlap between its called peaks (convert to Hg38 coordinates using liftOver [53] if necessary) and the predefined CREs for each cell. If a given CRE overlaps with at least 1 bp of any peak in the cell, it is marked as "active" (value of 1); otherwise, it is designated as "inactive" (value of 0). This generates a binary chromatin accessibility vector of length 1,033,239 for each cell.

Following feature mapping, we apply quality control filters to genes, peaks, and cells. For each dataset, genes or peaks detected in fewer than 10 cells are removed. Cells expressing fewer than 200 or more than 20,000 genes, or with fewer than 500 or more than 50,000 active peaks, are also excluded. These filtering steps are set to enhance the model performance.

Feature duplication and selection

The sparsity of single-cell chromatin accessibility data introduces bias during model training, encouraging to minimize loss by predicting uniformly low or zero across all loci. Furthermore, using the full set of genes and CREs as input or output for each cell would result in substantial memory consumption and computational inefficiencies. To address these challenges, we propose a novel feature duplication and selection strategy.

RNA to ATAC: For RNA modality, we randomly sample 2048 genes with non-zero expression values for each cell to construct the input RNA sequence. For ATAC modality, we randomly select 1024 active CREs and 1024 inactive CREs to form the target ATAC sequence. To increase the coverage of the vocabulary in training samples, this sampling step is repeated multiple times depending on the sparsity of data.

ATAC to RNA: For the ATAC modality, we randomly select 10,000 active CREs per cell as the input ATAC sequence. To construct the predicted RNA sequence, 3000

genes with non-zero expression values are randomly selected. In contrast to the RNA-to-ATAC direction, repetition of the sampling step is unnecessary, since the number of selected genes or CREs is nearly equivalent to the total number of expressed genes or active CREs.

Cisformer model

Peak index encoding

The CRE vocabulary is exceptionally large, comprising 1,033,239 unique elements. Traditional embedding approaches would require storing a large trainable matrix of size (vocabulary size × embedding dimension), which is computationally expensive and memory-intensive. To mitigate this issue while preserving the uniqueness and independence of each token, we present a biologically informed and computationally efficient embedding method termed peak index encoding. This strategy proceeds as follows:

Index extraction and padding: for each CRE, we first determine its index within the CRE vocabulary. Given that the vocabulary contains 1,033,239 entries, the maximum index length is 7 digits. Each index is padded with leading zeros to ensure a fixed 7-digit representation. For example, index 32,488 is converted to 0032488.

Digit decomposition: the padded index is decomposed into a sequence of 7 digits, ordered from the most significant (millions place) to the least significant (units place). Using the example above, the index 0032488 is transformed into the sequence [0, 0, 3, 2, 4, 8, 8].

Digit embedding and reconstruction: each digit (ranging from 0 to 9) is embedded using a shared learnable embedding matrix of shape (10, embedding_dim/7), where embedding_dim denotes the model's embedding dimension. This yields a tensor of shape (7, embedding_dim/7) for each index sequence. The resulting tensor is then flattened to generate a final embedding vector of size embedding_dim.

Positional encoding

Given the relatively small number of unique gene tokens and their associated expression values compared to CREs, Cisformer employs a standard token embedding strategy for gene inputs. Each gene is represented by its fixed position in the gene vocabulary. It converts gene identifiers from textual form into integer indices, which are subsequently used to retrieve embeddings via the standard embedding layer. This strategy preserves the uniqueness of each gene while enabling efficient lookup operations.

Let $G = [g_1, g_2, \ldots, g_n]$ be the gene token sequence, and let $E_{gene} \in \mathbb{R}^{V \times d}$ be the gene embedding matrix, where V is the vocabulary size and d is the embedding dimension. The gene embedding is computed as:

$$H_{gene} = E_{gene}[G]$$

where the indexing operation corresponds to a torch.nn.Embedding lookup.

Value embedding

Gene expression values are embedded separately and fused with the gene token embeddings. Strategies are applied depending on the generation direction:

RNA to ATAC: raw gene expression values are capped at 64 to avoid extreme outliers and rounded down to the nearest integer:

$$x_i = \min(\lfloor e_i \rfloor, 64)$$

where e_i is the raw expression value of gene g_i, and $x_i \in \{0,1,\ldots,64\}$. A learnable embedding matrix $E_{val} \in \mathbb{R}^{65\times d}$ is used to encode the expression magnitude:

$$H_{val} = E_{val}[x]$$

ATAC to RNA: expression value of each gene e_i is first log-transformed:

$$x_i' = \log(1 + e_i)$$

Let M be the maximum transformed value across the dataset. We define 7 equal-width bins between 0 and M, with bin edges:

$$B = \{0 = b_0 < b_1 < \cdots < b_7 = M\}, \quad b_k = \frac{k}{7} \cdot M$$

Each transformed value x_i' is assigned to a bin $b \in \{0,\ldots,7\}$, where bin 0 is reserved for values where $e_i = 0$. The corresponding bin index is then embedded using a learnable embedding matrix $E_{bin} \in \mathbb{R}^{8\times d}$:

$$H_{val} = E_{bin}[b]$$

Cross-attention mechanism

At each Transformer layer, cross-attention integrates features across two input modalities (RNA and ATAC) by computing the attention of a *query* sequence from one modality over the *key* and *value* sequences of the other modality. Given a query matrix $\mathbf{Q} \in \mathbb{R}^{\mathbf{L_q}\times\mathbf{d}}$, key matrix $\mathbf{K} \in \mathbb{R}^{\mathbf{L_k}\times\mathbf{d}}$, and value matrix $\mathbf{V} \in \mathbb{R}^{\mathbf{L_k}\times\mathbf{d}}$, the scaled dot-product attention is computed as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d}}\right)V$$

In *multi-head attention*, this operation is performed across h parallel heads, allowing the model to capture information from different representation subspaces:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O$$

$$\text{head}_i = \text{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right)$$

where W^O is the weight matrix.

After computing attention, residual connections and layer normalization are applied:

$$H' = \text{LayerNorm}(H + \text{MultiHead}(Q, K, V))$$

Position-wise feedforward network (FFN)

Each Transformer block comprises a two-layer feedforward neural network applied independently to each token:

$$\mathbf{FFN}(\mathbf{x}) = \mathbf{max}(0, \mathbf{x}\mathbf{W}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2$$

It is followed by a second residual connection and normalization step:

$$\mathrm{H}_{\mathrm{out}} = \mathrm{LayerNorm}\left(\mathrm{H}' + \mathrm{FFN}\left(\mathrm{H}'\right)\right)$$

Memory-efficient attention with FlashAttention2

To further reduce the computational overhead and memory usage during model training, Cisformer adopts FlashAttention2, a state-of-the-art algorithm for memory-efficient attention computation [54]. Unlike the standard attention mechanism that materializes large attention matrices, FlashAttention2 performs the softmax and matrix multiplication in a fused and block-wise streaming manner. It significantly reduces memory usage and improves throughput on modern GPU architectures.

Parameter settings

Cisformer uses the same model architecture for both RNA-to-ATAC and ATAC-to-RNA generation tasks, but adopts task-specific parameter settings and training strategies.

From RNA to ATAC, we use 6 cross-attention layers, a 210 embedding dimension, and 6 attention heads. Each RNA and ATAC sequence is truncated or subsampled to a fixed length of 2048 tokens. RNA sequences are embedded by combining positional encodings and value embeddings, while ATAC sequences are represented via peak index encoding and serve as the query for attention computation. The output sequences are processed through a single-layer MLP (multi-layer perceptron) with a sigmoid activation function to generate binary probabilities for ATAC peaks.

From ATAC to RNA, we adopt 4 cross-attention layers, 280 embedding dimensions, and 7 attention heads. Each ATAC sequence is capped at 10,000 and embedded by peak index encoding.

Each RNA sequence is limited to a maximum of 3000, and represented using positional encoding. The output from Transformer layers is passed through an MLP to produce a prediction matrix of shape 3000×8. Each row of the output matrix represents the predicted bin probabilities corresponding to the expression level of a gene, with expression values categorized into 8 discrete expression bins.

Loss functions

Cisformer employs two types of loss functions depending on the prediction task: binary cross-entropy (BCE) loss for RNA-to-ATAC prediction and categorical cross-entropy (CCE) loss for ATAC-to-RNA prediction.

For predicting the binary status (active or inactive) of ATAC peaks, the BCE loss is used. Given the predicted probability $\widehat{y}_i$ and the true label $y_i \in \{0,1\}$ for each peak i, the BCE loss is defined as:

$$\mathcal{L}_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\widehat{y}_i) + (1 - y_i)\log(1 - \widehat{y}_i)]$$

where N denotes the number of peaks.

For predicting gene expression levels discretized into 8 bins, the CCE loss is used. Given the predicted probability vector $\widehat{y}_i \in \mathbb{R}^8$ for each gene i, and the one-hot encoded ground truth label vector $y_i \in \{0,1\}^8$, the CCE loss is defined as:

$$\mathcal{L}_{CCE} = -\frac{1}{M}\sum_{i=1}^{M}\sum_{c=1}^{8} y_{i,c} \log(\widehat{y}_{i,c})$$

where M is the number of predicted genes, c is index of the expression bin, and $y_{i,c}$ is 1 if gene i belongs to bin c, and 0 otherwise.

Cisformer training

Cisformer is trained using the Adam optimizer [55] for both RNA-to-ATAC and ATAC-to-RNA directions. A StepLR scheduler from PyTorch [56] is employed to decay the learning rate periodically, and gradient clipping with a maximum norm of 1 is applied to prevent gradient explosion.

For the RNA-to-ATAC direction, the initial maximum learning rate is set to 1×10^{-3}, and the learning rate decays by a factor of 0.9 every 5 epochs. We use training batch size of 16, and apply early stopping based on the validation loss. Training is terminated if no improvement is observed for five consecutive epochs. For the ATAC-to-RNA direction, we set the initial maximum learning rate to 5×10^{-4}, and the training batch size to 96. The learning rate decays by a factor of 0.6 every 4 epochs. Training step is terminated if the validation loss does not decrease for two consecutive epochs.

To enhance training efficiency and reduce memory consumption, mixed-precision training (FP16) is adopted, utilizing the default FP16 policy provided by the Hugging Face Accelerate framework (https://github.com/huggingface/accelerate). Model training is conducted in a distributed fashion across at least two NVIDIA A800 GPUs. During training, model checkpoints are saved at the end of every epoch, provided that the early stopping condition is not triggered at that epoch. The saved checkpoints store the model weights, the optimizer state, and the learning rate scheduler state, allowing training to be resumed if interrupted. Among all saved checkpoints, the one corresponding to the lowest validation loss is selected as the final model for downstream evaluation and testing.

Cisformer prediction

In the RNA-to-ATAC generation task, the primary goal is to generate complete scATAC-seq profiles solely based on scRNA-seq input. To maximize predictive coverage, we increase the input RNA sequence length to the max number of expressed genes in a cell from scRNA-seq data and perform ATAC peak prediction across all CREs from scATAC-seq data. Each predicted CRE is assigned a probability between 0 and 1

computed via the sigmoid activation function. Predicted probability more than 0.5 is set to 1, and 0 otherwise.

In the ATAC-to-RNA prediction task, Cisformer is designed to predict a fixed-length subset of nonzero gene expression profiles (default maximum: 3000 genes, adjustable depending on dataset size). In this setting, the emphasis is not on reconstructing complete gene expression profiles but rather on accurately modeling the regulatory links between CREs and genes.

Cisformer evaluation

To assess model performance in translating gene expression into chromatin accessibility, we compare Cisformer against BABEL and scButterfly, which are two state-of-the-art cross-modal prediction methods. All models are trained on the same training dataset and evaluated on the same test set. The predicted peak matrices are analyzed using SnapATAC2 [57], and cell clustering quality is evaluated by computing AMI, NMI, ARI, and HOM. In addition, we use precision, recall, and F1 score to assess the accuracy of peak identification at the cell level, and apply the Pearson correlation coefficient to evaluate the consistency between predicted and experimentally measured signal intensities at the cell-type level.

For ATAC-to-RNA prediction, Cisformer is benchmarked against ArchR and Scarlink, two leading tools capable of inferring regulatory connections from single-cell multiome data. To ensure a fair comparison, predictions from ArchR and Scarlink are post-processed. Genes with a true expression level of 0 are manually assigned a value of 0 in the predicting outcomes. All predicted gene expression profiles are processed using the Scanpy pipeline [58] for cell clustering and the uniform manifold approximation and projection (UMAP) visualization. Cell clustering quality is evaluated by AMI, NMI, ARI, and HOM. To validate CRE-gene pairs predicted by Cisformer, we compute the Pearson correlation coefficient between gene expression and peak intensity at the cell sub-type level using the Scanpy pipeline [58].

Attention matrix generation and normalization

For the ATAC-to-RNA translation task, the cross-attention score between the key matrix K (encoded from CRE inputs) and the query matrix Q (encoded from gene inputs) can be interpreted as the regulatory strength of CREs on genes. Formally, the raw attention A is computed as:

$$\mathrm{A} = \mathrm{softmax}\left(\frac{\mathrm{QK}^{\mathrm{T}}}{\sqrt{\mathrm{d}}}\right)$$

where d denotes the dimensionality of the model embeddings.

Since regulatory elements typically influence nearby genes, attention calculations are restricted to CREs within ± 250 kb of each gene. We first identify these CREs and store them in a mask dictionary. Then we randomly sample 100 cells to build a representative attention matrix for each cell type. For each sampled cell, we compute the average attention across all heads to obtain a single matrix. To ensure comparability across cells, rank normalization is performed both row-wise and column-wise on each cell's attention

matrix, followed by min–max normalization to scale the attention values between 0 and 1. Using the constructed mask dictionary, the attention matrix is aggregated at the cell type level for selected CREs. The attention scores, either gene-wise or CRE-wise, are derived from the attention matrix as needed. In the RNA-to-ATAC direction, the general steps of attention matrix processing follow the same procedure as in the ATAC-to-RNA direction, with minor modifications. In this case, the key matrix K is derived from gene inputs, and the query matrix Q is encoded from CRE inputs. Additionally, no filtering of CREs is performed, and log normalization is applied in place of rank normalization. The gene-wise attention scores are derived from the attention matrix as needed.

Cisformer Inferred peak enrichment

To compare the attention scores of TF genes inferred from Cisformer with corresponding protein chromatin binding profiles, we first collect ChIP-seq datasets for major cell types ($CD4^+$ T cells, B cells, and monocytes) from the ReMap database [59], which serves as an approximate ground truth. The enrichment score is calculated as the ratio of overlap between the true peak sets and attention score-derived or randomly shuffled peak sets for each cell subtype. Attention score-derived peaks correspond to the top 10,000 peaks ranked by gene-wise attention scores.

Causal GWAS variant enrichment

To systematically assess the accuracy of CRE–gene associations inferred by Cisformer, we use a fine-mapped GWAS variant dataset from the UK Biobank, covering 94 traits and 693,744 variants. For each CRE, we calculate the sum of attention scores to all genes located within ± 250 kb. The top-ranked CREs are then considered as potential functional CREs. To quantify enrichment, we compute the mean posterior inclusion probability (PIP) of variants falling on the selected CREs and compare it to the mean PIP of an equal number of variants randomly sampled. The causal GWAS variant enrichment score is defined as follows:

$$\text{Enrichment score} = \frac{\text{Mean PIP of variants on top CREs}}{\text{Mean PIP of variants on random CREs}}$$

For each GWAS trait, we calculate a trait-specific enrichment score, and the average across all traits is taken as the final enrichment score for the dataset. The cell-type-specific enrichment score can be computed using the aggregated attention matrix of each respective cell type.

CRE number determination

To investigate the relationship between gene expression levels and the number of strongly associated CREs, we first determine a significance cutoff for CRE–gene links by taking the 90th percentile of all CRE–gene attention scores within each cell type. For each gene, we count the CREs whose regulatory strength toward the gene exceeded this cutoff. Genes are then categorized into 12 groups based on the CRE number (from 0 to 10, and 10+). For each group, we show the distribution of gene expression levels using the log1p-transformed expression values.

Cell-Type-Specific transcription factor identification

To identify cell-type-specific transcription factors, we first aggregate CRE–gene regulatory strength by summing the attention scores from each CRE to nearby genes within ± 250 kb for each cell type. These summed attention scores of CREs across cell types are assembled into a CRE-by-cell-type matrix. For each CRE, we normalize the values across cell types to the range from − 1 to 1 based on its minimum and maximum values, and divide the range into five equal intervals, assigning a rank from 1 to 5 accordingly. We identify CREs that achieved the highest rank exclusively in a single cell type as cell-type-specific CREs. TF motif enrichment analysis for identification of cell-type-specific regulatory factors is performed using the tl.motif_enrichment function from SnapATAC2 [57].

Aging-related CRE and transcription factor identification

We use Epitrace [43] to estimate cell age from Cisformer-inferred scATAC-seq data, and compute CRE-age association by the AssociationOfPeaksToAge function for two mouse kidney epithelial cell types (TAL and PCT). CREs are clustered into four groups using the K-means clustering algorithm, and three clusters with high correlation in at least one cell type are selected (PCT-specific CREs, TAL-specific CREs, and shared CREs). Motif enrichment analysis is performed using SnapATAC2 [57] to identify cell-type-specific and aging-associated transcription factors.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-025-03823-z.

Additional file 1. Fig S1: Visualization of data dimensionality change upon processing by Cisformer. Fig S2: Performance evaluation of Cisformer in cross-modality translation. Fig S3: Cisformer uncovers transcriptome-epigenome interaction. Fig S4: Cisformer identifies cell-type-specific regulatory factors in the tumor microenvironment. Fig S5: Cisformer enables identification of aging-associated TFs in the mouse kidney.

Acknowledgements
We thank members of the Wang laboratory for insightful comments and suggestions.

Peer review information
Claudia Feng was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Authors' contributions
L.J. and C.W. conceived the project. L.J. and Q.Z. developed the Cisformer algorithm, and performed data analysis and investigation. K.T. conducted cell-type annotation for fibroblasts and macrophages in the pan-cancer datasets. L.J., Q.Z. and C.W. wrote the manuscript, with input from all authors on the text and figures. C.W. supervised the entire project and secured funding. All authors reviewed and approved the final manuscript.

Funding
This work was supported by the National Key R&D Program of China (2022YFA1106000 to C.W.), the National Natural Science Foundation of China grants (32222026 and 32170660 to C.W.; 32300429 to L.J.), the Natural Science Foundation of Shanghai (24ZR1492800 to C.W.), Tongji University Spark-X Program, Shanghai Pilot Program for Basic Research, Tongji University Medicine-X Interdisciplinary Research Initiative, the Fundamental Research Funds for the Central Universities (22120240435), Postdoctoral Fellowship Program of CPSF (GZC20231945 to L.J.), and China Postdoctoral Science Foundation (2024M762404 to L.J.).

Data availability
In this work, all datasets used were obtained from public data repositories. The human PBMC (https://www.10xgenomics.com/datasets/pbmc-from-a-healthy-donor-granulocytes-removed-through-cell-sorting10-k-1-standard-2–0-0), BCL (https://www.10xgenomics.com/datasets/fresh-frozen-lymph-node-with-b-cell-lymphoma-14-k-sorted-nuclei-1-standard-2–0-0), brain (https://www.10xgenomics.com/datasets/frozen-human-healthy-brain-tissue-3-k-1-standard-2–0-0), and mouse kidney (https://www.10xgenomics.com/datasets/mouse-kidney-nuclei-isolated-with-chromium-nuclei-isolation-kit-saltyez-protocol-and-10x-complex-tissue-dp-ct-sorted-and-ct-unsorted-1-standard-2–0-0) multiome datasets were downloaded from 10X Genomics website. The human BMMC multiome datasets can be accessed in NCBI Gene Expression Omnibus (GEO) with the accession number GSE194122 [60, 61]. The SHARE-seq data of K562 cell line

is available at GEO accession number GSE140203 [3, 62]. Human pancancer datasets were collected through the HTAN DCC Portal (https://data.humantumoratlas.org/) under the HTAN WUSTL Atlas [17, 63]. Mouse kidney scRNA-seq dataset was downloaded from CZ CELLxGENE Discover (https://cellxgene.cziscience.com/collections/0b9d8a04-bb9d-44da-aa27-705bb65b54eb) [42, 64]. UK Biobank GWAS data with fine-mapping were downloaded from the Finucane lab (https://www.finucanelab.org/data). The source code and pre-trained models for Cisformer is available at GitHub (https://github.com/wanglabtongji/Cisformer) [65] and Zenodo (https://doi.org/10.5281/zenodo.16991152) [66] under the MIT license. Detailed instructions for installation and usage are included in the repository, along with all the necessary files for training, prediction, and interpretation of the Cisformer model using the provided demo dataset.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 30 May 2025 Accepted: 30 September 2025
Published online: 06 October 2025

References

1. Sun F, Li H, Sun D, Fu S, Gu L, Shao X, et al. Single-cell omics: experimental workflow, data analyses and applications. Sci China Life Sci. 2025;68(1):5–102.
2. Chen S, Lake BB, Zhang K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. Nat Biotechnol. 2019;37(12):1452–7.
3. Ma S, Zhang B, LaFave LM, Earl AS, Chiang Z, Hu Y, et al. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. Cell. 2020;183(4):1103–16.
4. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, et al. Simultaneous epitope and transcriptome measurement in single cells. Nat Methods. 2017;14(9):865–8.
5. Peterson VM, Zhang KX, Kumar N, Wong J, Li L, Wilson DC, et al. Multiplexed quantification of proteins and transcripts in single cells. Nat Biotechnol. 2017;35(10):936–9.
6. Liu Z, Chen Y, Xia Q, Liu M, Xu H, Chi Y, et al. Linking genome structures to functions by simultaneous single-cell Hi-C and RNA-seq. Science. 2023;380(6649):1070–6.
7. Zhou T, Zhang R, Jia D, Doty RT, Munday AD, Gao D, et al. GAGE-seq concurrently profiles multiscale 3D genome organization and gene expression in single cells. Nat Genet. 2024;56(8):1701–11.
8. Wu KE, Yost KE, Chang HY, Zou J. Babel enables cross-modality translation between multiomic profiles at single-cell resolution. Proc Natl Acad Sci U S A. 2021;118(15):e2023070118.
9. Zhang R, Meng-Papaxanthos L, Vert JP, Noble WS. Multimodal single-cell translation and alignment with semi-supervised learning. J Comput Biol. 2022;29(11):1198–212.
10. Cao Y, Zhao X, Tang S, Jiang Q, Li S, Li S, et al. Scbutterfly: a versatile single-cell cross-modality translation method via dual-aligned variational autoencoders. Nat Commun. 2024;15(1):2973.
11. Wittkopp PJ, Kalay G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat Rev Genet. 2012;13(1):59–69.
12. Granja JM, Corces MR, Pierce SE, Bagdatli ST, Choudhry H, Chang HY, et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat Genet. 2021;53(3):403–11.
13. Mitra S, Malik R, Wong W, Rahman A, Hartemink AJ, Pritykin Y, et al. Single-cell multi-ome regression models identify functional and disease-associated enhancers and enable chromatin potential analysis. Nat Genet. 2024;56(4):627–36.
14. Yan J, Nie K, Mathew S, Tam Y, Cheng S, Knowles DM, et al. Inactivation of BANK1 in a novel IGH-associated translocation t (4; 14)(q24; q32) suggests a tumor suppressor role in B-cell lymphoma. Blood Cancer J. 2014;4(5):e215.
15. Kroenke MA, Eto D, Locci M, Cho M, Davidson T, Haddad EK, et al. Bcl6 and Maf cooperate to instruct human follicular helper CD4 T cell differentiation. J Immunol. 2012;188(8):3734–44.
16. Cheung KL, Zhang F, Jaganathan A, Sharma R, Zhang Q, Konuma T, et al. Distinct roles of Brd2 and Brd4 in potentiating the transcriptional program for Th17 cell differentiation. Mol Cell. 2017;65(6):1068–80.
17. Terekhanova NV, Karpova A, Liang WW, Strzalkowski A, Chen S, Li Y, et al. Epigenetic regulation during cancer transitions across 11 tumour types. Nature. 2023;623(7986):432–41.
18. Peng P, Qin S, Li L, He Z, Li B, Nice EC, et al. Epigenetic remodeling under oxidative stress: mechanisms driving tumor metastasis. MedComm – Oncology. 2024;3(4):e70000.
19. Pan X, Na F, Chen X. Deciphering transcriptional bursts and enhancer dynamics: advancing cancer therapeutics through single-cell global run-on sequencing. MedComm – Oncology. 2024;3(3):e88.
20. Wang Z, Xie F, Zhou F. RREB1: a critical transcription factor, integrates TGF-β and RAS signals to drive cancer metastasis via regulation of enhancers. MedComm – Oncology. 2025;4(1):e70016.
21. Zhang XK, Moussa O, LaRue A, Bradshaw S, Molano I, Spyropoulos DD, et al. The transcription factor Fli-1 modulates marginal zone and follicular B cell development in mice. J Immunol. 2008;181(3):1644–54.

22. Zhang PP, Wang YC, Cheng C, Zhang F, Ding DZ, Chen DK. Runt-related transcription factor 2 influences cell adhesion-mediated drug resistance and cell proliferation in B-cell non-Hodgkin's lymphoma and multiple myeloma. Leuk Res. 2020;92:106340.
23. Spender LC, Whiteman HJ, Karstegl CE, Farrell PJ. Transcriptional cross-regulation of RUNX1 by RUNX3 in human B cells. Oncogene. 2005;24(11):1873–81.
24. Huang X, Brown C, Ni W, Maynard E, Rigby AC, Oettgen P. Critical role for the Ets transcription factor ELF-1 in the development of tumor angiogenesis. Blood. 2006;107(8):3153–60.
25. Sivina M, Yamada T, Park CS, Puppi M, Coskun S, Hirschi K, et al. The transcription factor E74-like factor controls quiescence of endothelial cells and their resistance to myeloablative treatments in bone marrow. Arterioscler Thromb Vasc Biol. 2011;31(5):1185–91.
26. Harel S, Sanchez V, Moamer A, Sanchez-Galan JE, Abid Hussein MN, Mayaki D, et al. ETS1, ELK1, and ETV4 transcription factors regulate angiopoietin-1 signaling and the angiogenic response in endothelial cells. Front Physiol. 2021;12:683651.
27. Kim TM, Lee RH, Kim MS, Lewis CA, Park C. ETV2/ER71, the key factor leading the paths to vascular regeneration and angiogenic reprogramming. Stem Cell Res Ther. 2023;14(1):41.
28. Belk JA, Daniel B, Satpathy AT. Epigenetic regulation of T cell exhaustion. Nat Immunol. 2022;23(6):848–60.
29. Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Göttgens B, et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol. 2019;20:1–9.
30. Shan Q, Li X, Chen X, Zeng Z, Zhu S, Gai K, et al. Tcf1 and Lef1 provide constant supervision to mature CD8+ T cell identity and function by organizing genomic architecture. Nat Commun. 2021;12(1):5863.
31. Yu B, Zhang K, Milner JJ, Toma C, Chen R, Scott-Browne JP, et al. Epigenetic landscapes reveal transcription factors that regulate CD8+ T cell differentiation. Nat Immunol. 2017;18(5):573–82.
32. Singh A, Svaren J, Grayson J, Suresh M. CD8 T cell responses to lymphocytic choriomeningitis virus in early growth response gene 1-deficient mice. J Immunol. 2004;173(6):3855–62.
33. Li C, Liu Z, Wang Z, Yim WY, Huang Y, Chen Y. BATF and BATF3 deficiency alters CD8+ effector/exhausted T cells balance in skin transplantation. Mol Med. 2024;30(1):16.
34. Zhu L, Zhou X, Gu M, Kim J, Li Y, Ko CJ, et al. Dapl1 controls NFATc2 activation to regulate CD8+ T cell exhaustion and responses in chronic infection and cancer. Nat Cell Biol. 2022;24(7):1165–76.
35. Dahabieh MS, DeCamp LM, Oswald BM, Kitchen-Goosen SM, Fu Z, Vos M, et al. NRF2-dependent regulation of the prostacyclin receptor PTGIR drives CD8 T cell exhaustion. bioRxiv. 2024.06.23.600279.
36. Li E, Cheung HC, Ma S. CTHRC1+ fibroblasts and SPP1+ macrophages synergistically contribute to pro-tumorigenic tumor microenvironment in pancreatic ductal adenocarcinoma. Sci Rep. 2024;14(1):17412.
37. Contreras O, Soliman H, Theret M, Rossi FM, Brandan E. TGF-β-driven downregulation of the transcription factor TCF7L2 affects Wnt/β-catenin signaling in PDGFRα+ fibroblasts. J Cell Sci. 2020;133(12):jcs242297.
38. Rinn JL, Bondre C, Gladstone HB, Brown PO, Chang HY. Anatomic demarcation by positional variation in fibroblast gene expression programs. PLoS Genet. 2006;2(7):e119.
39. Nugteren S, Goos JA, Delis-van Diemen PM, Simons-Oosterhuis Y, Lindenbergh-Kortleve DJ, van Haaften DH, et al. Expression of the immune modulator secretory leukocyte protease inhibitor (SLPI) in colorectal cancer liver metastases and matched primary tumors is associated with a poorer prognosis. Oncoimmunology. 2020;9(1):1832761.
40. Cortés M, Sanchez-Moral L, de Barrios O, Fernández-Aceñero MJ, Martínez-Campanario MC, Esteve-Codina A, et al. Tumor-associated macrophages (TAMs) depend on ZEB1 for their cancer-promoting roles. EMBO J. 2017;36(22):3336–55.
41. Sheban F, San Phan T, Xie K, Ingelfinger F, Gur C, Itai YS, et al. ZEB2 is a master switch controlling the tumor-associated macrophage program. Cancer Cell. 2025;S1535–6108(25):00122–9.
42. Tabula Muris Consortium. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. Nature. 2020;583(7817):590–5.
43. Xiao Y, Jin W, Ju L, Fu J, Wang G, Yu M, et al. Tracking single-cell evolution using clock-like chromatin accessibility loci. Nat Biotechnol. 2025;43(5):784–98.
44. Gan Q, Huang J, Zhou R, Niu J, Zhu X, Wang J, et al. PPARγ accelerates cellular senescence by inducing p16INK4α expression in human diploid fibroblasts. J Cell Sci. 2008;121(13):2235–45.
45. Yang WT, Chen M, Xu R, Zheng PS. PRDM4 inhibits cell proliferation and tumorigenesis by inactivating the PI3K/AKT signaling pathway through targeting of PTEN in cervical carcinoma. Oncogene. 2021;40(18):3318–30.
46. Spengler D, Villalba M, Hoffmann A, Pantaloni C, Houssami S, Bockaert J, et al. Regulation of apoptosis and cell cycle arrest by Zac1, a novel zinc finger protein expressed in the pituitary gland and the brain. EMBO J. 1997;16(10):2814–25.
47. Dobson AJ, Boulton-McDonald R, Houchou L, Svermova T, Ren Z, Subrini J, et al. Longevity is determined by ETS transcription factors in multiple tissues and diverse species. PLoS Genet. 2019;15(7):e1008212.
48. Tanaka-Yano M, Sugden WW, Wang D, Badalamenti B, Côté P, Chin D, et al. Dynamic activity of Erg promotes aging of the hematopoietic system. bioRxiv. 2025. https://doi.org/10.1101/2025.01.23.634563.
49. Tao L, Zhou Y, Luo Y, Qiu J, Xiao Y, Zou J, et al. Epigenetic regulation in cancer therapy: from mechanisms to clinical advances. MedComm – Oncology. 2024;3(1):e59.
50. Du M, Zhang J, Wicha MS, Luo M. Redox regulation of cancer stem cells: biology and therapeutic implications. MedComm – Oncology. 2024;3(4):e70005.
51. Lara-Astiaso D, Weiner A, Lorenzo-Vivas E, Zaretsky I, Jaitin DA, David E, et al. Chromatin state dynamics during blood formation. Science. 2014;345(6199):943–9.
52. Zuo Z, Cheng X, Ferdous S, Shao J, Li J, Bao Y, et al. Single cell dual-omic atlas of the human developing retina. Nat Commun. 2024;15(1):6792.
53. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, et al. The UCSC genome browser database: update 2006. Nucleic Acids Res. 2006;34:D590–8.
54. Dao T. Flashattention-2: Faster attention with better parallelism and work partitioning. 2023. arXiv preprint arXiv:2307.08691.

55. Kingma DP, Ba J. Adam: a method for stochastic optimization. 2014. arXiv preprint arXiv:1412.6980.
56. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: An imperative style, high-performance deep learning library. 2019. arXiv preprint arXiv:1912.01703. 1912;10.
57. Zhang K, Zemke NR, Armand EJ, Ren B. A fast, scalable and versatile tool for analysis of single-cell omics data. Nat Methods. 2024;21(2):217–27.
58. Wolf FA, Angerer P, Theis FJ. Scanpy: large-scale single-cell gene expression data analysis. Genome Biol. 2018;19(1):15.
59. Hammal F, De Langen P, Bergon A, Lopez F, Ballester B. ReMap 2022: a database of human, mouse, *Drosophila* and *Arabidopsis* regulatory regions from an integrative analysis of DNA-binding sequencing experiments. Nucleic Acids Res. 2022;50(D1):D316-25.
60. Luecken MD, Burkhardt DB, Cannoodt R, Lance C, Agrawal A, Aliee H, et al. A sandbox for prediction and integration of DNA, RNA, and proteins in single cells. In: Vanschoren J, Yeung S, editors. Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks. vol. 1. La Jolla, CA: Neural Information Processing Systems Foundation, Inc; 2021.
61. Luecken MD, Burkhardt DB, Cannoodt R, Lance C, Agrawal A, Aliee H, et al. A sandbox for prediction and integration of DNA, RNA, and proteins in single cells. Datasets. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE194122 (2022).
62. Ma S, Zhang B, LaFave LM, Earl AS, Chiang Z, Hu Y, et al. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. Datasets. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE140203 (2020).
63. Terekhanova NV, Karpova A, Liang WW, Strzalkowski A, Chen S, Li Y, et al. Epigenetic regulation during cancer transitions across 11 tumour types. Datasets. The Human Tumor Atlas Network. https://humantumoratlas.org/explore?selectedFilters=%5B%7B%22value%22%3A%22HTAN+WUSTL%22%2C%22group%22%3A%22AtlasName%22%2C%22count%22%3A7156%2C%22isSelected%22%3Afalse%7D%5D (2023).
64. Tabula Muris Consortium. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. Datasets. CZ CELLxGENE Discover. https://cellxgene.cziscience.com/collections/0b9d8a04-bb9d-44da-aa27-705bb65b54eb (2020).
65. Ji L, Zou Q, Tang K, Wang C. Cisformer: a scalable cross-modality generation framework for decoding transcriptional regulation at single-cell resolution. GitHub. 2025. https://github.com/wanglabtongji/Cisformer.
66. Ji L, Zou Q, Tang K, Wang C. Cisformer: a scalable cross-modality generation framework for decoding transcriptional regulation at single-cell resolution. 2025. Zenodo. https://doi.org/10.5281/zenodo.16991152.

Publisher's Note