KD-CPT: A Knowledge-Driven Cellular Phenotypic Transdifferentiation Model

Lei Xin [*1,2] **Zhenglun Kong** [*3] **Xiaoshuo Yan** [4] **Sijia Yan** [5] **Zeheng Wang** [2] **Wuzhe Fan** [6] **Hao Tang** [1]

Abstract

Cell phenotype transition refers to the dynamic changes in cellular morphology, function, and marker expression under specific environmental or physiological conditions, driven by genomic information and external signals. It plays a key role in development, tissue repair, and immune responses. Traditional approaches, often hypothesis-driven, struggle to capture the inherent complexity and heterogeneity of these processes. We propose KD-CPT, a Markov process-based model for cell phenotype transition and differentiation, comprising two branches: a prediction branch for phenotype classification, evaluated via a multi-metric framework, and a screening branch that identifies key regulatory genes using a token pruning strategy. An enhanced multi-head attention mechanism is employed to strengthen information flow between full-sequence contexts and prioritized regulatory loci. The model further demonstrates strong performance in uncertainty quantification and confidence calibration. Gene knockout experiments reveal that disruption of critical genes significantly alters transition probabilities and can even terminate specific transition pathways, highlighting the model's utility in uncovering regulatory mechanisms.

1. Introduction

Traditional bioinformatics methods for addressing cell type transition often rely on hypothesis-driven models (Ritchie et al., 2015), which may not fully capture the complexity and heterogeneity of the transition processes (Eraslan et al., 2019). The rise of machine learning, particularly deep learning techniques (Vaswani et al., 2017; Wang et al., 2024) in recent years, has provided new opportunities to develop more accurate models of cell type transition (Wang et al., 2021). By learning potential patterns from large-scale single-cell data (He et al., 2020), machine learning methods can reveal interactions and dynamic changes between cells (You et al., 2020), thereby enhancing our understanding of the mechanisms underlying cell type transitions (Zhang et al., 2022).

To address the challenges associated with key gene identification and information integration in cellular phenotype transitions, we propose a biologically informed foundation model, KD-CPT. It leverages attention mechanisms alongside an enhanced multi-head attention module to integrate full-sequence contextual information with key regulatory loci, enabling sparse yet interpretable gene prioritization. An adaptive feature recalibration mechanism is further employed to dynamically fuse global transcriptomic patterns with localized regulatory signals. To evaluate the effectiveness of KD-CPT, we conduct experiments on datasets comprising both cancer and control groups. Given the non-directional nature of phenotype transitions and the intrinsic uncertainty of biological systems, we introduce a cosine similarity-based uncertainty index to quantify model reliability. Moreover, we design gene perturbation experiments to assess the model's robustness. Our work offers new insights into cellular dynamics and contributes to methodological advancements in the biomedical domain, promoting progress in health and disease management while underscoring the principles of scientific rigor and collaboration.

Our contribution is summarized as follows: i) **Paradigm shift:** For the first time, we have considered the regulatory roles of critical loci and developed a biologically-informed, inherently interpretable dual-branch architecture, ensuring that locus discovery extends beyond post-hoc interpretable validation. ii) **Cell selection module:** We designed a token selection module based on genetic regulatory mechanisms to identify key genes. iii) **Global-local information fusion:** We have designed an enhanced multi-head attention mechanism that better integrates global and local information by processing the global features from the prediction branch through 1×1 convolutions and the local features from the screening branch through 7×7 convolutions, before feeding them into the multi-head attention module.

2. Preliminary

In this section, we introduce a mathematical framework aimed at modeling cellular phenotype transitions, which is an essential biological process underlying differentiation,

[*]Equal contribution [1]State Key Laboratory of Multimedia Information Processing, School of Computer Science, Peking University [2]Wuhan University [3]Havard University [4]Shandong University [5]Tsinghua University [6]University of Sydney. Correspondence to: Hao Tang <haotangpku.edu.cn>.

Proceedings of the 41st International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

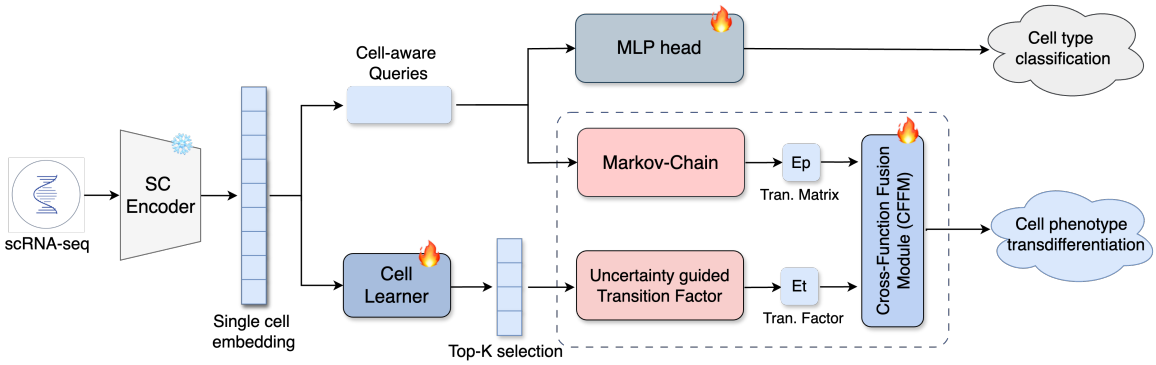

Figure 1. The overall pipeline of KD-CPT. The prediction branch performs sparse sequence representation compression based on bipartite graph matching, while the transition branch models the modulation effect through a token selection module.

reprogramming, and disease progression. Our goal is to predict changes in cell identity by integrating global expression profiles with localized regulatory signals derived from single-cell sequencing data.

Problem Formulation. Let $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ denote the set of gene expression profiles from n single cells, where each profile $\mathbf{x}_i \in \mathbb{R}^m$ represents the expression levels of m genes. We define two phenotypic states, $\mathcal{C}_1$ and $\mathcal{C}_2$, corresponding to distinct cellular identities. Each cell i is associated with a binary phenotype label $y_i \in \{0, 1\}$, where $y_i = 0$ indicates that the cell belongs to phenotype $\mathcal{C}_1$ and $y_i = 1$ indicates phenotype $\mathcal{C}_2$.

Full Sequence Modeling. To capture the underlying biological variability, we transform the raw gene expression data into a latent representation

$$\mathbf{Y} = f(\mathbf{X}; \theta_1), \tag{1}$$

where $f : \mathbb{R}^m \to \mathbb{R}^m$ is a learnable mapping model, θ_1 are the parameters of the model. The prediction for each cell is obtained by applying a classification function g:

$$\hat{y}_i^{\text{full}} = g(\mathbf{Y}_{\text{full},i}; \theta_g), \tag{2}$$

where g is typically a softmax function or a logistic regression model, and θ_g are the parameters for this classification.

Quantifying Regulatory Influence. A crucial aspect of phenotype transition is the modulation of gene expression by key regulatory genes. For each target gene g, we quantify its regulation by aggregating contributions from a selected set of k candidate regulatory genes. Let $\mathcal{R} = \{r_1, r_2, \ldots, r_k\}$ denote these regulators. We model the regulatory influence on gene g as:

$$\mathbf{R}_g = \sum_{j=1}^{k} w_j \cdot \phi(r_j), \tag{3}$$

where $\phi(r_j)$ is a function that captures the regulatory activity of gene r_j, w_j are learnable weights that quantify the strength of influence of each regulator on the target gene. This effectively integrates the diverse regulatory signals into a single measure of influence for each gene.

The output can be expressed as:

$$\mathbf{Y}_{\text{reg}} = f_{\text{reg}}(\mathbf{R}, \mathbf{X}; \theta_2), \tag{4}$$

where f_{reg} is a function that combines regulatory information with the original gene expression profiles, parameterized by θ_2.

3. Methodology

We design a dual-branch framework to capture both the deterministic patterns in gene expression and the regulatory influences that drive phenotype transitions discussed in Sec. 2. As shown in Figure 1, our model consists of two distinct branches: 1) *Prediction Branch* for capturing the full expression profiles of the cells; 1) *Selection Branch* model regulatory influences through a top-k selection mechanism. KD-CPT performs multi-omics sequence modeling by leveraging token pruning to efficiently integrate and process both scRNA-seq and regulatory site information. The dual-branch architecture is designed to support this integration, enabling both global expression modeling and targeted regulatory analysis.

3.1. Uncertainty-based Cellular Phenotype Prediction Branch

As shown in Figure 1, the feature representations of cells obtained from embeddings generated by a pretrained single-cell model capture the information about cell phenotype in various biological environments.

Specifically, we process single-cell RNA expression data through a pretrained single-cell model to obtain latent representations, which are subsequently fed into a multilayer perceptron (MLP) network for classification. The MLP architecture consists of three fully connected layers with ReLU activation functions, where the forward propagation

process can be formulated as:

$$h^{(l)} = \sigma\left(\mathbf{W}^{(l)} h^{(l-1)} + \mathbf{b}^{(l)}\right), \quad l \in \{1, 2, 3\} \tag{5}$$

where $h^{(l)}$ denotes the output of the l-th layer, $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$ represent the weight matrix and bias vector respectively, and $\sigma(\cdot)$ is the ReLU activation $\sigma(x) = \max(0, x)$.

3.2. Cell Learner

Token reduction reduces computational load by pruning unimportant tokens (Kong et al., 2025), thereby facilitating the processing of long input sequences. Previous methods (Bolya et al., 2022; Kong et al., 2022) were not specifically designed for modeling cellular phenotypic transdifferentiation (Lotfollahi et al., 2021). We proposed a Cell Learner module that identifies critical regulatory sites using attention scores from the Transformer encoder, where the self-attention mechanism captures sequence dependencies and aggregates scores across layers/heads (via mean or weighted summation), highlighting biologically significant regions. This aligns with the theory that attention weights reflect context-aware feature importance, and selected sites guide cellular transitions via learnable parameters. We determine Top-*k* tokens by integrating domain prior knowledge (e.g., expected transcription factor binding sites in genomics) rather than pure data optimization, a hybrid approach balancing biological plausibility and model flexibility supported by prior-guided sparse attention studies.

3.3. Regulatory Gene Screening Branch

In the process of cellular phenotype transition modeling, we assess the uncertainty of the trained Markov transition probability matrix using the information entropy of each category and the overall entropy.

Information Entropy is used to quantify the uncertainty of a random variable. Given the transition probability matrix P for a specific category C_i, the information entropy $H(C_i)$ can be calculated as follows:

$$H(C_i) = -\sum_{j=1}^{n} P_{ij} \log(P_{ij}) \tag{6}$$

where P_{ij} represents the probability of transitioning from state C_i to state C_j, and n is the total number of states. A higher information entropy indicates greater uncertainty in the state transitions of that category.

Overall Entropy is computed as a weighted average of the information entropy across all categories, reflecting the overall uncertainty of the system. The overall entropy H_{total} is calculated as follows:

$$H_{total} = -\sum_{i=1}^{m} \pi_i H(C_i) \tag{7}$$

where π_i is the prior probability of category C_i, and m is the total number of categories. This overall entropy provides a global perspective on the cell state transition process, revealing the uncertainty of transitions across different cellular phenotypes.

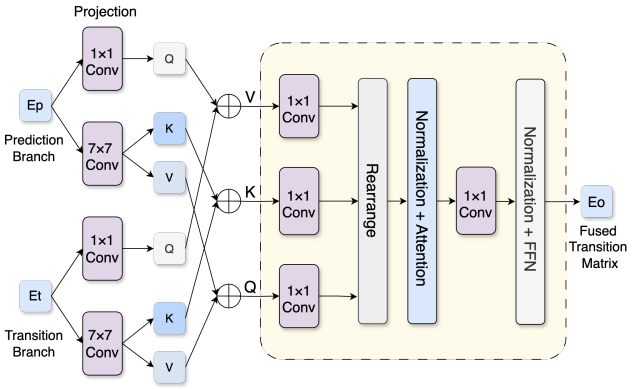

Figure 2. The overall pipeline of the proposed cross-function fusion module.

3.4. Cross-Function Fusion Module (CFFM)

We perform a fusion process that combines the transition matrix derived from the Markov chain with the Transition Factor, which is computed based on probability and uncertainty. As shown in Figure 2, the output embeddings from the two branches are denoted as $E_p \in \mathbb{R}^{d \times c}$ and $E_t \in \mathbb{R}^{d \times c}$, where d is the sequence length and c is the channel dimension. For multi-head attention, the channel dimension c can be divided into n subspaces. We first generate query/key/value projections for both branches:

$$Q_i = E_i W_q^i, \quad K_i = E_i W_k^i, \quad V_i = E_i W_v^i, \quad i \in \{p, t\}, \tag{8}$$

where $\{W_q^i, W_k^i, W_v^i\}$ are learnable projection matrices for each branch i. We use 1x1 Convolution to extract global features and 7x7 Convolution to extract local features of the input. We concatenate the projections and use cross-attention for fusion:

$$\begin{cases} Q = \text{Concat}(Q_p, Q_t), \\ K = \text{Concat}(K_p, K_t), \\ V = \text{Concat}(V_p, V_t) \end{cases} \tag{9}$$

The normalization operation is implemented through layer-wise transformations: $\bar{Q} = \text{LayerNorm}(Q)$, $\bar{K} = \text{LayerNorm}(K)$.

The attention score A can be calculated by:

$$A = \text{softmax}\left(\frac{\bar{Q}\bar{K}^\top}{\sqrt{c/n}}\right). \tag{10}$$

The fused representation is then obtained through value aggregation and the following feedforward layer. Finally,

Table 1. Ablation study of the proposed method.

Model	CFFM	Cell Learner	Confidence	Uncertainty	Accuracy	F1	Kappa
KD-CPT	×	×	89.16	28.52	90.24	89.89	88.36
	✓	×	89.56	28.01	90.83	90.89	89.91
	×	✓	90.53	23.43	90.53	90.42	88.76
	✓	✓	**91.82**	**21.28**	**91.60**	**91.43**	**89.94**

Table 2. Comparison of cellular phenotype classification.

Baseline	Accuracy	F1	Kappa
ResNet	88.82	88.68	86.79
Res2Net	88.95	88.79	86.94
ConNeXt	88.81	88.66	86.79
Informer	88.90	88.74	86.91
Metaformer	89.63	89.55	87.78
scBERT	45.16	42.13	36.59
scFoundation	90.13	88.31	86.53
Ours	**91.60**	**91.43**	**89.94**

Table 3. Comparison of XAI baseline

Baseline	Confidence	Uncertainty
Feature Importance	90.12	29.37
LIME	89.27	28.34
Diffrate	90.33	27.31
Token Learner	90.62	25.60
Ours	**91.82**	**21.28**

the output E_o is reconstructed back to the original spatial dimension through transposed convolution and projection. This architecture enables dynamic information flow between the two branches while preserving gradient stability through residual connections.

4. Experiments

4.1. Experimental Setup

Datasets. We utilize public datasets comprising over 590,000 samples (Dann et al., 2023) for single-cell foundation model training. Additionally, we employ three cell annotation datasets from scGPT (Cui et al., 2024) to perform transfer learning and conduct downstream tasks.

Baselines. We compare with multiple deep learning models including: 1) CNN-based models – ResNet (He et al., 2016), Res2Net (Gao et al., 2019), ConvNeXt (Liu et al., 2022); 2) Transformer-based models – Transformer (Vaswani et al., 2017), Informer (Zhou et al., 2021), Metaformer citeyu2022metaformer; 3) Specific Model – scBERT (Yang et al., 2022) and scGPT (Cui et al., 2024). Each model undergoes the same testing phase to ensure a consistent and fair comparison of their capabilities within the legal domain.

Evaluation Metrics. In the prediction branch, we evaluate the model's accuracy in cell phenotype prediction based on Accuracy, F1, and Kappa coefficient; in the screening branch, we assess the feasibility of the screening process based on accuracy, uncertainty, and confidence. Additionally, we optimize the confidence calculation process based on Top-K.

Implementation Details. We conducted the training of our model over 10 epochs using two L20 GPUs. Additionally, we configured the model's dropout rate at 0.1 and set the learning rate to e^{-3}.

4.2. Comparison Results

We compared the performance of six mainstream baseline models within the KD-CPT architecture, including ResNet, Res2Net, ConvNext, Transformer, Informer, and Metaformer. According to the experimental results in Table 2, we exhibit the best performance. Table 3 summarizes results from the explainable AI (XAI) baseline comparison. Our KD-CPT approach achieves the highest confidence (91.82%) and lowest uncertainty (21.28%) scores compared to conventional XAI methods.

4.3. Ablation Study

We conduct ablation experiments to evaluate the rationality of the screening branch and the design of CFFM. We compare four scenarios: with/without the CFFM module and with/without the Cell Learner module. As shown in Table 1, the ablation results demonstrate that the introduction of both the Cell Learner module and the CFFM improves the accuracy of cell phenotype prediction. These modules significantly reduce the model's uncertainty while causing only a minor performance decline.

5. Conclusion

This paper proposes a novel framework, KD-CPT, whose innovation lies in its ability to incorporate regulatory sites as critical biological priors during the training process, effectively circumventing biases introduced by traditional manual design approaches. Additionally, we have designed a global-local aware cross-attention module to fuse information from two modalities: the full sequence and regulatory sites. Extensive transfer learning experiments and gene knockout studies demonstrate that the KD-CPT model exhibits high robustness and accuracy.

References

Bolya, D., Fu, C.-Y., Dai, X., Zhang, P., Feichtenhofer, C., and Hoffman, J. Token merging: Your vit but faster. *arXiv preprint arXiv:2210.09461*, 2022.

Cui, H., Wang, C., Maan, H., Pang, K., Luo, F., Duan, N., and Wang, B. scgpt: toward building a foundation model for single-cell multi-omics using generative ai. *Nature Methods*, 21(8):1470–1480, 2024.

Dann, E. et al. Precise identification of cell states altered in disease using healthy single-cell references. *Nature Genetics*, 55(11):1998–2008, 2023.

Eraslan, G. et al. Deep learning: New computational modelling techniques for single-cell RNA-seq analysis. *Nature Reviews Genetics*, 20:296–312, 2019.

Gao, S.-H., Cheng, M.-M., Zhao, K., Zhang, X.-Y., Yang, M.-H., and Torr, P. Res2net: A new multi-scale backbone architecture. *IEEE transactions on pattern analysis and machine intelligence*, 43(2):652–662, 2019.

He, D. et al. A geometric deep learning framework for drug discovery and cellular response prediction. *Nature Communications*, 11(1):1–14, 2020.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Kong, Z., Dong, P., Ma, X., Meng, X., Niu, W., Sun, M., Shen, X., Yuan, G., Ren, B., Tang, H., et al. Spvit: Enabling faster vision transformers via latency-aware soft token pruning. In *European conference on computer vision*, pp. 620–640. Springer, 2022.

Kong, Z., Li, Y., Zeng, F., Xin, L., Messica, S., Lin, X., Zhao, P., Kellis, M., Tang, H., and Zitnik, M. Token reduction should go beyond efficiency in generative models–from vision, language to multimodality. *arXiv preprint arXiv:2505.18227*, 2025.

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., and Xie, S. A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11976–11986, 2022.

Lotfollahi, M. et al. Learning interpretable cellular responses to complex perturbations. *Nature Methods*, 18: 1183–1190, 2021.

Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7):e47, 2015.

Vaswani, A. et al. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pp. 5998–6008, Long Beach, CA, 2017. Curran Associates, Inc.

Wang, J. et al. Self-supervised contrastive learning for integrative single cell RNA-seq data analysis. *Bioinformatics*, 37(Supplement_1):i343–i351, 2021.

Wang, X., Chen, X., Ren, W., Han, Z., Fan, H., Tang, Y., and Liu, L. Compensation atmospheric scattering model and two-branch network for single image dehazing. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2024.

Yang, F., Wang, W., Wang, F., Fang, Y., Tang, D., Huang, J., Lu, H., and Yao, J. scbert as a large-scale pretrained deep language model for cell type annotation of single-cell rna-seq data. *Nature Machine Intelligence*, 4(10):852–866, 2022.

You, J. et al. Graph structure of neural networks. In *International Conference on Machine Learning*, pp. 10881–10891. PMLR, 2020.

Zhang, Z. et al. Deep learning in single-cell analysis. *ACM Transactions on Intelligent Systems and Technology*, 13 (4):1–25, 2022.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., and Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 11106–11115, 2021.