Data integration and inference of gene regulation using single-cell temporal multimodal data with scTIE

Yingxin Lin, Tung-Yu Wu, Xi Chen, et al.

Genome Res. published online December 15, 2023
Access the most recent version at doi:10.1101/gr.277960.123

Data integration and inference of gene regulation using single-cell temporal multimodal data with scTIE

Yingxin Lin[1,2,3§], Tung-Yu Wu[4§], Xi Chen[4§], Sheng Wan[5], Brian Chao[6], Jingxue Xin[4], Jean Y.H. Yang[1,2,3], Wing H. Wong[4,7,8*], and Y. X. Rachel Wang[1*]

[1]School of Mathematics and Statistics, The University of Sydney, NSW, Australia.

[2]Charles Perkins Centre, The University of Sydney, NSW, Australia.

[3]Laboratory of Data Discovery for Health Limited (D24H), Science Park, Hong Kong SAR, China

[4]Department of Statistics, Stanford University, CA, USA.

[5]Institute of Electronics, National Yang Ming Chiao Tung University, Hsinchu, Taiwan.

[6]Department of Electrical Engineering, Stanford University, CA, USA.

[7]Department of Biomedical Data Science, Stanford University, CA, USA.

[8]Bio-X Program, Stanford University, CA, USA.

[§]Equal contribution.

[*]To whom correspondence should be addressed. Email: Y. X. Rachel Wang, rachel.wang@sydney.edu.au; Wing H. Wong, whwong@stanford.edu.

Keywords

Single cell multiome, Temporal data integration, Context-specific gene regulatory network

Abstract

Single-cell technologies offer unprecedented opportunities to dissect gene regulatory mechanisms in context-specific ways. Although there are computational methods for extracting gene regulatory relationships from scRNA-seq and scATAC-seq data, the data integration problem, essential for accurate cell type identification, has been mostly treated as a standalone challenge.

Here we present scTIE, a unified method that integrates temporal multimodal data and infers regulatory relationships predictive of cellular state changes. scTIE uses an autoencoder to embed cells from all time points into a common space using iterative optimal transport, followed by extracting interpretable information to predict cell trajectories. Using a variety of synthetic and real temporal multimodal datasets, we demonstrate scTIE achieves effective data integration while preserving more biological signals than existing methods, particularly in the presence of batch effects and noise. Furthermore, on the exemplar multiome dataset we generated from differentiating mouse embryonic stem cells over time, we demonstrate scTIE captures regulatory elements highly predictive of cell transition probabilities, providing new potentials to understand the regulatory landscape driving developmental processes.

Introduction

In eukaryotic cells, gene expressions are intricately regulated through complex interactions of transcription factors (TFs), various regulatory elements and target genes. Deciphering the functions of gene regulatory networks (GRNs) in shaping cell identity and cell fate is one of the central quests in understanding the mapping from genomic blueprints to phenotypes. Over the past decades, much effort has been devoted to developing statistical and computational methods for inferring GRNs from tissue-level bulk data containing genome-wide profiling of gene expression, TF binding, and 3D chromatin structure. More recently, the advent of single-cell sequencing technologies has propelled the study of GRNs into a new era, in which context-specific regulation mechanisms can be investigated. Unlike global GRNs, which inherently aggregate gene interactions over all biological conditions present in a given dataset, context-specific GRNs are tailored to a particular biological setting. These specialized networks detail the regulatory interactions that occur in unique circumstances, such as within specific cell types, lineages, tissues, or under certain environmental conditions. Alongside new opportunities, the sparse and noisy nature of these single-cell data also brings new challenges to the statistical and computational analyses.

A growing number of methods have been developed to extract GRNs from data generated by assays of single-cell RNA-sequencing (scRNA-seq) and single-cell transposase-accessible chromatin sequencing (scATAC-seq). Most of these methods infer the relationships between TFs and target genes by estimating their interactions with *cis*-regulatory elements (CREs) as an inter-

mediate, using information including TF motif enrichment, marginal or conditional correlations between genes and CRE accessibility, and physical proximity between different elements (Duren et al. 2022; Y Jiang et al. 2022; Kartha et al. 2022; Tran et al. 2022; L Zhang et al. 2022). These methods typically work with multimodal data that provide joint profiling of scRNA-seq and scATAC-seq from the same cells, or unpaired data from a matched population of cells, possibly measured over a time course. However, they do not directly address the data integration problem accompanying such data, in which noise, sparsity, and batch effects can obscure identification of cell types and affect the downstream inference of context-specific GRNs. Furthermore, to compare how GRNs dynamically evolve in developmental data, features (e.g., genes, CREs) that are different between time points (or pseudotime points) are identified using differential expression (DE) / accessibility (DA) analyses. While this captures marginal correlations, the features found are not necessarily predictive of the developmental changes.

On a separate front, an increasing number of computational methods have been proposed to perform data integration for single-cell multiomics data from unpaired measurements (Cao and Gao 2022; Gong et al. 2021; Y Lin et al. 2022; Z Zhang et al. 2022). As more technologies capable of multimodal profiling start to emerge (S Chen et al. 2019; Ma et al. 2020; Plongthongkum et al. 2021), integration methods designed for paired data (Argelaguet et al. 2020; Ashuach et al. 2023; Y Hao et al. 2021; Jin et al. 2020) have also attracted significant research interests. However, most of these integration methods do not directly address the immediate downstream problem of inferring GRNs; one exception is GLUE (Cao and Gao 2022), although the GRNs inferred there remain global and not context-specific. One difficulty lies in the fact that most of these methods rely on finding a low-dimensional representation of the datasets across modalities and data batches, and how to extract interpretable biological signals from blackbox methods such as neural networks is a challenging problem. Neural networks offer a conceptual advantage over methods built on linear models, including cross correlation analysis and non-negative matrix factorization, as their superior representation power can capture complex nonlinear interactions in the feature space. However, this comes with the drawback that the relationships between the measured features (e.g., genes) and cellular phenotypes in trained models become more difficult to interpret. Although alternative architectures have been proposed involving linearizing part of the neural network (Svensson et al. 2020), a tradeoff remains between the network's representation power and interpretability.

Here, we propose scTIE, an autoencoder-based method for integrating multimodal profiling of scRNA-seq and scATAC-seq data over a time course and inferring context-specific GRNs. Unlike existing GRN inference methods that study cell type-specific or condition-specific GRNs, scTIE focuses on cellular state transitions and aims to infer GRNs that predict cellular changes along a developmental process. To the best of our knowledge, scTIE provides the first unified framework for the integration of temporal data and the inference of context-specific GRNs that predict cell fates. We achieve this through three main innovations in the architecture design of the autoencoder and the interpretation of a blackbox neural network method. Firstly, scTIE uses iterative optimal transport (OT) fitting to align cells in similar states between different time points and estimate their transition probabilities. scTIE incorporates OT into the loss function of the autoencoder so that the alignment of cells is updated iteratively throughout training to achieve a desirable balance between time point alignment and cell type separation. This is in contrast to many widely used applications of OT in trajectory inference of scRNA-seq data (Schiebinger et al. 2019; Forrow and Schiebinger 2021), where most of the methods solve OT only once on suitably constructed cell distance matrices. Secondly, scTIE removes the need for selecting highly variable genes (HVGs) as input through a pair of coupled batchnorm layers to account for large variations in gene expression levels, making it more robust and generalizable. Thirdly, scTIE provides the means to extract interpretable features from the common embedding space by linking the developmental trajectories of cell representations to their measured features (genes and peaks). We formulate a trajectory prediction problem using the estimated transition probabilities from OT and use gradient-based saliency mapping (P Yang et al. 2021; Ciortan and Defrance 2021) to identify genes and peaks that are potentially driving the cellular state changes. Compared to most GRN inference methods, which focus on developing new ways to construct network relationships among features selected through DE/DA analysis, the main innovation of scTIE lies in selecting these informative features based on their ability to predict cellular changes.

To demonstrate the performance of scTIE on developmental data, we have chosen to focus on multimodal time-course data, as this emerging form of data provides better opportunities to understand the key transcriptional regulatory activities driving a developmental process. To assess scTIE's integration performance against other existing methods, we constructed a variety of synthetic datasets using a mouse early organogenesis multiome dataset. Furthermore, we generated an exemplar dataset comprising paired scRNA-seq and scATAC-seq measurements from $\sim 11,000$ differentiating mouse embryonic stem cells (mESCs) over a time course. Using these datasets, our primary aim was to assess scTIE's ability to integrate multimodal developmental

data for better cell type identification and uncover key regulatory elements predictive of cell fate in a unified framework.

Results

Overview of scTIE

scTIE consists of two main steps. In the first step, scTIE uses modality-specific encoders and decoders to project high dimensional input data from all time points into a lower dimensional common embedding space and reconstruct them in the original space (Fig. 1A). Each encoder-decoder pair is designed to preserve the original information in the input data with minimal information loss, with appropriate loss functions to guide the integration process. For scATAC-seq, accessibility peaks are used as input without conversion to gene activity scores. The encoder and decoder for scRNA-seq use an additional pair of coupled batchnorm layers to handle heterogeneity in gene expression levels and achieve high-fidelity reconstruction of the signals without the need for selecting HVGs. Between consecutive time points, scTIE models cell trajectories using the principle of OT based on the current embeddings and computes an OT loss using the transport cost matrix. The OT loss is incorporated into the total loss function to update the embedded features, aligning cells by their estimated transition probabilities in the trajectories; the cost matrix itself is also updated iteratively throughout training. In addition to the OT loss, a modality alignment loss is used to ensure the projected feature vectors from the two modalities (RNA and ATAC) are close in distance for the same cell.

In the second step, scTIE finetunes the learned embeddings to build a supervised model for predicting cellular transition probabilities for user-selected subgroups of cells (Fig. 1B). Genes and peak regions highly predictive of the cellular transitions are selected by backpropagating the gradients, allowing us to construct GRNs responsible for developmental changes.

We demonstrate the advantages of scTIE on a number of synthetic and real datasets. On synthetic datasets constructed from a mouse early organogenesis multiome dataset, our results reveal that scTIE can effectively align cells from different time points and mitigate batch effects, achieving an optimal balance between time alignment, modality alignment, and cell type separation. Furthermore, our analysis of an exemplar multiome dataset from differentiating mESCs show its superior capacity to capture biological signals from each modality and achieve better day alignment when compared to other methods, resulting in identification of distinct cell sub-

populations. Finally, using developmental transitions from anterior primitive streak as a case study, we demonstrate scTIE's ability to construct lineage-specific GRNs consisting of regulatory elements with a high predictive power of cell fate and identify key regulatory signals that would be missed by DE or DA-based analysis.

scTIE outperforms existing methods in integrating temporal multimodal data.

We first evaluated the data integration performance of scTIE against recent methods designed to integrate paired multimodal data, including Seurat (Y Hao et al. 2021), scAI (Jin et al. 2020), multiVI (Ashuach et al. 2023) and MOFA (Argelaguet et al. 2020). We generated four synthetic datasets by introducing batch effects and noise into a mouse early organogenesis multiome dataset (Argelaguet et al. 2022) (Fig. 2A, Supplementary Fig. S1-S5). As shown in the UMAP plots of the data with synthetic batch effects introduced in RNA and noise introduced in ATAC (Fig. 2A), scTIE effectively removed the batch effects while also better revealing the cell type signals.

Next, we compared the performance of these methods from three aspects quantitatively, namely batch effect removal, time point alignment and their ability to capture cell type signals (Fig. 2B-C). We quantify the quality of batch removal and time point alignment using purity scores, which calculate the proportion of cells from the same batch/sampling time among neighbors of given cells. A lower purity score indicates a better mixing of batch/time points. We measured the cell type preservation using adjusted rand index (ARI) with the cell type annotations provided in the original paper as the ground truth. An ideal embedding should mix well cells from different batches and different time points, while maintaining well-separated cell types. These three metrics are summarized in Fig. 2C, where scTIE encloses the largest area, thus outperforming the other methods in the overall performance (Supplementary Fig. S1). Furthermore, scTIE's superior performance is robust against the number of neighbors used in the purity score calculation (Supplementary Fig. S2). We observe similar trends across the other three synthetic scenarios, where scTIE consistently exhibits better performance than the other methods (Supplementary Fig. S3-S5). Together, we demonstrate the superiority of scTIE in data integration, enabling better capture of biological signals through batch effect removal and time point alignment.

scTIE enables identification of cellular subpopulations via modality and time point alignment with robust performance.

Encouraged by scTIE's performance in data integration, we next generated a temporal single-cell multimodal dataset and leveraged scTIE for the integration of cells across time points and annotation of cell types. We performed single-cell multiome sequencing from mESCs treated with Activin A/Lithium Chloride and measured on Day 2, 4 and 6, using the 10x Chromium Single Cell Multiome platform. After quality control filtering (Supplementary Fig. S6), we obtained high quality measurements of RNA and ATAC from a total of 11,440 cells, with a median detection of 4,130 genes expressed per cell and a median of 11,267 peaks detected per cell.

By clustering on the joint embeddings produced by scTIE, we identified 17 clusters with either distinct transcription or chromatin accessibility profiles that include cell types from all the three germ layers as well as from extra-embryonic layers of embryonic development (Fig. 3A-C). We annotated these clusters based on the key markers identified in the two previous studies (Mittnenzweig et al. 2021; Pijuan-Sala et al. 2019) (Fig. 3C), and confirmed them by label transfer using a public reference (Y Lin et al. 2020; Mittnenzweig et al. 2021) (Supplementary Fig. S7). Further explorations of the motif enrichment of regions with DA in specific clusters highlight the cluster-specific TFs of the annotated cell types (Fig. 3D-E). Additionally, we quantitatively assessed the clustering results using evaluation metrics. Our findings demonstrate that compared with the existing methods, scTIE better preserves biological signals in each modality and achieves better alignment in days, further supporting our annotation of the cells using the integrated data from scTIE (Supplementary Figs. S8-S9). Furthermore, we performed the same training and clustering procedure on two pseudo-replicates constructed by randomly splitting the data into two halves and demonstrated the consistency of the cell type annotation results. The UMAP visualizations for these two subsets are mostly consistent with an overall accuracy rate of 81% across cell types (Supplementary Fig. S10-S11).

Notably, scTIE identifies three distinct clusters of definitive endoderm (Cluster 3, 4 and 7) (Supplementary Fig. S12A). We find that Cluster 4 uniquely expresses several Wnt pathway direct targets (*Vcan*, *Nrcam* and *Ccnd2*) and Wnt TF (LEF1), and has lower expressions in Wnt inhibitors *Dkk1* and some definitive endoderm markers (*Hhex* and *Sox17*) (Supplementary Fig. S12B). The activation of Wnt signaling of this group of cells could be linked to primordial lung

specification progenitors (Ikonomou et al. 2020). Cluster 3 and Cluster 7 have similar expression profiles to each other. Compared with Cluster 3, we find Cluster 7 with majority of cells from Day 6 has lower expressions in Nodal signaling genes *Nodal* and *Tdgf1*, but higher expressions in genes that negatively regulate the Nodal pathway (*Cer1* and *Lefty1*) (Supplementary Fig. S12B).

An inspection of the epiblast subsets further demonstrates that scTIE enables cellular subpopulation identification (Supplementary Fig. S13A). We find that one of the epiblast clusters (Cluster 12) has up-regulation of genes related to Hypoxia (*Adm*, *Anxa2*, *Ddit4* and *Gbe1*), which could enhance the defintive endoderm differentiation, as suggested in (LF Chu et al. 2016; Pimton et al. 2015) (Supplementary Fig. S13B). In addition, we find that Cluster 1 is enriched with anterior epiblast markers (*Pou3f1*, *Enpp3*, *Pten* and *Slc7a3*), while Cluster 10 highly expresses posterior epiblast markers (*Lhx1*, *Ifitm1*) (Supplementary Fig. S13B) (Peng et al. 2016), with down-regulation of the TFs POU5F1 and SOX2 but up-regulation of the TFs FOXA1 and FOXA2 (Supplementary Fig. S13C).

Finally, we examine the stability of our results in both modality alignment and cluster identification, with respect to key tuning parameters in scTIE, including the weight of OT in the loss function, the number of nodes in hidden layer and the updating frequency of OT. We find that the weight of the OT loss is an important parameter to reach a balance between the alignment of modalities and time points, with a larger weight resulting in a better alignment in time points but poorer performance in modality integration (Supplementary Fig. S14A, E, Supplementary Fig. S15A). In this sense, the choice of this parameter can be guided by the performance in modality alignment, since the pairing information for all cells is known and serves as the ground truth. The two other tuning parameters have a small impact on our results (Supplementary Fig. S14B-D, F-H, Supplementary Fig. S15B-D).

Together, we demonstrate that scTIE is able to capture distinct cellular subpopulations by preserving information from both epigenomic and transcriptomic profiles, while also aligning the cells from different time points.

scTIE embeddings capture interpretable biological features.

To interpret the embedding space projected by scTIE, we deconvoluted the latent representation by backpropagating the gradient of each dimension in the embedding layer with respect to gene and peak input, followed by ranking the features. We then computed the enrichment scores of the cell type marker list for the feature rankings of each embedding dimension (see Methods). We find that each dimension exhibits distinct patterns of enrichment of cell type markers, and at the same time the cell types from the same lineage share similar enrichment patterns across the dimensions, indicating that scTIE captures diverse and biologically meaningful information from the data (Fig. 4A). We further observe that the enrichment results of RNA and ATAC share similar patterns, illustrating that scTIE is able to link the transcriptomic profiles with the chromatin accessibility through the common embeddings (Fig. 4A).

The embedding gradients can be further interpreted in terms of known biological functions, based on their Gene Ontology (GO) enrichment. As illustrated in Fig. 4B, we find that the embedding dimensions enriched with definitive endoderm cell type markers can be associated with different pathways. We observe that dimension 39 is uniquely enriched with Activin receptor signaling, as confirmed by the top ranking genes including *Lefty1*, *Fst*, and *Nodal* from this pathway (Fig. 4C). Consistently, the nearest genes of the top ranking peaks also include genes associated with the Activin pathway, such as *Nodal*, *Lefty1* and *Fgf9*. Since treatment by Actinvin is a key component of our differentiation protocol (see Methods), it is comforting to see that the relevance of this pathway is captured by the fitted model. Together, we demonstrate that scTIE is able to project the two modalities into a joint embedding space that captures interpretable biological signals of the data.

Lastly, we find that the above results are robust to the choice of dimension size (i.e., number of nodes in the embedding layer). We trained scTIE and performed the same gradient calculations with the number of dimensions set to 32 and 96 (vs. the current choice of 64) and found qualitatively similar enrichment patterns (Supplementary Fig. S16). Selected embeddings also show enrichment of GO pathways related to the definitive endoderm development, similar to Fig. 4B (Supplementary Fig. S17).

scTIE uncovers cell fate-specific regulatory networks.

scTIE constructs lineage-defining GRNs by combining information across different dimensions of the embedding layer to predict the cell transition probabilities between time points. As a case study, we investigate the transitions of cells from anterior primitive streak on earlier days into endoderm, mesoderm, as well as remaining as anterior primitive streak on later days. The primitive streak is a transient embryonic structure which marks bilateral symmetry, helps confer anterior-posterior spatial information during gastrulation, and initiates germ layer formation (Mikawa et al. 2004). A distinct group of cells located at anterior primitive streak, the node, forms the axial mesodermal structures and definitive endoderm cells (Hoodless et al. 2001).

In each of the above three possible cell fates, we fine-tuned the trained embeddings using a prediction layer with weight regularization and backpropagate the gradients from the prediction layer to select the top 200 genes and 500 peak regions as the most predictive features of the lineage. Compared with the conventional approach that uses DE / DA analysis to select the top features, scTIE selects genes and peak regions with significantly better prediction performance (Fig. 5A). The superior prediction performance is consistent across a range of tuning parameters, including the regularization weights and the number of top features, evaluated via cross validation (Supplementary Fig. S18). To demonstrate the benefit of jointly modeling RNA and ATAC data, we considered the alternative approach of only integrating the RNA data across the time points. Then, we used the same gradient approach to select the top genes for each of the three lineages and selected top peaks by physical proximity and correlation. The predictive power of these features decreased compared to joint modeling (Supplementary Fig. S19). Conceptually, we note that joint modeling allows us to train a separate autoencoder for the ATAC modality and backpropagate the gradients from the prediction layer to select the most informative peaks for predicting the transition probabilities. Thus, the framework of scTIE is capable of jointly finding the most predictive features in both modalities.

In addition, we assessed the stability of the gradient analysis through subsampling. For both the definitive endoderm and mesoderm lineages, we randomly subsampled 60% of the cells considered for each lineage analysis, finetuned the trained neural network and calculated the feature gradients for RNA and ATAC the same way as we did for the full data. The correlations of gradients between the subsampling approach (averaged over 50 repetitions) and the full set demon-

strate a high level of agreement across all genes and peaks used as input to scTIE (Supplementary Fig. S20).

To annotate the top peaks, we overlapped the selected peaks with the published enhancer database from 12 tissues of seven developmental stages from 11.5 days after conception until birth (Gorkin et al. 2020), quantified by the Jaccard index. We find that the top peaks associated with mesoderm transition potential are enriched with facial prominence and limb enhancers at E11.5, while endoderm transition-related peaks identified by scTIE show higher enrichment and distinct overlap with stomach enhancers at E14.5, E15.5 and P0 (Fig. 5B). In contrast, the peaks selected by DA analysis show enrichments in tissues that are much less specific to predicted lineages of mesoderm or endoderm (Supplementary Fig. S21). Together, these results illustrate that scTIE is able to identify peaks that are specific to lineage transition.

The identification of genes and peaks that are predictive of cell transition further allows us to infer GRN for each of the lineages: anterior primitive streak, endoderm and mesoderm (see Methods). In the GRN of anterior primitive streak (Fig. 5C, left panel), we identified a few TFs that play key roles in jointly governing anterior mesendoderm and the node development (LHX1, OTX2 and SMAD4) (GC Chu et al. 2004; Costello et al. 2015), as well as a TF related to axial mesendoderm morphogenesis and patterning (MIXL1) (Hart et al. 2002). When focusing on the endoderm GRN (Fig. 5C, middle panel), we find that besides identifying TFs that are central regulators for the formation of definitive endoderm development (SOX17, GATA4, GATA6, and GSC) (F Li et al. 2011; Fisher et al. 2017; Bossard and Zaret 1998; Kanai-Azuma et al. 2002; Heslop et al. 2021), scTIE also captures TFs that are associated with early mesendoderm differentiation (RUNX1) (VanOudenhove et al. 2016) and morphogenetic movement (LHX1) (Tam and Loebel 2007).

Lastly, we examined the mesoderm GRN (Fig. 5C, right panel) which identifies a few key TFs (HHEX, SOX17, SMAD3, ZIC3, TWIST1 and NFAT5) that are associated with mesoderm lineages. Notably, most of these TFs have insignificant p-values under DE analysis (Table S1), illustrating that scTIE captures key regulatory signals in this lineage that would be missed otherwise. More specifically, the mesoderm GRN highlights TFs that are associated with cardiac development such as ZIC3 in early mesodermal patterning (Z Jiang et al. 2013; Sutherland et al. 2013); HHEX that is involved in mediating the SOX17 for cardiac mesoderm formation in

mESC (Y Liu et al. 2014) and NFAT5 for cardiomyogenic during mesodermal induction through regulating the canonical Wnt pathway (Adachi et al. 2012). We also identify TFs that are essential for mesoderm formation and patterning (SMAD3) (Dunn et al. 2004) and cranial mesoderm development (TWIST1) (Bildsoe et al. 2016).

Discussion

While the rapidly increasing collection of single-cell multiomics data provides a wealth of information for examining context-specific regulatory mechanisms, accurate characterization of cell identities remains the first hurdle to be overcome in such tasks. scTIE provides a unified framework for the integration and joint modeling of temporal multimodal data and the subsequent visualization, cell type identification and inference of key regulatory modules predictive of the developmental transitions of cells. Incorporating OT into the training of an autoencoder, scTIE alternates between updating the alignment of cells at different time points and using the current alignment for training the projections into the common embedding space, thus achieving a better balance between integrating time points and maintaining cell type specific signals. As we have demonstrated on the real and synthetic datasets, scTIE outperforms existing paired methods in terms of integration performance.

Different from existing integration methods that also utilize the notion of a common embedding space, scTIE directly exploits the information in this space produced by the nonlinear projections of a neural network, linking it to interpretable features such as genes and peak regions. scTIE extracts context-specific gene regulatory relationships through the identification of features that are predictive of cell transition probabilities, which quantify how likely a collection of cells on earlier days will transit to a certain cell state on later days, relative to other cells. These sets of cells can be flexibly defined, allowing users to investigate any cell transition process of interest. In addition to cell transition probabilities derived from OT, the current framework can also be adapted to select features that are predictive of other types of response variables, such as pseudotime and perturbation, which potentially enables the construction of differential GRN under continuous cell differentiation and in perturbed conditions.

scTIE is designed for temporal multimodal data, which is ideal for studying single-cell genomics in developmental trajectories. Paired measurements from the same cells remove the need

for computational pairing, which can introduce errors into the downstream GRN analysis if cells of different cell types are paired, and the issue of cell type imbalance between different modalities. The integration of unpaired developmental data across multiple time points remains an open problem itself. For datasets taken from a matched population, a loss function performing global alignment between modalities, such as the one used in Z Zhang et al. 2022, can be potentially incorporated into the training of scTIE. However, the problem is more challenging if cells are sampled at different time points or develop at a different rate across the modalities, and we will pursue this in future work.

Although a large number of methods exist for inferring pseudotime ordering of cells from a static snapshot of a developmental process, pseudotime inference assumes that a continuum of cellular states is observed at the sampled time, and thus may not capture the entire transition process (Tritschler et al. 2019). An interesting extension would be combining pseudotime inference and experimental time points to create a finer temporal resolution. However, we note that this would also increase the computation time of scTIE, since iterative OT estimation is performed between consecutive time points; efficient and accurate OT algorithms remain an active area of research.

We have focused on scRNA-seq and scATAC-seq as common modalities from multimodal profiling technologies. Other modalities such as methylation and protein levels (Mimitou et al. 2021; Swanson et al. 2021; Y Wang et al. 2021) can be easily incorporated into scTIE through appropriate encoder-decoder pairs. Since transcriptional regulation involves interactions of protein complexes, histone modifications and other microenvironmental factors, we expect the addition of such information will allow us to build a more accurate prediction model for cellular state changes. Furthermore, emerging single-cell perturbation assays (Rubin et al. 2019) can either be used to validate the top candidates found in our predictive model, or built into the neural network architecture as a prior knowledge graph (Cao and Gao 2022).

In summary, scTIE provides an integrative framework for analyzing temporal multimodal data, which is an emerging form of data we expect will become more readily available as interests in characterizing GRNs at single-cell resolution continue to rise. On real and synthetic developmental datasets, scTIE is shown to provide effective integration of cells from all time points and select key regulatory elements with superior performance in predicting cellular state

changes. We envision that advances in single-cell technologies generating new forms of temporal data will enable us to further expand the functionalities of scTIE, paving the way towards a holistic understanding of cellular transitions and responses in development and disease.

Methods

Synthetic data construction

The 10x Genomics multiome data of mouse early organogenesis, along with its cell type annotation, was obtained from the Gene Expression Omnibus database under accession number GSE205117 (Argelaguet et al. 2022). The dataset comprises 59,132 cells from a time course of mouse embryonic development, spanning 5 time points from E7.5 to E8.75.

To construct synthetic data that could be processed by most of the methods within their computational capacity, we subset the data to 24,188 cells by selecting only one sample at each time point. We filtered out genes expressed in less than 1% of cells and peaks expressed in less than 5% of cells, resulting in 15,754 genes and 81,108 peaks. To introduce noise and batch effects to the data, we used the `downsampleReads()` function in the DropletUtils R package (R Core Team 2022) to downsample the reads. We generated five synthetic scenarios: (1) subsample 10% for all cells in ATAC; (2) subsample 10% for all cells in ATAC and 50% for all cells in RNA; (3) subsample 50% for half of cells in RNA to create the synthetic batch effect in the data; (4) subsample 50% for half of cells in both RNA and ATAC to create the synthetic batch effect in the data; and (5) subsample 10% for all cells in ATAC, subsample 50% for half of the cells in RNA and 25% for the other half of the cells.

mESC data generation

Cell culture

Mouse embryonic stem cell line R1 was obtained from ATCC. The cells were first expanded on an MEF feeder layer previously irradiated. Then, subculturing was carried out on 0.1% bovine gelatin-coated tissue culture plates. The cells were propagated in mESC medium consisting of Knockout DMEM supplemented with 15% Knockout Serum Replacement, 100 μM nonessential amino acids, 0.5 mM beta-mercaptoethanol, 2 mM GlutaMax, and 100 U/mL Penicillin-Streptomycin with the addition of 1,000 U/mL of LIF (ESGRO, Millipore).

Cell differentiation

mESCs were differentiated using the hanging drop method (X Wang and P Yang 2008). Trypsinized cells were suspended in chemically defined medium CDM (F Li et al. 2011) to a concentration

of 37,500 cells/mL. CDM consists of 75% Iscove's modified Dulbecco's medium (IMDM, Invitrogen), 25% Ham's F12 medium (Invitrogen), 1X N2 supplements (Invitrogen), 0.05% bovine serum albumin (BSA, Invitrogen), 2 mM Glutamax-1 (Invitrogen), 0.5 mM ascorbic acid (Sigma-Aldrich), and 4.5 x 10^4 M MTG (Sigma-Aldrich). 20 μL drops ($\sim$750 cells per drop) were then placed on the lid of a bacterial plate and the lid was upside down. After 48 h incubation at 37°C incubator with 5% CO_2, Embryoid bodies (EBs) formed at the bottom of the drops were collected and placed in the well of a 6-well ultra-low attachment plate (Corning) with fresh CDM medium containing 50 ng/mL Activin A (R&D Systems, 338-AC-050/CF) and 2 mM Lithium Chloride (LiCl, Sigma-Aldrich) for up to 6 days, with the medium being changed daily.

Single cell multiome library

We followed 10x Genomics single cell multiome library preparation protocol. The EBs were collected at Day 2, 4, and 6 after Activin A/Lithium Chloride treatment. For each time point, the cells were first treated with StemPro Accutase Cell Dissociation Reagent (Thermo Fisher Scientific) at 37°C for 10-15 min with pipetting. Single cell suspension was obtained by passing through 37 μM cell strainer (STEMCELL Technologies) twice. After measuring cell concentration, approximately 1 million of cells were centrifuged at 300 rcf for 5 min. Nuclei were isolated by following the protocol provided by 10x Genomics (Nuclei isolation for single cell multiome ATAC + Gene expression sequencing, CG00365, Rev A). The final nuclei concentration was adjusted to 3000 cell/μL in 1X Nuclei Buffer (10x Genomics). The sample was immediately submitted to Stanford Genomics Service Center (SGSC) for single cell sorting using 10x Chromium Controller (target cells: 5000 per replicate, total 2-3 replicates per time point). The singe cell multiome library was generated using Chromium Next GEM Single Cell Multiome ATAC + Gene Expression Reagent Bundle Kit (10x Genomics, PN-1000283).

Data preprocessing

10x Genomics Cell Ranger arc v2.0.0 was used to process the raw FASTQ files for each multiome single-cell dataset separately. The reference genome and transcriptome for alignment and annotation was version arc-mm10-2020-A-2.0.0. To integrate all filtered count matrices for scRNA-seq and scATAC-seq from different replicates and time points, the cellranger-arc aggr command was applied with default depth normalization method.

Next, we performed quality control on the cell level. We removed cells based on the following

criteria in scRNA-seq: (1) with the total number of UMI (nUMI) less than 6000 on Day 2, 3000 on Day 4 and Day 6; (2) with nUMI greater than 100,000; (3) with the number of genes less than 2000 on Day 2, 1800 on Day 4 and 1500 on Day 6 and (4) mitochondrial reads greater than 25%. We further removed cells based on the following criteria in scATAC-seq: (1) with less than 500 total ATAC fragments and (2) with less than 500 peaks detected. After quality control, we retained 11440 cells (Day 2: 2896 cells; Day 4: 2796 cells and Day 6: 5748 cells). We then performed the quality control on the feature level, removing the genes that are not expressed in any cells and the peaks that are expressed at least 5% of cells, resulting in 26717 genes and 61744 peaks as input in scTIE.

Architecture and training of scTIE

scTIE uses an autoencoder structure to project high dimensional feature vectors (i.e., gene expression levels and accessibility peaks) from all time points into a lower dimensional common embedding space and reconstruct the features in the original high dimensional space. Each modality has its own encoder and decoder (Table 1). For RNA, the architecture has an additional pair of coupled batchnorm layers, where the final reconstructed output uses the moving average μ and standard deviation σ stored in the first batchnorm layer of the encoder to perform rescaling. This accounts for the high variability in gene expression levels without the need for selecting HVGs, and allows us to significantly improve the performance in reconstruction correlation, modality and day alignment, and clustering quality (Supplementary Fig. S22). The pairing between feature vectors from the same cell is enforced through a modality loss function minimizing their distance in the embedding space. An OT matrix is used to construct cell trajectories between each pair of consecutive time points. In contrast to existing methods using OT for trajectory inference, we integrate an OT loss into the autoencoder training process and estimate the OT matrix iteratively throughout. A larger weight on the OT loss leads to better alignment between days (Supplementary Fig. S15A).

Let $X^{(t,s)}$ denote the data matrix from time point t and modality s, where $t = 1, \ldots, T$ and $s = 1, 2$ for RNA and ATAC respectively. Each time point t provides measurements for N_t cells; thus in this case, $X^{(t,1)} \in \mathbb{R}^{D_1 \times N_t}$ with $D_1 =$ number of genes and $X^{(t,2)} \in \mathbb{R}^{D_2 \times N_t}$ with $D_2 =$ number of peak regions. In each iteration, a mini-batch of data is sampled by taking equal-sized subsets of cells from each time point, that is, $\mathcal{B} = \{\mathcal{B}^{(t)}\}_{t=1}^{T}$, where each subset $\mathcal{B}^{(t)}$ has B cells. Three loss functions are applied to the mini-batch.

1. *Reconstruction loss.* (f_s, g_s) represents the encoder-decoder pair for modality s. Compared with the architecture for ATAC, the RNA part has a pair of coupled batchnorm layers, starting with a batchnorm layer in the encoder to remove scale variations in genes and prevent the gradients from being dominated by a small number of highly expressed genes (Table 1). Let $x_i^{(t,1)}$ denote the gene expression vector from cell i at time t and $\tilde{x}_i^{(t,1)}$ denote the normalized output from the first batchnorm layer, then $\tilde{x}_i^{(t,1)} = (x_i^{(t,1)} - \mu)/\sigma$, where μ and σ are the moving average and standard deviation of the genes saved in the batchnorm layer throughout training. The reconstruction loss is applied to the normalized data and the output from the decoder, defined as

$$L_{\text{recon}}^{(1)} = \frac{1}{TB} \sum_{t=1}^{T} \sum_{i \in \mathcal{B}^{(t)}} \|\tilde{x}_i^{(t,1)} - g_1(f_1(x_i^{(t,1)}))\|_2^2.$$

   For ATAC, the first layer in the encoder is a fully connected layer and the reconstruction loss is computed on the input $x_i^{(t,2)}$ and output $g_2(f_2(x_i^{(t,2)}))$ as usual. The overall L_{recon} is the sum of $L_{\text{recon}}^{(1)}$ and $L_{\text{recon}}^{(2)}$.

2. *Optimal transport loss.* We leverage OT to effectively align cells from all time points in the embedding space. For notational convenience, we will suppress the dependence on modality s for now, with understanding that the following steps are performed for each modality. For any two adjacent time points t and $t+1$, a transport cost matrix $C^{(t,t+1)} \in \mathbb{R}^{N_t \times N_{t+1}}$ can be computed using the current embeddings, where the (k,l)-th entry is given by $C^{(t,t+1)}(k,l) = ||f(x_k^{(t)}) - f(x_l^{(t+1)})||_2$ for the k-th cell from t and the l-th cell from $t+1$. With the cost matrix, Waddington-OT (Schiebinger et al. 2019) is then used as the algorithm to estimate a transport matrix $\gamma^{(t,t+1)} \in R^{N_t \times N_{t+1}}$. Each row in $\gamma^{(t,t+1)}$ sums to 1, representing the transition probabilities of a cell in time step t to all the other cells in time step $t+1$. Given T time steps, we need to maintain a total of $T-1$ transport matrices throughout the autoencoder training process. For a given mini-batch $\mathcal{B}$ in each iteration, a submatrix version of $C^{(t,t+1)}$ is computed using the rows and columns specified in $\mathcal{B}$ and is denoted by $\tilde{C}^{(t,t+1)}$. Similarly, a mini-batch version $\tilde{\gamma}^{(t,t+1)}$ of $\gamma^{(t,t+1)}$ is calculated by taking the appropriate submatrix and rescaling the rows to unit sum. The batch-wise feature alignment loss (for each modality s) is defined as

$$L_{\text{ot}} = \frac{1}{T-1} \sum_{t=1}^{T} \left(\sum_{k=1}^{B} \sum_{l=1}^{B} (\tilde{C}^{(t,t+1)} \odot \tilde{\gamma}^{(t,t+1)})(k,l) \right),$$

where $\odot$ is the Hadamard product. The final L_{ot} is the sum over modalities s.

3. *Modality alignment loss*. For each mini-batch, the modality alignment loss is simply defined as the L2 distance between feature vectors from the same cell in the embedding space, which is to be minimized:

$$L_{\text{modality}} = \frac{1}{TB} \sum_{t=1}^{T} \sum_{i \in \mathcal{B}^{(t)}} \| f_1(x_i^{(t,1)}) - f_2(x_i^{(t,2)}) \|_2^2 .$$

The total loss in each iteration is $L = \lambda_{\text{recon}} L_{\text{recon}} + \lambda_{\text{ot}} L_{\text{ot}} + L_{\text{modality}}$ where the λ's are tuning parameters controlling the relative weighting of the losses. For every K epochs, the transport matrices (for each modality s) $\gamma_s^{(t,t+1)}, 1 \leq i \leq T-1$ are updated by computing OT on the current embedding features.

The functionalities of each loss function in L:

1. The **reconstruction loss** preserves the original data signals (i.e., distinct cell type signals) at each time point by encouraging the autoencoders to learn a low-dimensional embedding that can reproduce the data input.

2. The **OT loss** aligns embeddings between consecutive time points by calculating an alignment cost function derived from the estimated transition probabilities. To reduce the alignment cost, cell pairs with high transition probabilities should be near each other in the embedding space and vice versa. The transition probabilities and embeddings are iteratively refined. Additionally, the loss aids in mitigating batch effects, as OT can corss-align cells from different batches when mapping cells between consecutive time points. As we demonstrated on the synthetic data with batch effects in both RNA and ATAC data (Supplementary Fig. S4), the pre-training stage (see Training details below), which only trains the RNA autoencoder using the OT loss and the reconstruction loss, already removes most of the batch effects in RNA data (Supplementary Fig. S23).

3. The **modality alignment loss** makes use of the pairing information between RNA and ATAC so that the final embeddings take into account signals in both modalities.

Training details

scTIE took a collection of peak matrices from scATAC-seq data and raw couns matrices from scRNA-seq data from multiple time points as input. For ATAC, the peak matrices were trans-

formed to binary matrices, where one represents any non-zero original values. For RNA, the raw count matrices were sized-factor normalized and then log-transformed. For the overall multimodal training, we first pre-trained the RNA autoencoder f_1, g_1 for 500 epochs (excluding L_{modality}). Then, we fixed the weights of the pretrained RNA model to train the ATAC model for 300 epochs with the overall loss L. Finally, the two models were jointly trained for 200 epochs using the full algorithm as detailed in Algorithm 1. The final joint embeddings were calculated by taking the averages of $f_1(x_i^{(t,1)})$ and $f_2(x_i^{(t,2)})$ for each cell i from time t, followed by computing the final $\gamma^{(t,t+1)}$ from the joint embeddings. Throughout training, we used Adam as the optimizer with learning rate set to 0.1, batch size $B = 256$, tuning parameters $\lambda_{\text{recon}} = 1$, $\lambda_{\text{ot}} = 0.1$, and OT was updated every 10 epochs.

We note here that due to the pre-training of the RNA autoencoder, the biological signals utilized by scTIE to produce the common embeddings were mostly driven by the RNA modality. However, complementary signals from scATAC-seq still play a role in generating the embeddings since the modality alignment loss is affected by both RNA and ATAC positions in the embedding space. Pre-training with RNA signals is essential for stable training of the neural network because (i) the RNA modality generally contains stronger signals for cell type identification and (ii) the dimension of ATAC input (number of peaks) is much larger than that of the RNA modality (number of genes).

Algorithm 1 Multimodal OT Autoencoder (two-modality case)

Data matrices $X^{(t,s)}$, training iterations M, batch size B, autoencoder f_1, g_1, f_2, g_2 with weights θ, learning rate α, loss weight tuning parameters $\lambda_{\text{recon}}, \lambda_{\text{ot}}$, OT update frequency K.

Initialize all $\gamma_s^{(t,t+1)}, 1 \le t \le T-1$ matrices with zero matrices.

for $iteration = 1, 2, \ldots, M$ **do**

Sample cells $\mathcal{B} = \{\mathcal{B}^{(t)}\}_{t=1}^{T}$, where each subset $\mathcal{B}^{(t)}$ has B cells.

Compute L_{recon}, L_{ot}, L_{modality}

Compute $L = \lambda_{\text{recon}} L_{recon} + \lambda_{ot} L_{ot} + L_{modality}$

Perform gradient descent step on autoencoder weights $\theta \leftarrow \theta - \alpha \nabla_\theta L$

if $M \% K == 0$ **then**

Update $\gamma_s^{(t,t+1)}, 1 \le t \le T-1, s = 1, 2$ using current embeddings.

end if

end for

Estimation of long-range transition probabilities

Long-range transition probabilities can be estimated by multiplying the transport matrices. For example, $\gamma^{(t,t+2)}$ can be calculated as $\gamma^{(t,t+1)}\gamma^{(t+1,t+2)}$. An alternative approach is to compute $\gamma^{(t,t+2)}$ directly from OT. However, since OT interpolates between two observed datasets by finding the shortest path in the space of distributions, one has to implicitly assume that the cells do not change their expression or accessibility by large amounts over the two time points. It is generally recommended that long-range time couplings are estimated by multiplying the gamma matrices (Schiebinger et al. 2019). On the mESC dataset, these two ways of estimation give positively correlated results, with the mode of correlations lying around 0.6 (Supplementary Fig. S24).

Cell type annotation of mESC data

Cell clustering of scTIE

To identify the clusters on the common embedding of scTIE, we first constructed a shared nearest neighbor graph using `buildSNNGraph` in R package `scran` (Lun et al. 2016) (v 1.23.0), with the number of nearest neighbor set as 15 with weighted scheme set as jaccard. Next we performed Leiden community detection (Traag et al. 2019) on the shared nearest graph with resolution 1.8 and number of iterations 50, implemented in R package `leidenAlg` (v 1.0.3), resulting in 17 clusters in total.

Motif enrichment

We used `Signac` (Stuart et al. 2021) to calculate the over-represented motif of each cluster based on the differential accessible peaks. The motif position frequency matrices are obtained from Cis-BP (Weirauch et al. 2014). We used `limma-trend` (Ritchie et al. 2015) to perform differential accessibility analysis between the cells in one cluster and the remaining cells, where the top 500 peaks of each cluster with log fold change greater than 0.1 and adjusted p-value less than 0.001 are selected. We then performed the motif enrichment analysis using `FindMotifs` to find motifs over-represented in the selected set of peaks.

Benchmarking and evaluation metrics

Settings used in other methods

We benchmarked the performance of scTIE against four other methods designed for single-cell paired multimodal data integration: Seurat, scAI, MultiVI and MOFA. We compared scTIE's performance in terms of visualisation of the latent space, alignment of the days and clustering in the latent space against these methods.

- **Seurat**. R package Seurat v4.1.0 (Y Hao et al. 2021) was used. We ran Seurat (WNN) using `FindMultiModalNeighbors`, with the reduction list input as the first 50 components of LSI reduced dimension of scATAC-seq (with the first dimension excluded) and 50 top PCs of scRNA-seq, with other parameters set as default.

- **scAI**. R package scAI v1.0.0 (Jin et al. 2020) was used. We ran scAI using `run_scAI` by setting the rank of the inferred factor set as 64 and `nrun = 5`, with other parameters set as default.

- **MultiVI.** Python package scvi v0.15.0 (Ashuach et al. 2023) was used. We ran MultiVI using `MULTIVI` by setting the `fully_paried = True`, `n_hidden = 256` and `n_latent = 64`, with other parameters set as default. The model was then trained with `max_epochs = 200`.

- **MOFA**. R package MOFA2 v1.7.0 (Argelaguet et al. 2020) was used. We ran MOFA using `run_mofa` by setting the number of factors as 64, with other parameters set as default.

Benchmarking of mESC data

Modality alignment: We used two metrics to measure scTIE's performance in the alignment of the two modalities, namely FOSCTTM and paired data proportion.

- *FOSCTTM*. FOSCTTM refers to Fraction of Samples Closer than True Match, which is first introduced in MMD-MA (J Liu et al. 2019) to quantify the alignment of multi-omics data. To evaluate the modal alignment of scTIE using FOSCTTM, we first calculated the Euclidean distance between the ATAC embedding and RNA embedding. Then for each modality we calculated one FOSCTTM score, which summarizes the proportion of cells that are closer to the ground truth matched cells based on the distance matrix. Finally we

summarized the FOSCTTM scores from the two modalities into one score by taking the average.

- *Paired data proportion*. Paired data proportion (used in Cobolt (Gong et al. 2021)) calculated the proportion of cells whose ground truth matched cells are included within a certain number of neighbors, based on the Euclidean distance between the ATAC embedding and RNA embedding. We varied the number of neighbors from 1 to the total number of cells in the data.

Day alignment: We quantified the alignment of data sampled on different days using neighborhood purity using `neighborPurity` in R package `bluster` (v1.5.1), which calculated the proportion of cells from the same day among a certain number of neighbors, based on the UMAP coordinates generated from the common latent embeddings.

Comparison with single-modality clustering: We benchmarked clustering results from scTIE against other paired data integration methods by evaluating how similar the results are compared to clustering dimension-reduced scRNA-seq (PCA space) or scATAC-seq (LSI space) alone. On the latent space of each method or the dimension-reduced space from scRNA-seq or scATAC-seq, we performed Leiden clustering on the shared nearest neighbor graphs constructed, with the same parameter settings as mentioned in Section *Cell clustering*. Note that for Seurat, we performed Leiden clustering directly on the weighted nearest neighbor graph it outputs. We used two metrics to quantify the results, Adjusted Rand Index and silhouette coefficient.

- *Adjusted Rand Index (ARI)*. We computed the ARI scores of clustering results from each data integration method and clustering results from scRNA-seq or scATAC-seq alone.

- *Silhoutte coefficient*. For each clustering result, we computed the silhouette coefficient based on the Euclidean distance calculated from the UMAP coordinates generated from the dimension-reduced scRNA-seq or scATAC-seq.

For both metrics, higher values indicate a method better captures the clustering information in a single modality.

Benchmarking of synthetic data

We benchmarked the data integration performance of scTIE with the other paired data integration methods in terms of three evaluation metrics: (1) ARI scores of the cell type annotation provided

by the original study and the Leiden clustering results from each method; (2) neighborhood purity of days; and (3) neighborhood purity of batch for scenarios with synthetic batch effects.

Enrichment analysis for embedding dimensions

Upon completion of training, scTIE has projected the high dimensional feature vectors (genes and peaks) into a 64 dimensional embedding space. Treating each dimension as a representation unit, for each cell type, we backpropagate the gradient of each unit with respect to gene and peak input to select features with the largest impact. More specifically, for each cell in cell type G, we pass its gene expression vector through the autoencoder to obtain its embedding vector y and compute $\frac{\partial y_j}{\partial x_i}$ for each dimension j and gene input node i. The gradients are averaged over all cells in G to obtain the mean gradient for each gene. We then take the variability of gene expression into account by multiplying each mean gradient by its corresponding gene standard deviation, so that the final gradients are equivalent to gradients after the first batchnorm layer. Finally, we rank the genes by their gradient values and calculate the enrichment scores of the top 200 genes from the DE analysis of cell type G, where the DE analysis is performed using `limma-trend` (Ritchie et al. 2015) between the cells in one cluster and the remaining cells. Similar steps are performed for the peaks and the top 500 peaks are selected for enrichment score calculation.

We used `fgsea` function in the R package `fgsea` (Korotkevich et al. 2021) to perform the gene set enrichment analysis (GSEA) on the pathways related to mouse embryonic stem cells (as listed in **Fig. 4B**). Significant pathways are defined with adjusted p-value less than 0.05.

GRN inference

Selecting features with high predictive power

By building a prediction framework on the obtained transition probabilities, scTIE selects genes and peaks jointly with high predictive power for developmental outcomes. In the mESC data, we consider how a group of cells from earlier days, denoted as G_0, develops into two other groups G_1 and G_2 on later days.

The transition probabilities are obtained from $\gamma^{(t,t+1)}$ ($t = 1, 2$ in our data) so that each cell i in G_0 is associated with a probability vector (p_{i1}, p_{i2}) indicating its probabilities of becoming G_1 and G_2 (See Section *Cell transition probability calculation*). We finetune a one-layer classifier on the pretrained features in the embedding space of cells in G_0 to predict their transition prob-

abilities. A simple linear classifier is sufficient to partition the cell feature space into G_1 and G_2 when the pretrained features are representative enough. Concretely, let q be the linear classifier and $\mathcal{B}$ be a mini-batch of cells from G_0 of size B, we employ a batch-wise KL divergence loss defined

$$L_{kl} = \frac{1}{B} \sum_{j \in \mathcal{B}} D_{KL}(q(f(x_j)) || P_j),$$

where f is the trained encoder, $P_j = (p_{j1}, p_{j2})$. This loss enforces the classifier q to output transition probability distributions close to those in P_j's. We also include the modality alignment loss L_{modality}, with weight default set as 0.1. The classifier is trained with Adam setting learning rate to 0.001, training epochs to 200, batch size to 256 and L1 regularization.

After training, gradients from the two classification nodes are backpropagated to each gene (or peak) input the same way as in computing embedding gradients. The gene gradients are then scaled by multiplying with the gene-wise standard deviations. A positive gradient for gene (or peak) j with respect to the node for G_1 means increasing the input feature value tend to increase the cells' probabilities of becoming G_1, while a negative value indicates more contribution to G_2. The final feature ranking is based on the average gradients by repeating this procedure 20 times with different seeds.

Selection of G_0, G_1, G_2

As a case study in this paper, we focus on the transition of cells from anterior primitive streak on Day 2 and Day 4 into endoderm, mesoderm, as well as remaining as anterior primitive streak on Day 4 and Day 6.

First, we considered the cells that are annotated as anterior primitive streak (Cluster 6) on Day 2 and Day 4 as G_0. G_1 and G_2 are then selected from the cells on Day 4 and Day 6 that are more likely to be the descendants of G_0, as quantified by the descendant scores. The descendant scores are defined similarly as in WOT (Schiebinger et al. 2019). Recall $\gamma^{(t,t+1)}$ is the N_t by N_{t+1} transition probability matrix between time points t and $t+1$, let $s_t \in R^{N_t}$ be the vector of descendant scores for all cells at time point t, then we can calculate

$$s_{t+1} = s_t \gamma^{(t,t+1)}, \text{ where } s_t(i) = \begin{cases} \frac{1}{|G_0|}, & \text{if cell } i \text{ is in } G_0, \\ 0, & \text{otherwise.} \end{cases}.$$

This formula can then be pushed forward again to calculate the descendant scores for the next time point $t+2$, and so on. For all cells in G_0 at time point t (here $t = 1$ or 2), we calculated

the descendant scores s_{t+k} of all cells at the later time point $t+k$, for $k = 1, ..., T-t$. We then considered the cells with descendant scores greater than the median of all cells at a certain time point as the potential descendants, i.e., cells with $s_{t+k}(i) > \text{median}(s_{t+k})$. Among these descendant cells, we selected three pairs of G_1 and G_2 corresponding to the three cell fates we have analyzed: G_1 that are annotated as (1) anterior primitive streak or (2) definitive endoderm or (3) mesoderm; for each selection of G_1, G_2 always represents the remaining descendant cells.

Cell transition probability calculation

For each cell $i \in G_0$ on Day t, and G_1, G_2 on Day $k \in K$, where $K = \{k : t < k \leq T\}$, the transition probability vector $(p_{i1}^{(t)}, p_{i2}^{(t)})$ are calculated as the following,

$$
\begin{aligned}
p_{i1}^{(t,k)} &= \sum_{y \in G_1} \gamma^{(t,k)}(i, y), \\
p_{i2}^{(t,k)} &= \sum_{y \in G_2} \gamma^{(t,k)}(i, y), \\
p_{ij}^{(t,k)} &= \frac{p_{ij}^{(t)}}{\sum_j p_{ij}^{(t)}}, j = 1, 2, \\
p_{ij}^{(t)} &= \frac{1}{|K|} \sum_k p_{ij}^{(t,k)}.
\end{aligned}
$$

(p_{i1}, p_{i2}) is then the concatenated vector of $(p_{i1}^{(t)}, p_{i2}^{(t)})$.

Evaluation of cell transition probability prediction

To evaluate the predictive power of the selected features to the transition probability, we performed support vector machine (SVM) with radial kernel to predict the transition probability using Day 2 and 4 anterior primitive streak gene expression of the top selected genes and peak matrix of the top selected peaks. The performance are quantified by root mean squared error (RMSE) from a 20 repeated 5 fold cross validation. We benchmarked the predictive power of the features selected by gradients with different regularization weights (0, 1, 10, 100), against the features selected by DE/DA analysis using limma-trend (Ritchie et al. 2015).

Gene regulatory network construction

To construct the gene regulatory network for each cell fate (anterior primitive streak, definitive endoderm and mesoderm), we focus on the top 500 genes based on the gradient ranking. For each

gene, we consider the open chromatin regions that are within 250kb upstream and downstream of its transcription start site (TSS) as well as ranked top 2000 according to the gradients as the distal candidate functional regions, which results in 396, 404 and 339 gene-peak pairs for the three cell fates respectively. We next filtered the pairs based on the gene-peak correlation, calculated from the meta-cells. The meta-cells were constructed using the following strategies: we first randomly selected 100 cells from the anterior primitive streak cells on Day 2. For each cell, we looked for its 5 nearest neighbors based on the euclidean distances of the common embeddings and aggregated them as a meta-cell. Then, we calculated Pearson's correlation of the gene-peak pairs for these 100 meta-cells. This procedure is repeated 20 times and the gene-peak pairs with an absolute average correlation greater than 0.2 are retained (APS: 35, DE: 38 and MES: 17 pairs remained).

To link the peak region with the TF, we identified the enriched TF using *matchMotifs* function in R package *motifmatchr* of the peaks from the selected gene-peak pairs based on Cis-BP database (Weirauch et al. 2014). We only consider if the TF are the top 500 genes. Finally, by linking the TF-region and peak-gene relationships, we construct the TF-gene regulatory networks that are associated cell fate probabilities.

In the alternative approach of only integrating RNA across time, we selected the peaks that are within 250kb upstream and downstream of the transcription start site (TSS) of the top ranking genes, with Pearson's correlation greater than 0.2.

Data access

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE223041.

scTIE is available at `https://github.com/SydneyBioX/scTIE` and as part of the Supplementary Code.

Acknowledgements

We would like to thank Michael Blanco and Dhananjay Wagh from Stanford Genomics Service Center (SGSC) for their kind help on the preparation of 10x Genomics single cell multiome libraries. We also want to thank Xuhuai Ji form SGSC for providing sequencing services. The Illumina HiSeq 4000 was purchased using a NIH S10 Shared Instrumentation Grant (S10OD018220). The Illumina NovaSeq 6000 was also purchased using a NIH S10 Shared Instrumentation Grant (1S10OD02521201). The authors gratefully acknowledge the following funding sources: Research Training Program Tuition Fee Offset and Stipend Scholarship and Chen Family Research Scholarship to Y.L.; AIR@innoHK programme of the Innovation and Technology Commission of Hong Kong to J.Y.H.Y. and Y.L.; the UT Austin Harrington Faculty Fellowship to Y.X.R.W; NIH grants R01 HG010359 and P50 HG007735 to W.H.W.

Author contributions: T.W., W.H.W. and Y.X.R.W. conceived and designed this project; X.C. performed the mESC multiome experiment; Y.L., T.W., S.W., B.C., and J.X. performed data preprocessing, model development, and evaluation of results; J.Y.H.Y., W.H.W. and Y.X.R.W. supervised the execution; Y.L., B.C., J.X., J.Y.H.Y., W.H.W. and Y.X.R.W. wrote the manuscript. All authors read and approved the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

References

Adachi A, Takahashi T, Ogata T, Imoto-Tsubakimoto H, Nakanishi N, Ueyama T, and Matsubara H. 2012. NFAT5 regulates the canonical Wnt pathway and is required for cardiomyogenic differentiation. *Biochemical and biophysical research communications*. **426**: 317–323.

Argelaguet R, Arnol D, Bredikhin D, Deloro Y, Velten B, Marioni JC, and Stegle O. 2020. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome biology*. **21**: 1–17.

Argelaguet R, Lohoff T, Li JG, Nakhuda A, Drage D, Krueger F, Velten L, Clark SJ, and Reik W. 2022. Decoding gene regulation in the mouse embryo using single-cell multi-omics. *bioRxiv*. 2022–06.

Ashuach T, Gabitto MI, Koodli RV, Saldi GA, Jordan MI, and Yosef N. 2023. MultiVI: deep generative model for the integration of multimodal data. *Nature Methods*. **20**: 1222–1231.

Bildsoe H, Fan X, Wilkie EE, Ashoti A, Jones VJ, Power M, Qin J, Wang J, Tam PP, and Loebel DA. 2016. Transcriptional targets of TWIST1 in the cranial mesoderm regulate cell-matrix interactions and mesenchyme maintenance. *Developmental biology*. **418**: 189–203.

Bossard P and Zaret KS. 1998. GATA transcription factors as potentiators of gut endoderm differentiation. *Development*. **125**: 4909–4917.

Cao ZJ and Gao G. 2022. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nature Biotechnology*. 1–9.

Chen S, Lake BB, and Zhang K. 2019. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nature biotechnology*. **37**: 1452–1457.

Chu LF, Leng N, Zhang J, Hou Z, Mamott D, Vereide DT, Choi J, Kendziorski C, Stewart R, and Thomson JA. 2016. Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. *Genome biology*. **17**: 1–20.

Chu GC, Dunn NR, Anderson DC, Oxburgh L, and Robertson EJ. 2004. Differential requirements for Smad4 in TGFβ-dependent patterning of the early mouse embryo.

Ciortan M and Defrance M. 2021. Explainability methods for differential gene analysis of single cell RNA-seq clustering models. *bioRxiv*.

Costello I, Nowotschin S, Sun X, Mould AW, Hadjantonakis AK, Bikoff EK, and Robertson EJ. 2015. Lhx1 functions together with Otx2, Foxa2, and Ldb1 to govern anterior mesendoderm, node, and midline development. *Genes & development*. **29**: 2108–2122.

Dunn NR, Vincent SD, Oxburgh L, Robertson EJ, and Bikoff EK. 2004. Combinatorial activities of Smad2 and Smad3 regulate mesoderm formation and patterning in the mouse embryo.

Duren Z, Chang F, Naqing F, Xin J, Liu Q, and Wong WH. 2022. Regulatory analysis of single cell multiome gene expression and chromatin accessibility data with scREG. *Genome biology*. **23**: 1–19.

Fisher J, Pulakanti K, Rao S, and Duncan S. 2017. GATA6 is essential for endoderm formation from human pluripotent stem cells. *Biology Open*. **6**: 1084–1095.

Forrow A and Schiebinger G. 2021. LineageOT is a unified framework for lineage tracing and trajectory inference. *Nature communications*. **12**: 1–10.

Gong B, Zhou Y, and Purdom E. 2021. Cobolt: integrative analysis of multimodal single-cell sequencing data. *Genome biology*. **22**: 1–21.

Gorkin DU, Barozzi I, Zhao Y, Zhang Y, Huang H, Lee AY, Li B, Chiou J, Wildberg A, Ding B, et al. 2020. An atlas of dynamic chromatin landscapes in mouse fetal development. *Nature*. **583**: 744–751.

Hao Y, Hao S, Andersen-Nissen E, Mauck III WM, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, et al. 2021. Integrated analysis of multimodal single-cell data. *Cell*. **184**: 3573–3587.

Hart AH, Hartley L, Sourris K, Stadler ES, Li R, Stanley EG, Tam PP, Elefanty AG, and Robb L. 2002. Mixl1 is required for axial mesendoderm morphogenesis and patterning in the murine embryo.

Heslop JA, Pournasr B, Liu JT, and Duncan SA. 2021. GATA6 defines endoderm fate by controlling chromatin accessibility during differentiation of human-induced pluripotent stem cells. *Cell reports*. **35**: 109145.

Hoodless PA, Pye M, Chazaud C, Labbé E, Attisano L, Rossant J, and Wrana JL. 2001. FoxH1 (Fast) functions to specify the anterior primitive streak in the mouse. *Genes & development*. **15**: 1257–1271.

Ikonomou L, Herriges MJ, Lewandowski SL, Marsland R, Villacorta-Martin C, Caballero IS, Frank DB, Sanghrajka RM, Dame K, Kańduła MM, et al. 2020. The in vivo genetic program of murine primordial lung epithelial progenitors. *Nature communications*. **11**: 1–17.

Jiang Y, Harigaya Y, Zhang Z, Zhang H, Zang C, and Zhang NR. 2022. Nonparametric single-cell multiomic characterization of trio relationships between transcription factors, target genes, and cis-regulatory regions. *Cell Systems*. **13**: 737–751.

Jiang Z, Zhu L, Hu L, Slesnick TC, Pautler RG, Justice MJ, and Belmont JW. 2013. Zic3 is required in the extra-cardiac perinodal region of the lateral plate mesoderm for left–right patterning and heart development. *Human molecular genetics*. **22**: 879–889.

Jin S, Zhang L, and Nie Q. 2020. scAI: an unsupervised approach for the integrative analysis of parallel single-cell transcriptomic and epigenomic profiles. *Genome biology*. **21**: 1–19.

Kanai-Azuma M, Kanai Y, Gad JM, Tajima Y, Taya C, Kurohmaru M, Sanai Y, Yonekawa H, Yazaki K, Tam PP, et al. 2002. Depletion of definitive gut endoderm in Sox17-null mutant mice.

Kartha VK, Duarte FM, Hu Y, Ma S, Chew JG, Lareau CA, Earl A, Burkett ZD, Kohlway AS, Lebofsky R, et al. 2022. Functional inference of gene regulation using single-cell multi-omics. *Cell genomics*. **2**: 100166.

Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, and Sergushichev A. 2021. Fast gene set enrichment analysis. *BioRxiv*. 060012.

Li F, He Z, Li Y, Liu P, Chen F, Wang M, Zhu H, Ding X, Wangensteen KJ, Hu Y, et al. 2011. Combined activin A/LiCl/Noggin treatment improves production of mouse embryonic stem cell-derived definitive endoderm cells. *Journal of cellular biochemistry*. **112**: 1022–1034.

Lin Y, Cao Y, Kim HJ, Salim A, Speed TP, Lin DM, Yang P, and Yang JYH. 2020. scClassify: sample size estimation and multiscale classification of cells using single and multiple reference. *Molecular systems biology*. **16**: e9389.

Lin Y, Wu TY, Wan S, Yang JY, Wong WH, and Wang Y. 2022. scJoint integrates atlas-scale single-cell RNA-seq and ATAC-seq data with transfer learning. *Nature Biotechnology*. **40**: 703–710.

Liu J, Huang Y, Singh R, Vert JP, and Noble WS 2019. Jointly embedding multiple single-cell omics measurements. In: *Algorithms in bioinformatics: International Workshop, WABI, proceedings. WABI (Workshop)*. Vol. 143. NIH Public Access.

Liu Y, Kaneda R, Leja TW, Subkhankulova T, Tolmachov O, Minchiotti G, Schwartz RJ, Barahona M, and Schneider MD. 2014. Hhex and Cer1 mediate the Sox17 pathway for cardiac mesoderm formation in embryonic stem cells. *Stem cells*. **32**: 1515–1526.

Lun AT, McCarthy DJ, and Marioni JC. 2016. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research*. **5**:

Ma S, Zhang B, LaFave LM, Earl AS, Chiang Z, Hu Y, Ding J, Brack A, Kartha VK, Tay T, et al. 2020. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell*. **183**: 1103–1116.

Mikawa T, Poh AM, Kelly KA, Ishii Y, and Reese DE. 2004. Induction and patterning of the primitive streak, an organizing center of gastrulation in the amniote. *Developmental dynamics: an official publication of the American Association of Anatomists*. **229**: 422–432.

Mimitou EP, Lareau CA, Chen KY, Zorzetto-Fernandes AL, Hao Y, Takeshima Y, Luo W, Huang TS, Yeung BZ, Papalexi E, et al. 2021. Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nature biotechnology*. **39**: 1246–1258.

Mittnenzweig M, Mayshar Y, Cheng S, Ben-Yair R, Hadas R, Rais Y, Chomsky E, Reines N, Uzonyi A, Lumerman L, et al. 2021. A single-embryo, single-cell time-resolved model for mouse gastrulation. *Cell*. **184**: 2825–2842.

Peng G, Suo S, Chen J, Chen W, Liu C, Yu F, Wang R, Chen S, Sun N, Cui G, et al. 2016. Spatial transcriptome for the molecular annotation of lineage fates and cell identity in mid-gastrula mouse embryo. *Developmental cell*. **36**: 681–697.

Pijuan-Sala B, Griffiths JA, Guibentif C, Hiscock TW, Jawaid W, Calero-Nieto FJ, Mulas C, Ibarra-Soria X, Tyser RC, Ho DLL, et al. 2019. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature*. **566**: 490–495.

Pimton P, Lecht S, Stabler CT, Johannes G, Schulman ES, and Lelkes PI. 2015. Hypoxia enhances differentiation of mouse embryonic stem cells into definitive endoderm and distal lung cells. *Stem cells and development*. **24**: 663–676.

Plongthongkum N, Diep D, Chen S, Lake BB, and Zhang K. 2021. Scalable dual-omics profiling with single-nucleus chromatin accessibility and mRNA expression sequencing 2 (SNARE-Seq2). *Nature Protocols*. **16**: 4992–5029.

R Core Team 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, and Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*. **43**: e47–e47.

Rubin AJ, Parker KR, Satpathy AT, Qi Y, Wu B, Ong AJ, Mumbach MR, Ji AL, Kim DS, Cho SW, et al. 2019. Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks. *Cell*. **176**: 361–376.

Schiebinger G, Shu J, Tabaka M, Cleary B, Subramanian V, Solomon A, Gould J, Liu S, Lin S, Berube P, et al. 2019. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*. **176**: 928–943.

Stuart T, Srivastava A, Madad S, Lareau CA, and Satija R. 2021. Single-cell chromatin state analysis with Signac. *Nature methods*. **18**: 1333–1341.

Sutherland MJ, Wang S, Quinn ME, Haaning A, and Ware SM. 2013. Zic3 is required in the migrating primitive streak for node morphogenesis and left–right patterning. *Human molecular genetics*. **22**: 1913–1923.

Svensson V, Gayoso A, Yosef N, and Pachter L. 2020. Interpretable factor models of single-cell RNA-seq via variational autoencoders. *Bioinformatics*. **36**: 3418–3421.

Swanson E, Lord C, Reading J, Heubeck AT, Genge PC, Thomson Z, Weiss MD, Li Xj, Savage AK, Green RR, et al. 2021. Simultaneous trimodal single-cell measurement of transcripts, epitopes, and chromatin accessibility using TEA-seq. *Elife*. **10**: e63632.

Tam PP and Loebel DA. 2007. Gene function in mouse embryogenesis: get set for gastrulation. *Nature Reviews Genetics*. **8**: 368–381.

Traag VA, Waltman L, and Van Eck NJ. 2019. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific reports*. **9**: 1–12.

Tran A, Yang P, Yang JY, and Ormerod JT. 2022. scREMOTE: Using multimodal single cell data to predict regulatory gene relationships and to build a computational cell reprogramming model. *NAR genomics and bioinformatics*. **4**: lqac023.

Tritschler S, Büttner M, Fischer DS, Lange M, Bergen V, Lickert H, and Theis FJ. 2019. Concepts and limitations for learning developmental trajectories from single cell genomics. *Development*. **146**: dev170506.

VanOudenhove JJ, Medina R, Ghule PN, Lian JB, Stein JL, Zaidi SK, and Stein GS. 2016. Transient RUNX1 expression during early mesendodermal differentiation of hESCs promotes epithelial to mesenchymal transition through TGFB2 signaling. *Stem Cell Reports*. **7**: 884–896.

Wang X and Yang P. 2008. In vitro differentiation of mouse embryonic stem (mES) cells using the hanging drop method. *JoVE (Journal of Visualized Experiments)*. e825.

Wang Y, Yuan P, Yan Z, Yang M, Huo Y, Nie Y, Zhu X, Qiao J, and Yan L. 2021. Single-cell multiomics sequencing reveals the functional regulatory landscape of early embryos. *Nature communications*. **12**: 1–14.

Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, et al. 2014. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell*. **158**: 1431–1443.

Yang P, Huang H, and Liu C. 2021. Feature selection revisited in the single-cell era. *Genome Biology*. **22**: 1–17.

Zhang L, Zhang J, and Nie Q. 2022. DIRECT-NET: An efficient method to discover cis-regulatory elements and construct regulatory networks from single-cell multiomics data. *Science Advances*. **8**: eabl7393.

Zhang Z, Yang C, and Zhang X. 2022. scDART: integrating unmatched scRNA-seq and scATAC-seq data and learning cross-modality relationship simultaneously. *Genome Biology*. **23**: 1–28.

Figure legends

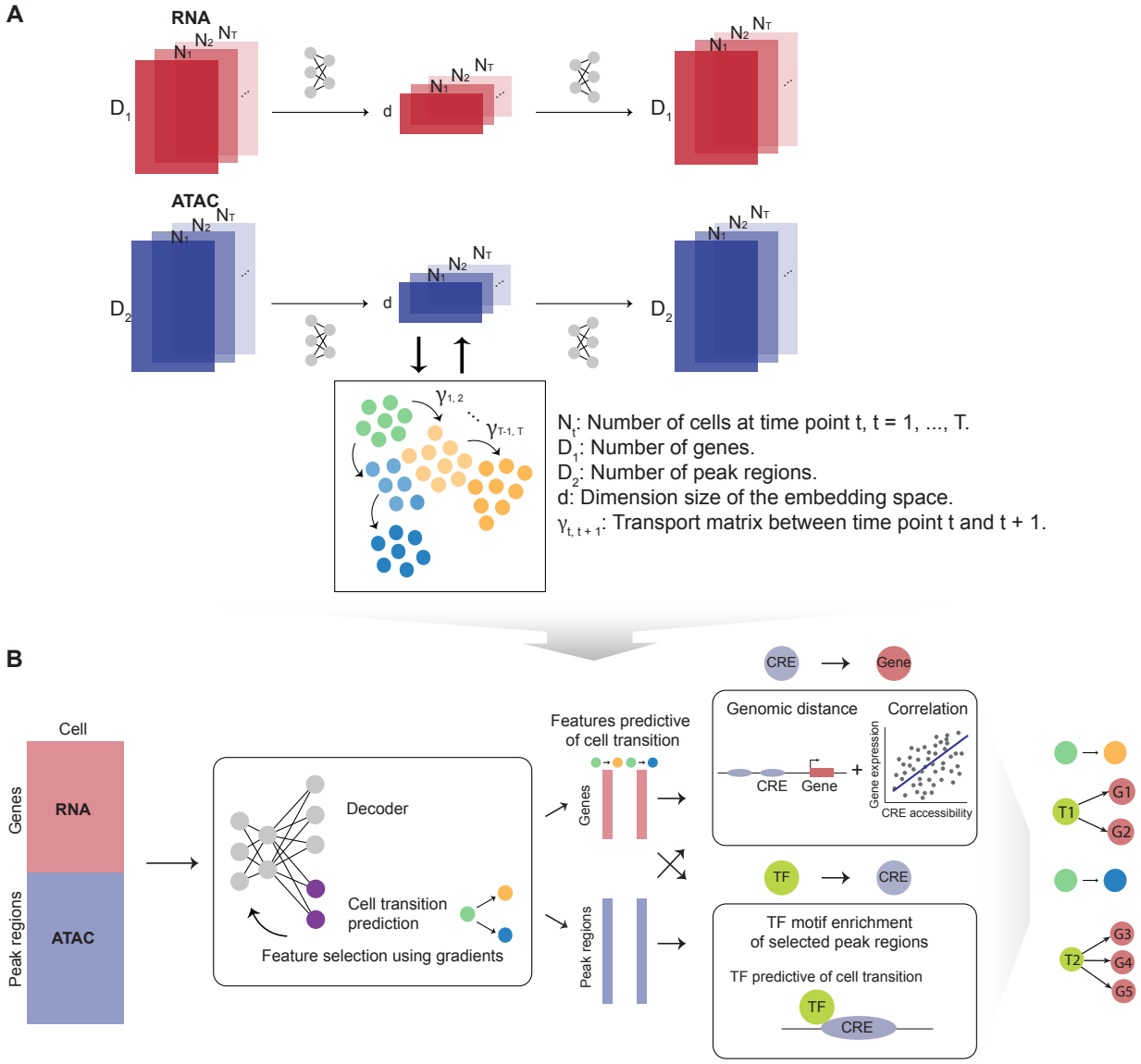


Figure 1: Overview of scTIE, a unified framework for the integration of temporal data and the inference of context-specific GRNs that predict cell fates. The input of scTIE consists of the gene expression matrix of scRNA-seq and peak matrix of scATAC-seq from single-cell multiome data over a time course. scTIE consists of two main steps: (A) In the first step, each cell, represented by a pair of gene and peak vectors, is projected into a common embedding space by separate encoders and decoders. The two modalities and time points are aligned by appropriate loss functions, while the transition probability matrix between cells from consecutive time points are iteratively estimated; (B) In the second step, users have the ability to select specific subgroups of cells whose transitions are of interest, fine-tune the previously trained neural network, identify features that are predictive of transition probabilities, and construct the corresponding GRNs.

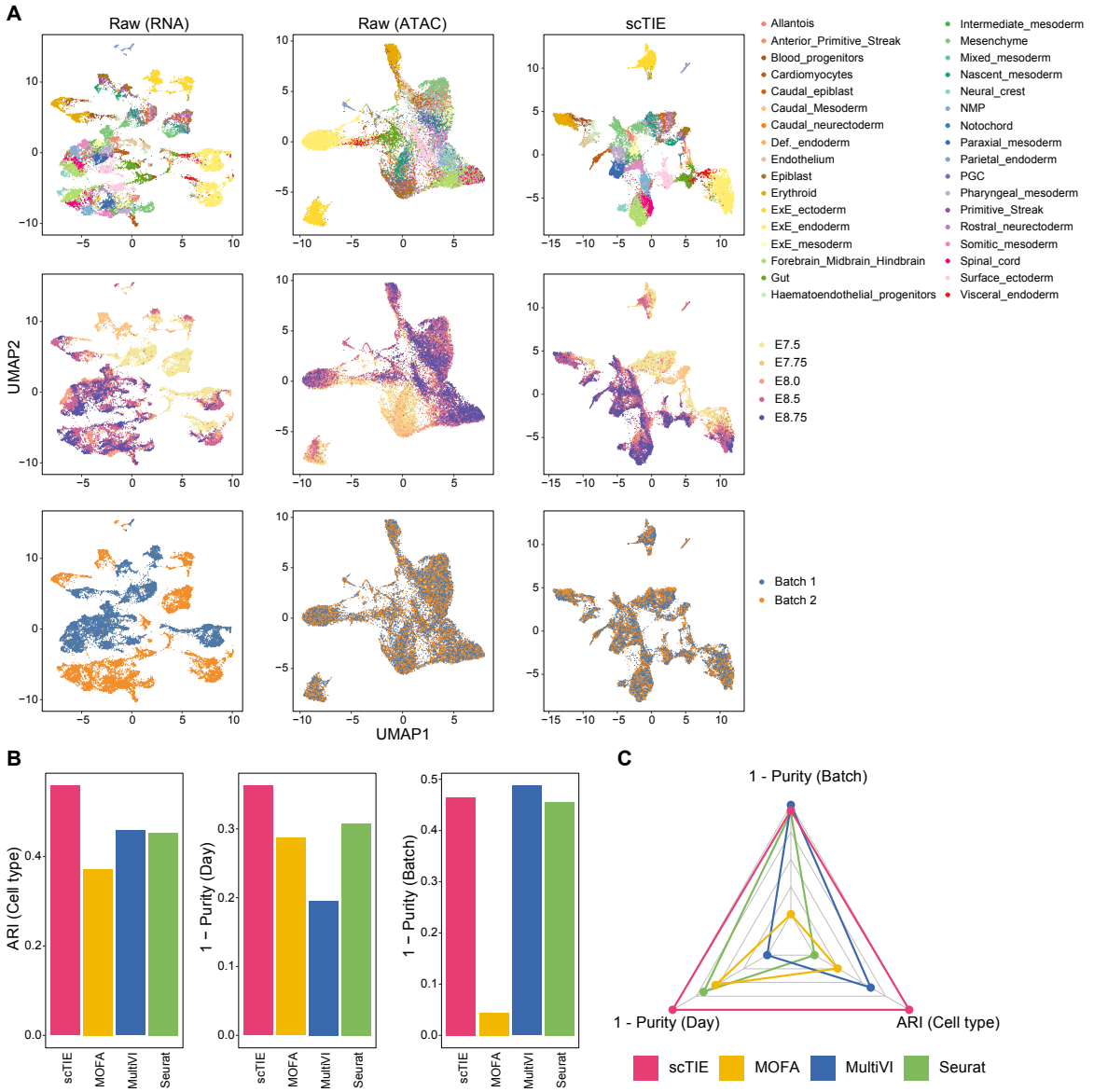


Figure 2: Performance benchmarking for integrating temporal multimodal data. (A) Joint visualization using UMAP of the synthetic dataset with batch effect in RNA and noise in ATAC, colored by cell type annotations (first row), sampling days (second row) and synthetic batch information (third row). Each dot represents a cell in the embedding space. (B) Bar plots showing the evaluation metrics of different data integration methods, including ARI values for clustering with annotations (left); 1 - average purity scores of sampling days with the number of neighbors equal to 50 (middle) and 1 - average purity scores of the synthetic batch with the number of neighbors equal to 50 (right). Higher values indicate better agreement with annotations and mixing of batches/days. (C) Radar plot summarizing the three evaluation metrics shown in (B), where each line represents the performance of one method, and each axis represents an evaluation metric, starting from the minimum value of all methods. It is noted that scAI was not included in this benchmarking due to its long computational time ($>$ 2 days).

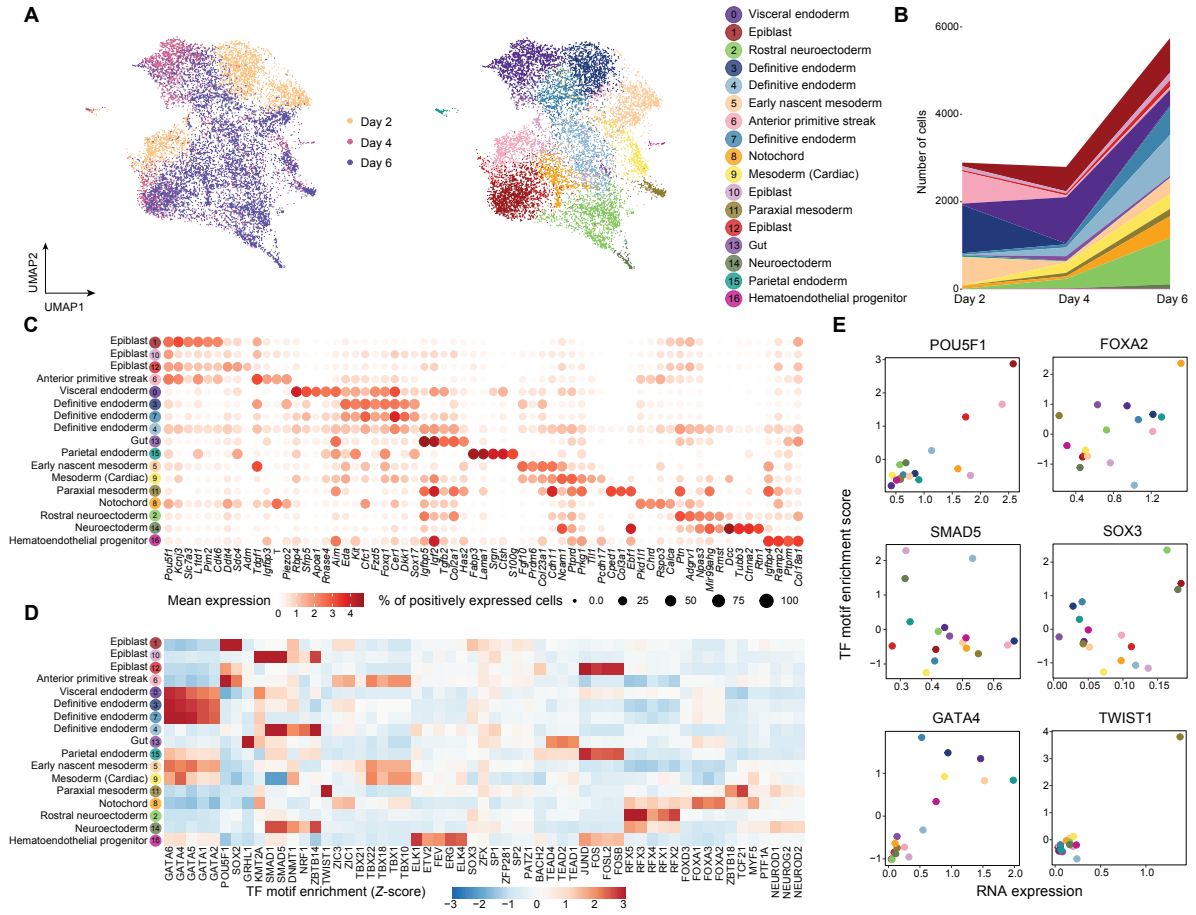


Figure 3: Integration and cell type identification of the mESC dataset by scTIE. (A) Joint visualization of the mESC dataset using UMAP, colored by sampling day and cell type annotations. Each dot represents a cell in the embedding space. (B) Cell type compositions per time point. (C) Dot plots of mean expression of RNA data. Rows represent cell types and columns indicate each genes. The color scale represents the expression level, and the size indicates proportion of positively expressed cells. The five most significantly expressed genes for each cluster are included. (D) Heatmap of the TF motif enrichment (Z-scores) of ATAC data. Rows represent cell types and columns indicate TFs. The five most significantly enriched TFs for each cluster are included. (E) Scatter plots of the mean RNA expression levels by clusters (x-axis) and the average TF motif enrichment scores of ATAC (y-axis) for the selected TFs. The dots are colored by the cell type annotations, with color legend consistent with Fig. 3A.

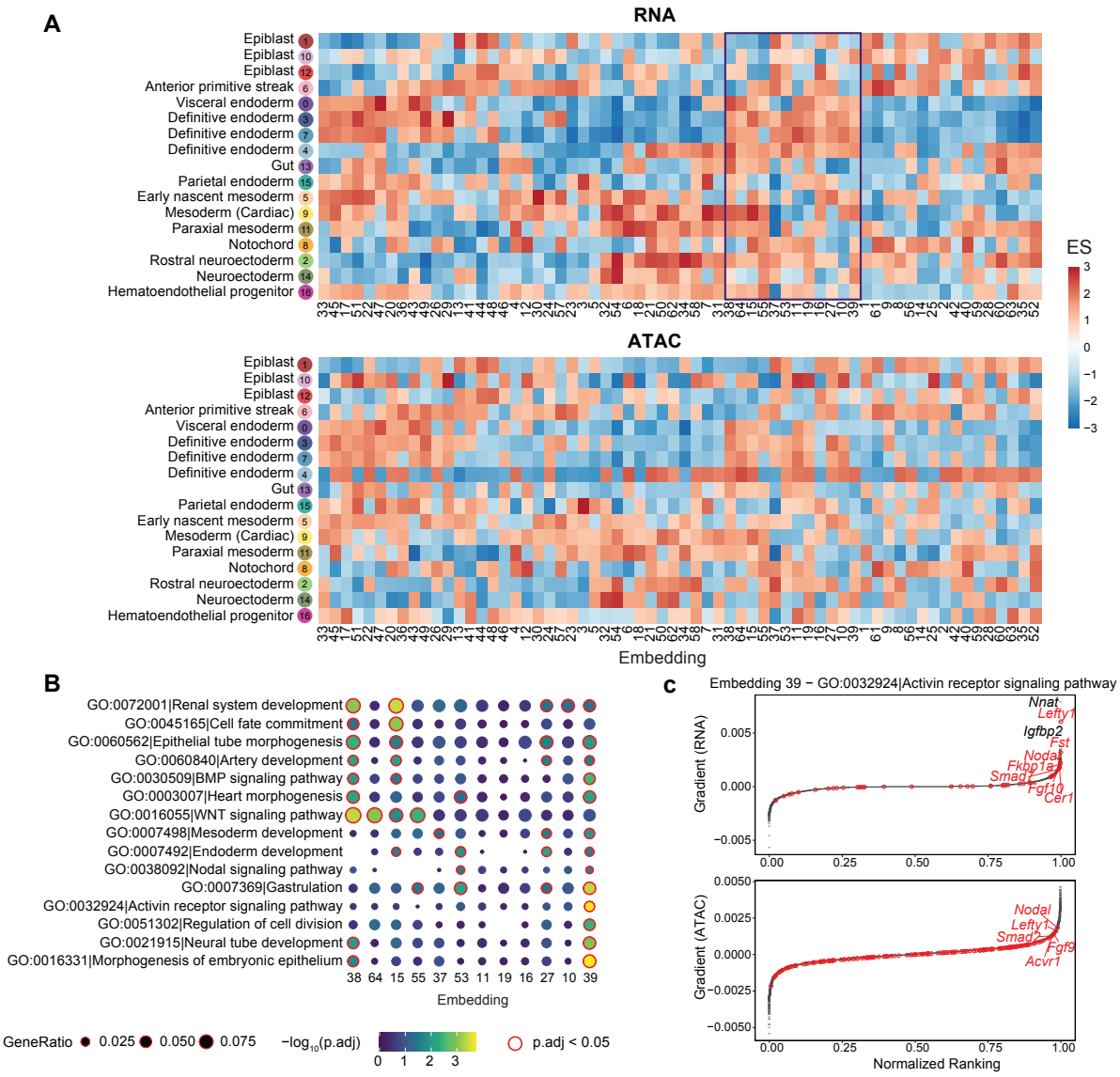


Figure 4: Biological signals in the mESC dataset captured by each embedding dimension of scTIE. (A) Enrichment scores of the gradient ranking in each embedding dimension using the RNA (top panel) and ATAC (bottom panel) marker list for each cell type. (B) Gene Ontology enrichment of selected pathways on the gradient ranking of a subset of embedding dimensions. (C) Gradient rankings for RNA (top panel) and ATAC (bottom panel) of embedding dimension 39, where genes/peaks are ranked based on the gradient values. The labeled points are genes in the selected gene set (Activin receptor signaling pathway).

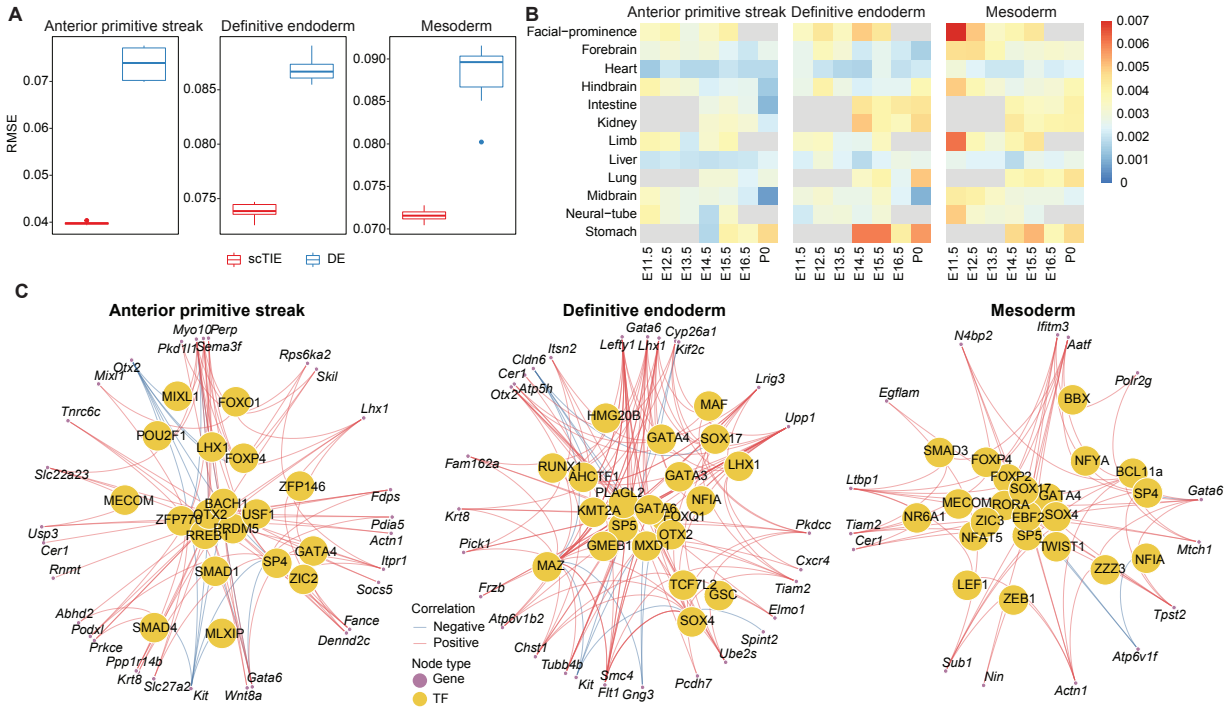


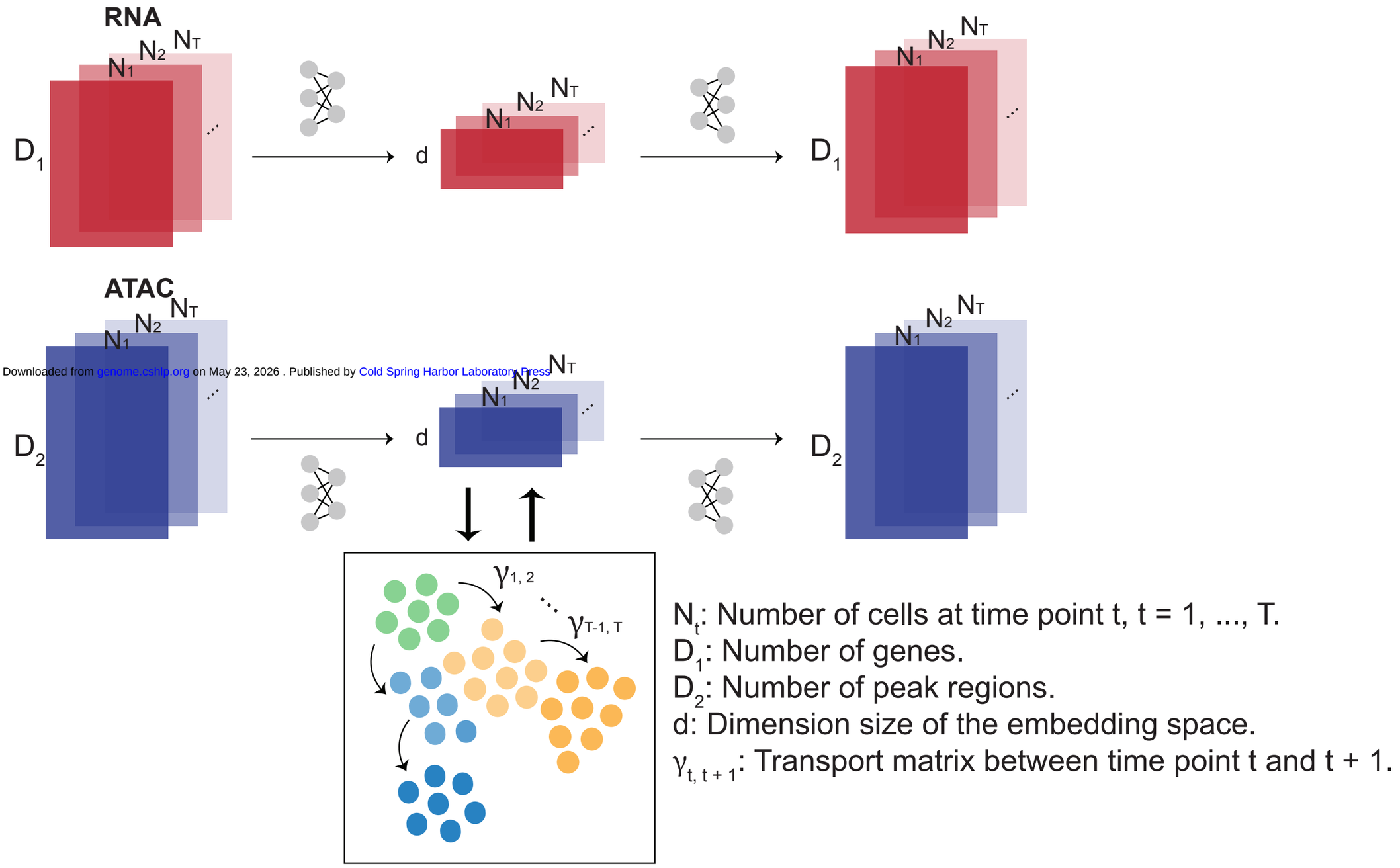
Figure 5: Lineage-specific regulatory elements selected by scTIE and the corresponding GRNs. (A) Performance of top genes and peaks selected by each method in predicting cell fate probabilities. (B) Similarity of top gradient peaks with enhancers of 12 tissues at seven developmental stages from known enhancer databases. (C) GRN of three cell fates.

Encoder
Batchnorm (26717)
Linear (26717, 1000)
Batchnorm (1000)
LeakyReLU (0.2)
Linear (1000, 1000)
Batchnorm (1000)
LeakyReLU (0.2)
Linear (1000, 64)
Decoder
Linear (64, 500)
Batchnorm (500)
LeakyReLU (0.2)
Linear (500, 1000)
Batchnorm (1000)
LeakyReLU (0.2)
Linear (1000, 26717)
Batchnorm (26717)
Multiply by σ and add μ

Encoder
Batchnorm (61744)
Linear (61744, 1000)
Batchnorm (1000)
LeakyReLU (0.2)
Linear (1000, 1000)
Batchnorm (1000)
LeakyReLU (0.2)
Linear (1000, 64)
Decoder
Linear (64, 500)
Batchnorm (500)
LeakyReLU (0.2)
Linear (500, 1000)
Batchnorm (1000)
LeakyReLU (0.2)
Linear (1000, 61744)

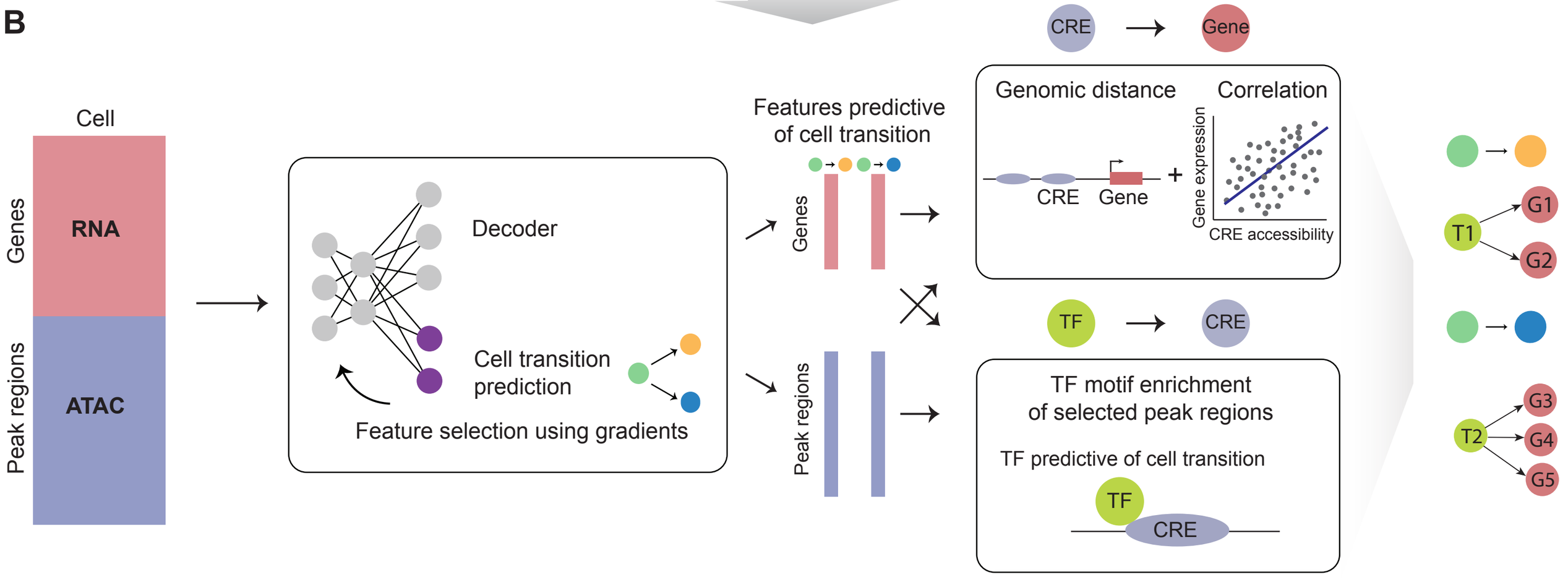
Table 1: Autoencoder architecture for RNA (left) and ATAC (right).

A



N_t: Number of cells at time point t, t = 1, ..., T.
D_1: Number of genes.
D_2: Number of peak regions.
d: Dimension size of the embedding space.
$\gamma_{t,\,t+1}$: Transport matrix between time point t and t + 1.

B



A

Raw (RNA)

Raw (ATAC)

scTIE

UMAP2

UMAP1

Allantois
Anterior_Primitive_Streak
Blood_progenitors
Cardiomyocytes
Caudal_epiblast
Caudal_Mesoderm
Caudal_neurectoderm
Def._endoderm
Endothelium
Epiblast
Erythroid
ExE_ectoderm
ExE_endoderm
ExE_mesoderm
Forebrain_Midbrain_Hindbrain
Gut
Haematoendothelial_progenitors
Intermediate_mesoderm
Mesenchyme
Mixed_mesoderm
Nascent_mesoderm
Neural_crest
NMP
Notochord
Paraxial_mesoderm
Parietal_endoderm
PGC
Pharyngeal_mesoderm
Primitive_Streak
Rostral_neurectoderm
Somitic_mesoderm
Spinal_cord
Surface_ectoderm
Visceral_endoderm

E7.5
E7.75
E8.0
E8.5
E8.75

Batch 1
Batch 2

B

ARI (Cell type)

1 – Purity (Day)

1 – Purity (Batch)

scTIE
MOFA
MultiVI
Seurat

C

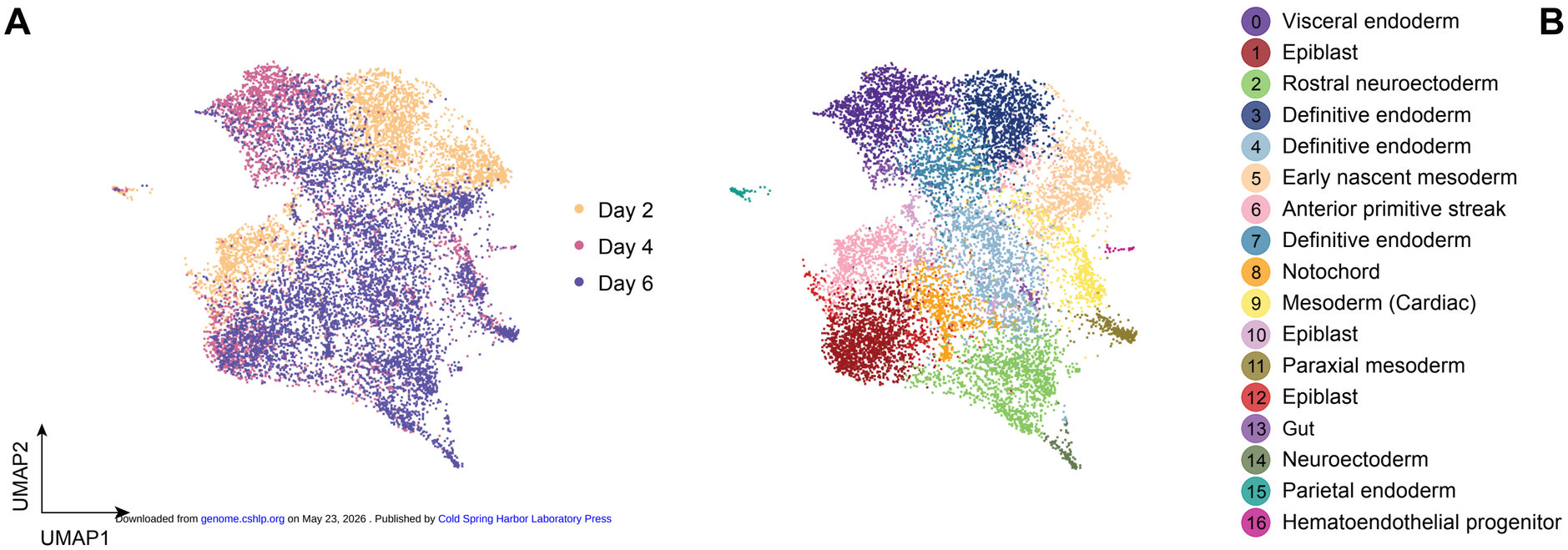
1 - Purity (Batch)

1 - Purity (Day)

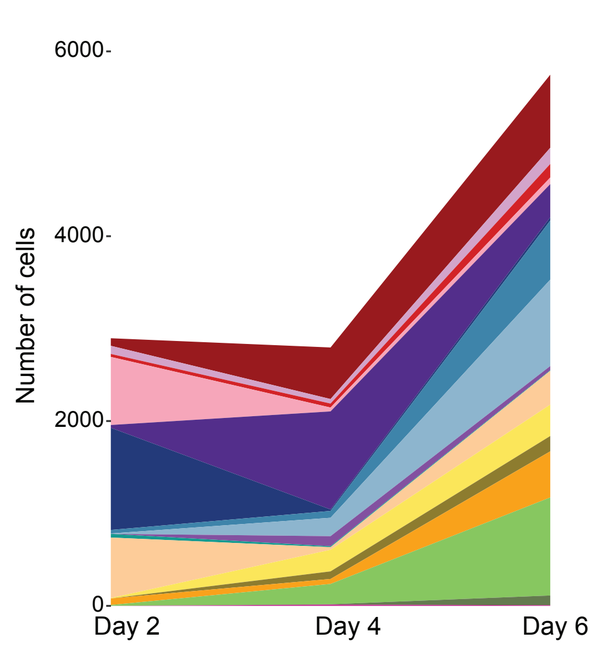
ARI (Cell type)

scTIE
MOFA
MultiVI
Seurat

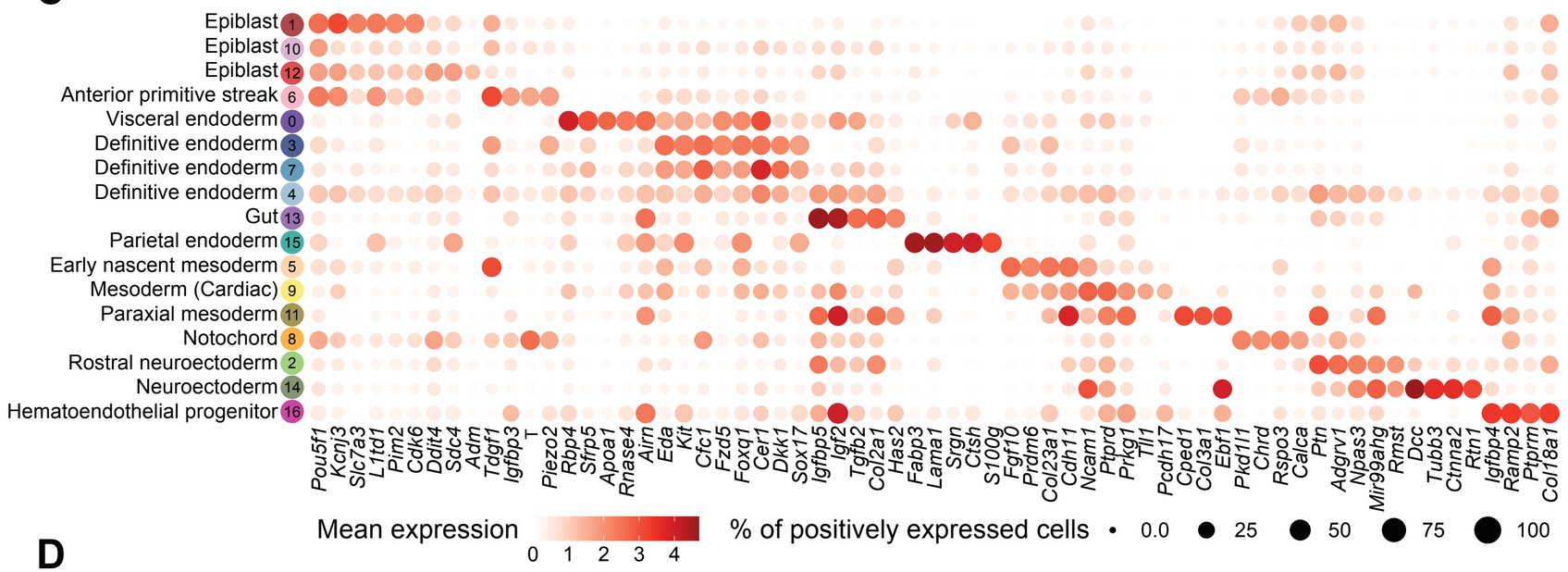
A



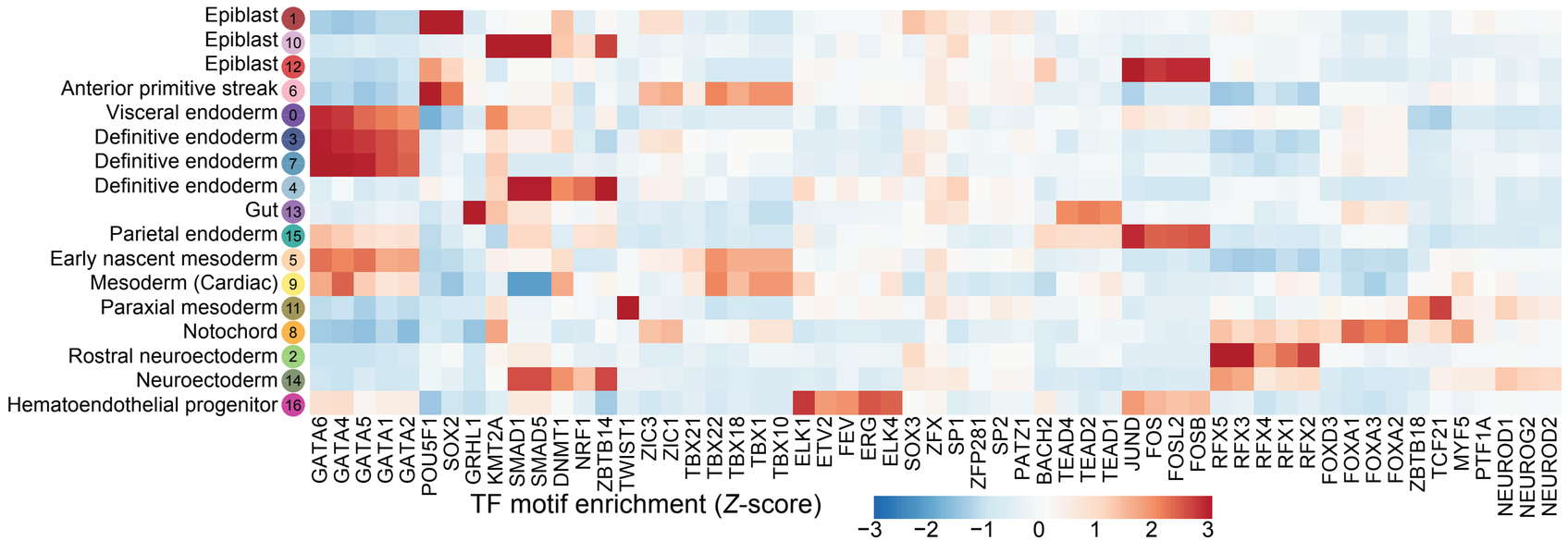
B



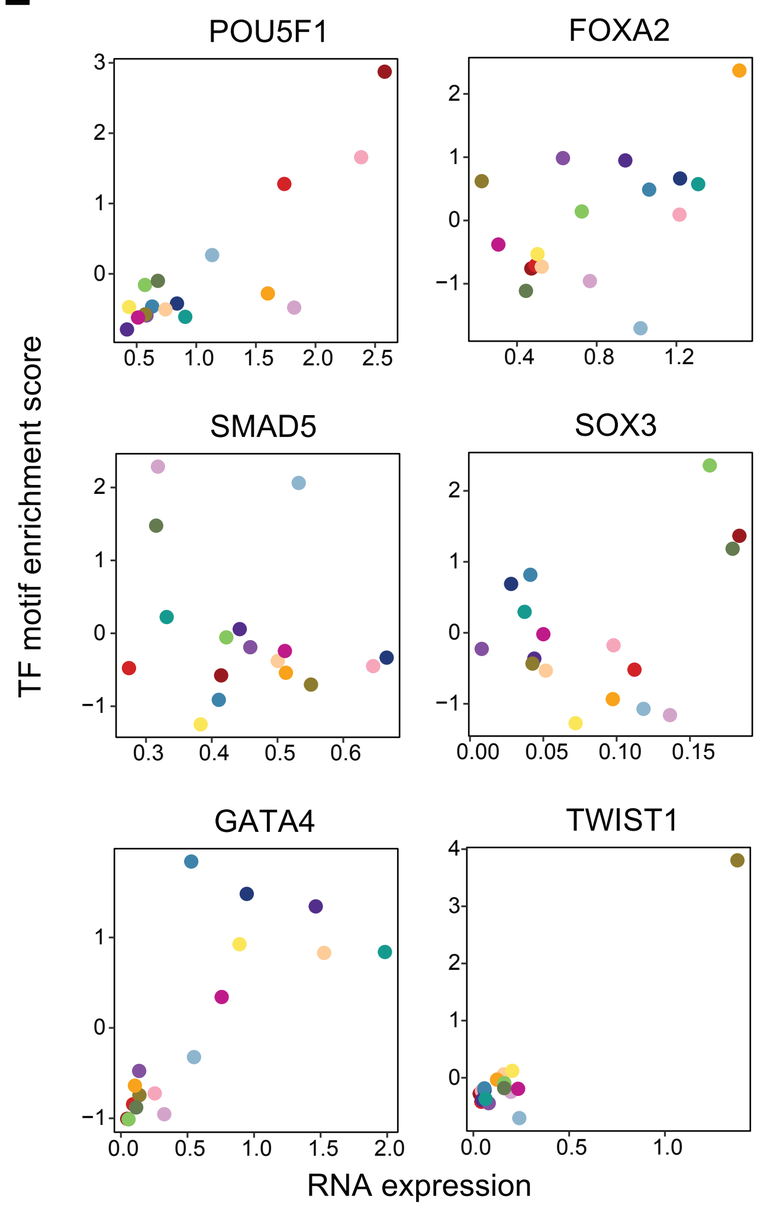
C



D



E



A

RNA

ATAC

Epiblast 1
Epiblast 10
Epiblast 12
Anterior primitive streak 6
Visceral endoderm 0
Definitive endoderm 3
Definitive endoderm 7
Definitive endoderm 4
Gut 13
Parietal endoderm 15
Early nascent mesoderm 5
Mesoderm (Cardiac) 9
Paraxial mesoderm 11
Notochord 8
Rostral neuroectoderm 2
Neuroectoderm 14
Hematoendothelial progenitor 16

33 45 17 51 22 47 20 36 43 49 26 29 13 41 44 48 46 4 12 30 24 57 23 3 5 32 54 6 18 21 50 62 34 58 7 31 38 64 15 55 37 53 11 19 16 27 10 39 1 61 9 8 56 14 25 2 42 40 59 28 60 63 35 52

Embedding

ES
3
2
1
0
−1
−2
−3

B

GO:0072001|Renal system development
GO:0045165|Cell fate commitment
GO:0060562|Epithelial tube morphogenesis
GO:0060840|Artery development
GO:0030509|BMP signaling pathway
GO:0003007|Heart morphogenesis
GO:0016055|WNT signaling pathway
GO:0007498|Mesoderm development
GO:0007492|Endoderm development
GO:0038092|Nodal signaling pathway
GO:0007369|Gastrulation
GO:0032924|Activin receptor signaling pathway
GO:0051302|Regulation of cell division
GO:0021915|Neural tube development
GO:0016331|Morphogenesis of embryonic epithelium

38 64 15 55 37 53 11 19 16 27 10 39

Embedding

GeneRatio 0.025 0.050 0.075

$-\log_{10}$(p.adj) 0 1 2 3

p.adj < 0.05

C

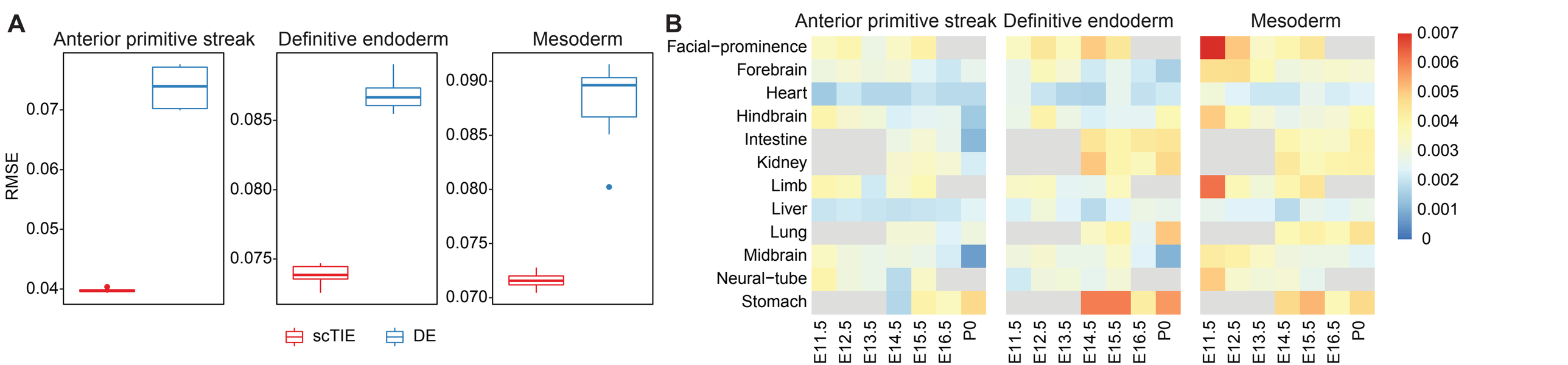
Embedding 39 − GO:0032924|Activin receptor signaling pathway

Gradient (RNA)
0.005
0.000
−0.005
Nnat
Lefty1
Igfbp2
Fst
Nodal
Fkbp1a
Smad7
Fgf10
Cer1

Gradient (ATAC)
0.0050
0.0025
0.0000
−0.0025
−0.0050
Nodal
Lefty1
Smad2
Fgf9
Acvr1

0.00 0.25 0.50 0.75 1.00

Normalized Ranking


A
Anterior primitive streak
Definitive endoderm
Mesoderm
RMSE
0.07
0.06
0.05
0.04
0.085
0.080
0.075
0.090
0.085
0.080
0.075
0.070
scTIE
DE
B
Anterior primitive streak
Definitive endoderm
Mesoderm
Facial−prominence
Forebrain
Heart
Hindbrain
Intestine
Kidney
Limb
Liver
Lung
Midbrain
Neural−tube
Stomach
E11.5
E12.5
E13.5
E14.5
E15.5
E16.5
P0
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0
C
Anterior primitive streak
Definitive endoderm
Mesoderm
Myo10
Perp
Pkd1l1
Sema3f
Rps6ka2
Otx2
Mixl1
Skil
MIXL1
FOXO1
Tnrc6c
Lhx1
POU2F1
LHX1
FOXP4
Slc22a23
ZFP146
MECOM
BACH1
Fdps
OTX2
USF1
ZFP770
PRDM5
Pdia5
Usp3
RREB1
Actn1
Cer1
GATA4
Itpr1
Rnmt
SP4
SMAD1
ZIC2
Socs5
Abhd2
Fance
Podxl
SMAD4
MLXIP
Dennd2c
Prkce
Ppp1r14b
Krt8
Slc27a2
Kit
Wnt8a
Gata6
Gata6
Cyp26a1
Lefty1
Lhx1
Kif2c
Itsn2
Cldn6
Cer1
Otx2
Atp5h
Lrig3
HMG20B
MAF
Upp1
GATA4
SOX17
Fam162a
RUNX1
LHX1
AHCTF1
GATA3
PLAGL2
NFIA
KMT2A
GATA6
Krt8
FOXQ1
Pkdcc
SP5
OTX2
Pick1
GMEB1
MXD1
Cxcr4
MAZ
Tiam2
Frzb
TCF7L2
GSC
Elmo1
Atp6v1b2
SOX4
Spint2
Chst1
Ube2s
Tubb4b
Pcdh7
Smc4
Kit
Flt1
Gng3
Correlation
Negative
Positive
Node type
Gene
TF
Ifitm3
N4bp2
Aatf
Egflam
Polr2g
BBX
SMAD3
FOXP4
NFYA
BCL11a
Ltbp1
FOXP2
SOX17
SP4
Gata6
GATA4
MECOM
RORA
NR6A1
ZIC3
EBF2
SOX4
Tiam2
NFAT5
SP5
Mtch1
Cer1
TWIST1
NFIA
ZZZ3
LEF1
ZEB1
Tpst2
Sub1
Atp6v1f
Nin
Actn1