DualCPT: Dual-branch Modeling for Cellular Phenotype Transition

Lei Xin[2*], Zhenglun Kong[3*], Fukang Chen[2], Yuhao Zheng[4] Zeheng Wang[2], Hao Tang[1†]

[1]Peking University [2]Wuhan University [3]Harvard University [4]University of Science and Technology of China
[†]corresponding author, haotang@pku.edu.cn

Abstract

Cell phenotype transition refers to the changes in the morphology, function, and surface markers of cells that occur under specific environmental conditions or physiological states, based on their genomic information and external signals. This process plays an important role in development, tissue repair, and responses to external stimuli such as infection or inflammation. Traditional bioinformatics methods for addressing cell type transition often rely on hypothesis-driven models, which may not fully capture the complexity and heterogeneity of the transition processes. In this paper, we introduce DualCPT, a cell phenotype transition and differentiation model based on Markov processes. Specifically, the model consists of a classification branch and a transition branch. The transition branch identifies regulatory genes involved in cell phenotype transition and differentiation. In the classification branch, we evaluate the model's overall performance on general cell type classification tasks using a comprehensive multi-metric evaluation framework; in the transition branch, we implement a token pruning-based approach for critical locus discovery and enhance information interaction between full-sequence contexts and prioritized regulatory sites via an improved multi-head attention mechanism. Cell phenotype transition tasks are further assessed by uncertainty quantification and confidence calibration. In particular, in gene knockout experiments, we found that knocking out important genes alters the probability of cell phenotype transition and differentiation, and knocking out a certain number of essential genes can terminate specific transition processes. Data, code, and checkpoints are publicly available at https://github.com/Ssupercoder/DualCPT.

Introduction

Cell phenotype transition refers to the changes in the morphology, function, and surface markers of cells that occur under specific environmental conditions or physiological states, based on their genomic information and external signals. This process plays a crucial role in various biological phenomena, including development, tissue repair, and responses to external stimuli such as infection or inflammation (Liu, Lou, and Xie 2019). Cell phenotype transition is not only a fundamental mechanism for organisms to adapt to environmental changes (Trapnell et al. 2014), but it is also a key factor in the occurrence and progression of many diseases. For example, in cancer, tumor cells often undergo phenotype transitions that grant them increased invasiveness and metastatic capabilities (Hanahan and Weinberg 2011). This transition enables cancer cells to evade the surveillance of the host immune system and grow into new tumors in other tissues (Lambert, Pattabiraman, and Weinberg 2017). Additionally, in fibrotic diseases such as pulmonary fibrosis, normal cells can transition into myofibroblasts under specific stimuli (Rockey et al. 2015), leading to excessive extracellular matrix deposition that ultimately impairs organ function. Therefore, understanding the mechanisms of cell phenotype transition is vital for revealing the biological basis of disease development and for identifying potential therapeutic targets (Stuart et al. 2019).

Traditional bioinformatics methods for addressing cell type transitions 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). With the support of machine learning methods, modeling cell phenotype transitions can help scientists gain a better understanding of this complex biological process. By analyzing single-cell sequencing data (Wolf, Angerer, and Theis 2018), researchers can identify expression patterns of key genes involved in phenotype transitions and their regulatory networks across different cellular states (Arlotta and Berninger 2012). Methods such as scBERT (Yang et al. 2022) and scGPT (Cui et al. 2024) have employed transformer architectures, demonstrating enhanced capabilities in capturing complex cellular states and phenotypic transitions by leveraging attention mechanisms to interpret high-dimensional transcriptomic data. Such modeling not only reveals how cells undergo phenotype transitions in varying environments (Efremova et al. 2020), but also predicts cellular behaviors under certain conditions (Lotfollahi et al. 2019), thereby providing new insights for disease prevention and treatment.

However, current approaches still struggle to fully capture the complexity of the biological processes underlying phenotype transitions (Kipf and Welling 2016). Although transformer-based models can handle sequence data (Vaswani et al. 2017; Zhou et al. 2021; Yu et al. 2022) and identify changes in cell states (Dijk et al. 2018), they still

[*]These authors contributed equally.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

face limitations in capturing how intracellular and extracellular signals jointly influence gene expression and cell fate determination (Schmidhuber 2015). These models often lack the integration of biological priors related to gene regulation, which limits their interpretability and constrains their ability to generalize across diverse biological contexts (Schmidhuber 2015).

Motivated by these challenges, we design a foundation model specifically designed for biologically informed cell phenotype transition (CPT). We introduce a dual-branch architecture, DualCPT, which integrates global transcriptomic features and local regulatory signals through specialized attention mechanisms. Specifically, the model comprises two branches: a *prediction branch* modeling comprehensive gene expression patterns, and a *selection branch* identifying critical regulatory genes via attention-based token selection. By embedding biological priors into attention mechanisms, the DualCPT model achieves interpretable gene prioritization and dynamic global-local feature fusion, thereby enhancing model transparency and accuracy.

Our DualCPT model has been extensively validated using large-scale public datasets, including more than 590,000 training samples (Dann et al. 2023), and systematically evaluated against state-of-the-art baseline models and specialized single-cell models. Given the non-directional nature of cellular phenotype transitions, although regulatory genes can serve as genetic signals for these transitions, the process remains influenced by inherent uncertainties in biological systems. To address this, we introduced uncertainty and confidence metrics to quantify the uncertainty index.

To further validate the reliability of the DualCPT model in modeling cellular phenotype transitions, we conducted extensive transfer learning experiments. In addition, to test the robustness of the model, we designed gene knockout experiments. This study not only deepens the understanding of cellular dynamics, but also aligns with the intrinsic logic of scientific innovation and technological development in the biomedical field. By revealing the molecular mechanisms underlying cell type transitions, this research provides technical support to understand the cell reprogramming process and promotes the development of regenerative medicine.

Our contribution is summarized as follows:

- **Paradigm shift:** For the first time, we consider the regulatory roles of critical loci and develop a biologically-informed, inherently interpretable dual-branch architecture, ensuring that locus discovery extends beyond post-hoc interpretable validation.
- **Cell selection module:** To better learn domain priors during modeling cell phenotype transdifferentiation, we designed a token selection module based on genetic regulatory mechanisms to identify key genes.
- **Global-local information fusion:** We design 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.

Related Work

Cellular Fate Modeling

Traditional Cellular Fate Models. In developmental biology, cells undergo various transformations during the differentiation process. Waddington landscape (Waddington 1957) is fundamental to understanding these critical transitions, as it illustrates how cells experience changes in their transcriptomes and epigenomes while differentiating into distinct cell types. Predicting early cell fate biases and uncovering the underlying mechanisms of cellular decision-making are vital for advancing Cell Reprogramming techniques (Wang et al. 2021). By deciphering the regulatory mechanisms that govern cell fate determination, we can reprogram cells into specific types for regenerative medicine applications, such as replacing damaged or lost cells in various tissues and organs.

Recent advances in single-cell technologies have enabled high-resolution profiling of individual cells (Ramachandran et al. 2020), but these methods provide only static snapshots, limiting insight into dynamic cellular changes. To infer temporal trajectories, pseudotime algorithms estimate cell order based on gene expression similarities (Manno et al. 2018), using strategies such as root-cell distances or entropy measures. RNA velocity can supplement these methods with directional cues (Bergen et al. 2020), though its accuracy is inconsistent. Despite their utility, existing approaches often fail to precisely detect key transitions or elucidate the underlying biological processes.

Deep Learning-Based Cellular Fate Models. In recent years, deep learning has become a powerful tool to predict changes in complex dynamic systems (Eraslan et al. 2019). In particular, neural networks can learn to predict state transitions by training on large datasets generated from noisy dynamic systems (Lotfollahi et al. 2021). Recent advances integrate dynamical systems theory with deep learning to directly predict cell fate decisions from transcriptomic dynamics (Sadria and Bury 2024), while (Sha et al. 2024) employs optimal transport and neural ODEs to reconstruct cellular growth trajectories and dynamic lineage progressions. However, the high dimensionality and sparsity commonly found in scRNA-seq data hinder the direct application of mainstream deep learning methods to this area of research.

Explainable Medical Artificial Intelligence

The integration of AI into healthcare offers significant potential for improving patient care and clinical decision-making. However, the black-box nature poses challenges regarding its interpretability and trustworthiness (Ribeiro, Singh, and Guestrin 2016). Recent research has focused on developing explainable AI (XAI) frameworks tailored for the healthcare domain. Techniques like LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) provide local explanations for individual predictions (Lundberg and Lee 2017), while attention mechanisms help identify key regions in medical images for diagnostic insight (Li et al. 2023). Beyond method development, there is a growing emphasis on the ethical implications of AI in medicine, advocating for frameworks that pri-

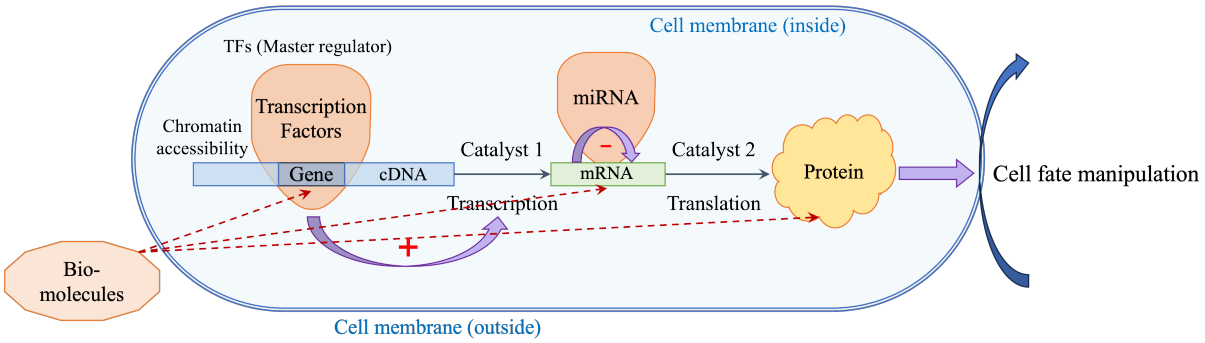

Figure 1: Biological Background of Cell Reprogramming. The cellular reprogramming process is regulated by three types of substances: transcription factors, miRNAs, and small biochemical molecules. This figure illustrates the binding processes of these three biochemical molecules to their targets, among them, cDNA designates chromosomal DNA.

oritize fairness, accountability, and transparency (Akinrinola et al. 2024). Interpretable models such as decision trees and rule-based systems are gaining interest due to their transparent logic (Rane, Choudhary, and Rane 2023). Counterfactual explanations further aid understanding by illustrating how changes in inputs affect predictions (Prosperi et al. 2020).

Methodology

Preliminary Section

We introduce a mathematical framework aimed at modeling cellular phenotype transitions, which is an essential biological process underlying differentiation, reprogramming, and disease progression. Our goal is to predict changes in cell identity by integrating global expression profiles with localized regulatory signals derived from scRNA-seq 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 $\mathbf{x}_i \in \mathbb{R}^d$ represents the expression levels of d genes. We define m phenotypic states, $\mathcal{C}_1$, $\mathcal{C}_2$,..., $\mathcal{C}_m$,corresponding to distinct cellular identities. Each cell i is associated with a categorical phenotype label $\mathbf{y}_i \in \{1, 2, \ldots, m\}$, where $y_i = k$ indicates that the cell belongs to phenotype $\mathcal{C}_k$.

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 $f : \mathbb{R}^d \to \mathbb{R}^d$ 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 weighted sum 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 function f_{reg} combines regulatory information with the original gene expression profiles, parameterized by θ_2.

Cell Phenotype Transdifferentiation Framework

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 the preliminary section. As shown in Figure 2, our model consists of two distinct branches: one for capturing the full expression profiles of the cells and the other for modeling regulatory influences through a top-k selection mechanism.

DualCPT 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. Specifically, the framework consists of the following components:

- **Prediction Branch:** This branch utilizes a pretrained single-cell model to obtain single-cell embeddings, and ultimately performs cell classification using a multilayer perceptron.
- **Selection Branch:** Important gene selection is conducted through our *Cell Learner* module. This is followed by multimodal information integration using cross-attention and the markov state transition matrix. The uncertainty of

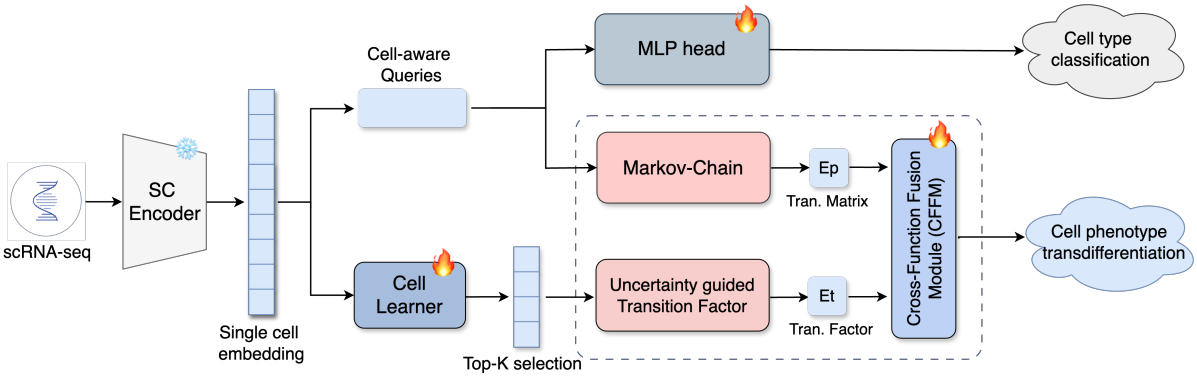

Figure 2: The overall pipeline of DualCPT consists of two branches: the classification branch, which extracts complete sequence information, and the transition branch, which employs the Cell Learner module to model the modulation effects of critical sites. A fusion process then combines the transition matrix derived from the Markov chain with the Transition Factor computed based on probability and uncertainty to perform the cell phenotype transition task.

the model is assessed based on the category information entropy and overall entropy.

The prediction branch symbolizes the deterministic role of coding regions in gene expression, while the selection branch represents the selection function of non-coding regions. These two branches together influence the gene expression process.

Cellular Phenotype Classification Branch

As shown in Figure 2, 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 function defined as $\sigma(x) = \max(0, x)$.

To address data distribution bias, we establish a multi-metric evaluation system incorporating Accuracy (ACC), F1-score, and Cohen's Kappa coefficient (κ). These metrics are computed as follows:

$$\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{6}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}, \tag{7}$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad \kappa = \frac{p_o - p_e}{1 - p_e}, \tag{8}$$

where $p_o = \text{ACC}$ and $p_e = \sum_{i=1}^{k} \frac{n_{i+} \times n_{+i}}{N^2}$. Here, TP/FP denote true/false positives, TN/FN represent true/false negatives, n_{i+} and n_{+i} are row/column marginal sums in the confusion matrix, and N is the total sample size. This ensures robustness against class imbalance and distribution shifts.

Regulatory Gene Transition Branch

During modeling of cellular phenotypic transitions, since the relationship between key regulatory sites and cell phenotype transitions is not reversible, we employ uncertainty and confidence metrics for each phenotype to assess the uncertainty level of the well-trained Markov transition probability matrix.

The Markov transition probability matrix P is derived from observed phenotype transitions in time-series data. For a system with n discrete phenotypes, each element P_{ij} in the matrix represents the probability of transitioning from state C_i to C_j. The matrix is constructed through two key steps: 1) *Phenotype Transition Counts:* Compute the frequency of transitions between phenotypes over all observed trajectories

$$N_{ij} = \sum_{t=1}^{T-1} \mathbb{I}(C(t) = C_i \wedge C(t+1) = C_j), \tag{9}$$

where N_{ij} is the number of transitions from C_i to C_j, T is the total time steps, and $\mathbb{I}$ is an indicator function. 2) *Probability Normalization:* Convert counts to probabilities by row normalization:

$$P_{ij} = \frac{N_{ij}}{\sum_{k=1}^{n} N_{ik}}, \tag{10}$$

ensuring $\sum_{j=1}^{n} P_{ij} = 1$ for each row i.

Uncertainty quantifies the unpredictability of transitions for a given category C_i, calculated directly from the probability distribution of transitions. For a category C_i, the uncertainty

$U(C_i)$ is defined as:

$$U(C_i) = -\sum_{j=1}^{n} P_{ij} \log(P_{ij} + \epsilon), \tag{11}$$

where P_{ij} represents the transition probability from state C_i to C_j, n is the total number of states, and ϵ (e.g., 10^{-10}) ensures numerical stability. A higher uncertainty value indicates greater unpredictability in transitions.

Confidence measures the certainty of the most probable transition path for category C_i, computed as:

$$\text{Confidence}(Q_i) = \max_j P_{ij}, \tag{12}$$

where $\max_j P_{ij}$ denotes the maximum probability in the transition distribution. Higher confidence reflects a stronger dominance of a specific transition path.

These metrics jointly characterize the robustness of the Markov transition probability matrix:

- *Low uncertainty and high confidence*: Transition paths are well-defined and stable, reflecting strong biological regulation or environmental constraints.
- *High uncertainty and low confidence*: Transitions are variable and context-dependent, potentially influenced by noise, incomplete data, or competing regulatory signals.

Cell Learner

The core innovation of Cell Learner lies in its *hierarchical attention mechanism* and *dynamic token selection strategy*, which together prioritize regulatory genes from high-dimensional single-cell expression profiles. The pipeline operates in two stages: 1) modeling gene regulatory interactions via attention weight calculation and 2) selecting key regulatory genes through token pruning.

Stage 1: Gene Regulatory Network (GRN) Modeling

- The input single-cell expression matrix $X \in \mathbb{R}^{N \times d}$ (N: genes, d: cells) is treated as a sequence of *gene tokens* and fed into a Transformer encoder.
- At each layer, a *gene–gene attention matrix* $A \in \mathbb{R}^{N \times N}$ is computed, where each entry A_{ij} quantifies the regulatory potential of gene i on gene j: Softmax $\left(QK^T/\sqrt{d_k}\right)$

Stage 2: Key Regulatory Gene Selection

- A gene importance score $S \in \mathbb{R}^N$ is derived by aggregating attention weights to capture each gene's regulatory centrality:

$$S_i = \underbrace{\sum_{j=1}^{N} A_{ij}}_{\text{Global regulatory strength}} + \lambda \cdot \underbrace{\text{Pos}(i)}_{\text{Positional bias}}, \tag{13}$$

  where the first term represents the total regulatory influence of gene i on all other genes, while the second term introduces an optional positional bias that incorporates prior biological knowledge (e.g., promoter or enhancer regions).

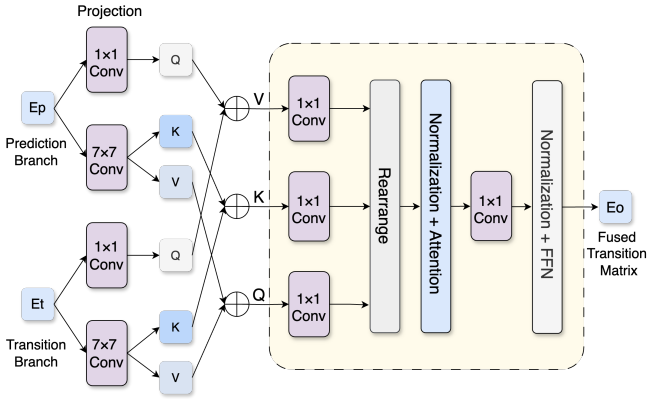

Figure 3: Cross-Function Fusion Module.

- The top-K genes with the highest S_i values are selected as the *candidate regulatory gene set* $\mathcal{G}_{\text{reg}}$, while others are pruned:

$$\mathcal{G}_{\text{reg}} = \text{TopK}(S, K). \tag{14}$$

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 3, 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:

$$Q_i = E_i W_q^i, \quad K_i = E_i W_k^i, \quad V_i = E_i W_v^i, \quad i \in \{p, t\}, \tag{15}$$

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 from the input. We first concatenate the projections by:

$$\begin{cases} Q = \text{Concat}(Q_p, Q_t), \\ K = \text{Concat}(K_p, K_t), \\ V = \text{Concat}(V_p, V_t). \end{cases} \tag{16}$$

The normalization operation is implemented through layer-wise transformations: $\bar{Q} = \text{LayerNorm}(Q)$ and $\bar{K} = \text{LayerNorm}(K)$. We then fuse the projections via cross-attention and obtain the attention score: $A = \text{softmax}\left(\bar{Q}\bar{K}^\top/\sqrt{c/n}\right)$.

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

Loss Function for Supervised Training.

We optimize the model using a composite loss that jointly enforces discriminative cell type classification and faithful modeling of transition dynamics. The individual components are defined as follows:

Multi-class Cross-Entropy Loss:

$$\mathcal{L}_{\text{CE}} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{c=1}^{C} y_{i,c}\log\left(\hat{y}_{i,c}^{\text{fusion}}\right), \tag{17}$$

where C is the total number of cell type categories, $y_{i,c} \in \{0,1\}$ is the one-hot encoded ground truth label for sample i, and $\hat{y}_{i,c}^{\text{fusion}}$ is the fused prediction probability for class c.

Transition MSE Loss:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{n}\sum_{i=1}^{n}\left\|\mathbf{M}_i - \hat{\mathbf{M}}_i\right\|_F^2. \tag{18}$$

In this equation, $\mathbf{M}_i \in \mathbb{R}^{K\times K}$ is the true Markov transition matrix for sample i, $\hat{\mathbf{M}}_i$ is its prediction, and $\|\cdot\|_F$ denotes the Frobenius norm.

Total Loss is expressed as a weighted combination of the above loss terms:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \lambda \cdot \mathcal{L}_{\text{MSE}}, \tag{19}$$

where λ is a hyperparameter (suggested range [0.1, 1.0]) that adjusts the relative importance of the Transition MSE term. This joint loss supports simultaneous optimization of:

- **Cell Type Discrimination:** Minimizing $\mathcal{L}_{\text{CE}}$ encourages parameters $\{\theta_1, \theta_2, \theta_g\}$ to learn discriminative features.
- **Transition Dynamics Consistency:** Minimizing $\mathcal{L}_{\text{MSE}}$ ensures that the network captures the correct biological transition patterns via parameters $\{\theta_1, \theta_2\}$.

The overall optimization objective is then given by:

$$\theta^* = \arg\min_{\theta_1,\theta_2,\theta_g} \mathcal{L}_{\text{total}}. \tag{20}$$

Experiments

Experimental Setups

Dataset. 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. To comprehensively evaluate our DualCPT model, we compare it against several state-of-the-art deep learning architectures categorized as follows: 1) CNN-based models: ResNet (He et al. 2016), Res2Net (Gao et al. 2019), and ConvNeXt (Liu et al. 2022); 2) transformer-based models: original Transformer (Vaswani et al. 2017), Informer (Zhou et al. 2021), and Metaformer (Yu et al. 2022); and 3) specialized single-cell models: scBERT (Yang et al. 2022) and scGPT (Cui et al. 2024). Each baseline model is rigorously trained and evaluated under identical experimental conditions to ensure a fair and reliable comparison.

Table 1: **Comparison of Cellular Phenotype Classification Methods.**

Method	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
scGPT	90.13	88.31	86.53
DualCPT	**91.60**	**91.43**	**89.94**

Table 2: **Comparison of XAI Baseline**.

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

Evaluation Metrics. For the prediction branch, the evaluation metrics include accuracy, F1-score, and the Kappa coefficient, which measure the effectiveness of our model in accurately predicting cell phenotypes. In the selection branch, we assess interpretability and reliability through confidence, uncertainty, and accuracy metrics, focusing specifically on the performance improvement gained from the optimized Top-K confidence calculation approach.

Implementation Details. We trained our model over 10 epochs using two L20 GPUs. We configured the model's dropout rate at 0.1 and set the learning rate to 1×10^{-3}.

Comparison Results

Cellular Phenotype Classification: Table 1 compares the cellular phenotype classification performance across baseline models and our proposed DualCPT architecture. Our DualCPT model consistently achieves the highest scores, surpassing conventional CNN-based models (ResNet, Res2Net, ConvNeXt), Transformer-based models (Informer, Metaformer), and single-cell models (scBERT, scGPT).

The performance of DualCPT (Accuracy: 91.60%, F1: 91.43%, Kappa: 89.94%) highlights the effectiveness of our dual-branch design, demonstrating the model's capability to integrate comprehensive global expression data with local regulatory information, thus providing accurate phenotype predictions. The significant improvement over scBERT underscores the limitations of traditional transformer-based language models in capturing complex cellular phenotype information without specialized biological priors.

Explainable AI (XAI) Comparison: Table 2 summarizes results from the explainable AI (XAI) baseline comparison. Our DualCPT approach achieves the highest con-

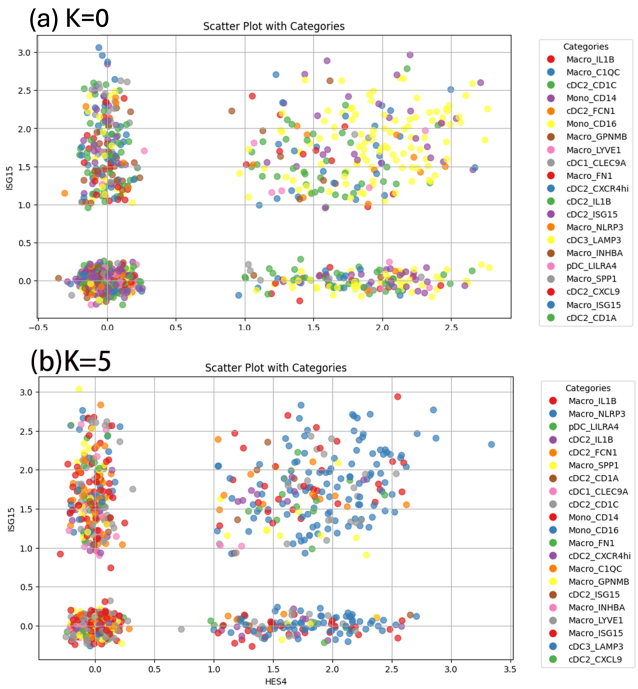

Figure 4: Data Distribution Scatter Plots.(a)and(b) illustrate the changes in single-cell data distribution following Top-1 and Top-5 gene knockouts, respectively.

Table 3: **Ablation Study on the impact of the Cell Learner module and the CFFM for the transition branch**.

CFFM	Cell Learner	Conf.	Uncer.	Acc.	F1	Kappa
×	×	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**

fidence (91.82%) and lowest uncertainty (21.28%) scores compared to conventional XAI methods such as Feature Importance, LIME, Diffrate, and Token Learner. This indicates that our biologically informed attention mechanism effectively selects the most informative genes, leading to more interpretable and reliable predictions. The improved interpretability of DualCPT can be attributed to its biologically motivated token selection and advanced attention mechanisms, offering clinicians actionable insights into critical regulatory genes influencing phenotype transitions.

Ablation Study

Here, we conduct ablation experiments to evaluate the rationality of the transition branch and the design of the global-local fusion module (CFFM). Specifically, we assess the impact of the Cell Learner module and the CFFM by testing four model variants: (1) baseline DualCPT without either component, (2) DualCPT with only CFFM, (3) DualCPT with only Cell Learner, and (4) the full model with both CFFM and Cell Learner. Results are presented in Table 3.

Figure 5: Box plot of the transfer learning experiments. To validate the generalization ability of the model, we introduced two additional datasets for transfer learning and evaluated the model performance based on accuracy, F1-score, and Kappa coefficient.

Table 4: **Hyperparametric Sensitivity**.

Keep rate	Confidence	Uncertainty	Acc.	F1	Kappa
1.0	84.86	25.15	92.10	92.80	90.52
0.9	86.69	24.66	91.31	92.31	90.17
0.8	**91.82**	**21.28**	**91.60**	**91.43**	**89.94**
0.7	89.23	24.76	89.47	89.38	87.61

Combining both CFFM and Cell Learner yields the best overall performance across all metrics: 91.60% accuracy, 91.43 F1, and the lowest uncertainty (21.28). Notably, the synergistic effect of combining both modules provides a compounded benefit—both modules contribute individually, but their joint usage leads to stronger, complementary improvements. The increase in confidence (+2.66 over baseline) and the highest Kappa score (89.94) further validate the model's reliability and agreement with the ground truth.

Hyperparametric Sensitivity

We conduct a comprehensive parameter sensitivity analysis of critical biological site screening using Cell Learner with a sparse attention mechanism. In Table 4, we systematically investigated four token retention ratios (1.0, 0.9, 0.8, 0.7) while monitoring three key evaluation metrics: prediction uncertainty, confidence level, and classification accuracy. The experimental results revealed distinct patterns across different retention settings. The results indicate that a keep rate of 0.8 is the optimal compromise, effectively harmonizing robust classification performance with reduced prediction uncertainty, making it the most balanced configuration for practical deployment.

Transfer Learning

To validate the generalization capability of the model in cell phenotype transdifferentiation tasks, we conducted transfer learning on two downstream cell type annotation task datasets from scGPT tasks. Specifically, we performed four repeated experiments on the hPancreas4 dataset and Multiple Sclerosis dataset, evaluating model performance using

Table 5: **The result of uncertainty analysis**

Top-K Gene	Confidence	Uncertainty
0	76.99	68.08
1	77.08	68.60
2	74.41	77.84
3	76.72	69.44
4	76.98	69.14
5	76.39	60.01

multiple metrics including Accuracy, F1-score, and Kappa coefficient, as shown in Figure 5.

Analysis

To validate the biological rationale for modeling cell phenotype transdifferentiation, we simulated gene knockout experiments. Specifically, we set the expression levels of key genes identified by the random forest algorithm to zero to simulate the gene knockout process and monitored changes in the transition probability matrix. The experimental results demonstrated that the knockout of key genes significantly altered cell type distribution, with the knockout of S100A9, FCN1, FCGR3A, S100A8, and C1QC even leading to a complete shift in cell type distribution.

Our findings demonstrate that the loss of specific genes disrupts established regulatory pathways, underscoring their essential roles in maintaining transdifferentiation dynamics,as shown in Figure 4. Additionally, as shown in Table 5, the model demonstrates good robustness. The results highlight the network-like interactions of genetic factors during phenotypic transitions, providing novel insights into the underlying biological mechanisms.

Conclusions

This paper proposes a novel framework, DualCPT, 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 DualCPT model exhibits high robustness and accuracy.

References

Akinrinola, O.; Okoye, C. C.; Ofodile, O. C.; and Ugochukwu, C. E. 2024. Navigating and reviewing ethical dilemmas in AI development: Strategies for transparency, fairness, and accountability. *GSC Advanced Research and Reviews*, 18(3): 050–058.

Arlotta, P.; and Berninger, B. 2012. Cellular Reprogramming: A New Approach to Disease Modeling. *Cell Stem Cell*, 10(6): 653–655.

Bergen, V.; et al. 2020. Generalizing RNA Velocity to Transient Cell States through Dynamical Modeling. *Nature Biotechnology*, 38: 1408–1414.

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

Dann, E.; et al. 2023. Precise Identification of Cell States Altered in Disease Using Healthy Single-Cell References. *Nature Genetics*, 55(11): 1998–2008.

Dijk, D. V.; et al. 2018. Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell*, 174(3): 716–729.

Efremova, M.; et al. 2020. CellPhoneDB: Inferring Cell-Cell Communication from Combined Expression of Multi-Subunit Ligand-Receptor Complexes. *Nature Protocols*, 15(4): 1484–1506.

Eraslan, G.; et al. 2019. Deep Learning: New Computational Modelling Techniques for Single-Cell RNA-Seq Analysis. *Nature Reviews Genetics*, 20: 296–312.

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

Hanahan, D.; and Weinberg, R. A. 2011. Hallmarks of Cancer: The Next Generation. *Cell*, 144(5): 646–674.

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

Kipf, T. N.; and Welling, M. 2016. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv preprint*. ArXiv:1609.02907.

Lambert, A. W.; Pattabiraman, D. R.; and Weinberg, R. A. 2017. Emerging Biological Principles of Metastasis. *Cell*, 168(4): 670–691.

Li, X.; Li, M.; Yan, P.; Li, G.; Jiang, Y.; Luo, H.; and Yin, S. 2023. Deep learning attention mechanism in medical image analysis: Basics and beyonds. *International Journal of Network Dynamics and Intelligence*, 93–116.

Liu, Z.; Lou, H.; and Xie, K. 2019. Single-Cell Transcriptome Analysis Reveals Cell Fate Determination in Developing Pancreas. *Cell Reports*, 28(11): 2874–2886.

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

Lotfollahi, M.; et al. 2019. scGen Predicts Single-Cell Perturbation Responses. *Nature Methods*, 16(8): 715–721.

Lotfollahi, M.; et al. 2021. Learning Interpretable Cellular Responses to Complex Perturbations. *Nature Methods*, 18: 1183–1190.

Lundberg, S. M.; and Lee, S.-I. 2017. A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, 30.

Manno, M. L.; et al. 2018. RNA Velocity of Single Cells. *Nature*, 560: 494–498.

Prosperi, M.; Guo, Y.; Sperrin, M.; Koopman, J. S.; Min, J. S.; He, X.; Rich, S.; Wang, M.; Buchan, I. E.; and Bian, J. 2020. Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nature Machine Intelligence*, 2(7): 369–375.

Ramachandran, P.; Matchett, K. P.; Dobie, R.; Wilson-Kanamori, J. R.; and Henderson, N. C. 2020. Single-cell technologies in hepatology: new insights into liver biology and disease pathogenesis. *Nature reviews Gastroenterology & hepatology*, 17(8): 457–472.

Rane, N.; Choudhary, S.; and Rane, J. 2023. Explainable artificial intelligence (XAI) in healthcare: Interpretable models for clinical decision support. *Available at SSRN 4637897*.

Ribeiro, M. T.; Singh, S.; and Guestrin, C. 2016. Why Should I Trust You? Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.

Ritchie, M. E.; et al. 2015. limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Research*, 43(7): e47.

Rockey, D. C.; et al. 2015. Fibrosis–A Common Pathway to Organ Injury and Failure. *New England Journal of Medicine*, 372(12): 1138–1149.

Sadria, M.; and Bury, T. M. 2024. FateNet: an integration of dynamical systems and deep learning for cell fate prediction. *Bioinformatics*, 40(9): btae525.

Schmidhuber, J. 2015. Deep Learning in Neural Networks: An Overview. *Neural Networks*, 61: 85–117.

Sha, Y.; Qiu, Y.; Zhou, P.; and Nie, Q. 2024. Reconstructing growth and dynamic trajectories from single-cell transcriptomics data. *Nature Machine Intelligence*, 6(1): 25–39.

Stuart, T.; et al. 2019. Comprehensive Integration of Single-Cell Data. *Cell*, 177(7): 1888–1902.

Trapnell, C.; et al. 2014. The Dynamics and Regulators of Cell Fate Decisions Are Revealed by Pseudotemporal Ordering of Single Cells. *Nature Biotechnology*, 32(4): 381–386.

Vaswani, A.; et al. 2017. Attention Is All You Need. In *Advances in Neural Information Processing Systems*, volume 30, 5998–6008. Long Beach, CA: Curran Associates, Inc.

Waddington, C. H. 1957. The Strategy of the Genes. *George Allen & Unwin*.

Wang, H.; Yang, Y.; Liu, J.; and Qian, L. 2021. Direct cell reprogramming: approaches, mechanisms and progress. *Nature Reviews Molecular Cell Biology*, 22(6): 410–424.

Wolf, F. A.; Angerer, P.; and Theis, F. J. 2018. SCANPY: Large-Scale Single-Cell Gene Expression Data Analysis. *Genome Biology*, 19(1): 15.

Yang, F.; Wang, W.; Wang, F.; Fang, Y.; Tang, D.; Huang, J.; Lu, H.; and Yao, J. 2022. 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.

Yu, W.; Luo, M.; Zhou, P.; Si, C.; Zhou, Y.; Wang, X.; Feng, J.; and Yan, S. 2022. Metaformer is actually what you need for vision. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10819–10829.

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

Reproducibility Checklist

1. General Paper Structure

1.1. Includes a conceptual outline and/or pseudocode description of AI methods introduced (yes/partial/no/NA) yes

1.2. Clearly delineates statements that are opinions, hypothesis, and speculation from objective facts and results (yes/no) yes

1.3. Provides well-marked pedagogical references for less-familiar readers to gain background necessary to replicate the paper (yes/no) yes

2. Theoretical Contributions

2.1. Does this paper make theoretical contributions? (yes/no) yes

If yes, please address the following points:

2.2. All assumptions and restrictions are stated clearly and formally (yes/partial/no) yes

2.3. All novel claims are stated formally (e.g., in theorem statements) (yes/partial/no) yes

2.4. Proofs of all novel claims are included (yes/partial/no) yes

2.5. Proof sketches or intuitions are given for complex and/or novel results (yes/partial/no) yes

2.6. Appropriate citations to theoretical tools used are given (yes/partial/no) yes

2.7. All theoretical claims are demonstrated empirically to hold (yes/partial/no/NA) yes

2.8. All experimental code used to eliminate or disprove claims is included (yes/no/NA) yes

3. Dataset Usage

3.1. Does this paper rely on one or more datasets? (yes/no) yes

If yes, please address the following points:

3.2. A motivation is given for why the experiments are conducted on the selected datasets (yes/partial/no/NA) yes

3.3. All novel datasets introduced in this paper are included in a data appendix (yes/partial/no/NA) NA

3.4. All novel datasets introduced in this paper will be made publicly available upon publication of the paper with a license that allows free usage for research purposes (yes/partial/no/NA) NA

3.5. All datasets drawn from the existing literature (potentially including authors' own previously published work) are accompanied by appropriate citations (yes/no/NA) yes

3.6. All datasets drawn from the existing literature (potentially including authors' own previously published work) are publicly available (yes/partial/no/NA) yes

3.7. All datasets that are not publicly available are described in detail, with explanation why publicly available alternatives are not scientifically satisficing (yes/partial/no/NA) NA

4. Computational Experiments

4.1. Does this paper include computational experiments? (yes/no) yes

If yes, please address the following points:

4.2. This paper states the number and range of values tried per (hyper-) parameter during development of the paper, along with the criterion used for selecting the final parameter setting (yes/partial/no/NA) yes

4.3. Any code required for pre-processing data is included in the appendix (yes/partial/no) yes

4.4. All source code required for conducting and analyzing the experiments is included in a code appendix (yes/partial/no) yes

4.5. All source code required for conducting and analyzing the experiments will be made publicly available upon publication of the paper with a license that allows free usage for research purposes (yes/partial/no) yes

4.6. All source code implementing new methods have comments detailing the implementation, with references to the paper where each step comes from (yes/partial/no) yes

4.7. If an algorithm depends on randomness, then the method used for setting seeds is described in a way sufficient to allow replication of results (yes/partial/no/NA) yes

4.8. This paper specifies the computing infrastructure used for running experiments (hardware and software), including GPU/CPU models; amount of memory; operating system; names and versions of relevant software libraries and frameworks (yes/partial/no) NA

4.9. This paper formally describes evaluation metrics used and explains the motivation for choosing these metrics (yes/partial/no) yes

4.10. This paper states the number of algorithm runs used to compute each reported result (yes/no) yes

4.11. Analysis of experiments goes beyond single-

dimensional summaries of performance (e.g., average; median) to include measures of variation, confidence, or other distributional information (yes/no) yes

4.12. The significance of any improvement or decrease in performance is judged using appropriate statistical tests (e.g., Wilcoxon signed-rank) (yes/partial/no) yes

4.13. This paper lists all final (hyper-)parameters used for each model/algorithm in the paper's experiments (yes/partial/no/NA) yes