Novelty Assessment Report

Paper: HYPED: A Multimodal HYbrid Perturbation Gene Expression and Imaging Dataset

PDF URL: https://openreview.net/pdf?id=HsJJDxjJ4R

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2025-12-29

Abstract

Integrating multimodal, high-resolution biological data is a useful way to characterize biological processes, such as how cells respond to perturbations. Cell perturbation prediction is a major experimental challenge and has motivated substantial research in machine learning for biology. In this work, we generated a multimodal benchmark dataset that captures the dynamic response of human fibroblasts to transient transcription factor perturbations. We performed time-series live cell imaging with fluorescent cell cycle reporters over 72 hours and collected long-read single-cell RNA sequencing data from the same population of cells. We release the processed dataset, preprocessing pipelines and benchmarking code along with the evaluation of existing models using our data as ground truth. This work supports the development and evaluation of machine learning methods for modeling dynamical systems from multimodal datasets. HYPED consists of RNA sequencing data from approximately 20,000 cells and 203 imaging timepoints across four experimental conditions, totaling 2030 imaging frames. HYPED makes the cell perturbation problem accessible to machine learning researchers with state-of-the-art experimental data.

Disclaimer

This report is **AI-GENERATED** using Large Language Models and WisPaper (a scholar search engine). It analyzes academic papers' tasks and contributions against retrieved prior work. While this system identifies **POTENTIAL** overlaps and novel directions, **ITS COVERAGE IS NOT EXHAUSTIVE AND JUDGMENTS ARE APPROXIMATE**. These results are intended to assist human reviewers and **SHOULD NOT** be relied upon as a definitive verdict on novelty.

Note that some papers exist in multiple, slightly different versions (e.g., with different titles or URLs). The system may retrieve several versions of the same underlying work. The current automated pipeline does not reliably align or distinguish these cases, so human reviewers will need to disambiguate them manually.

If you have any questions, please contact: mingzhang23@m.fudan.edu.cn

Core Task Landscape

This paper addresses: **Predicting Cellular Responses to Transcription Factor Perturbations**

A total of **50 papers** were analyzed and organized into a taxonomy with **18 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Computational Prediction Methods for Perturbation Response**
- **Regulatory Network Inference and Causal Discovery**
- **Experimental Perturbation Platforms and Resources**
- **Benchmarking and Evaluation Frameworks**
- **Mechanistic and Systems Biology Models**
- **Domain-Specific Applications and Case Studies**

Complete Taxonomy Tree

- Predicting Cellular Responses to Transcription Factor Perturbations Survey Taxonomy
- Computational Prediction Methods for Perturbation Response
  - Machine Learning Models for Gene Expression Prediction
  - Feature-Based Predictive Models (3 papers)
    - [1] Predicting which genes will respond to transcription factor perturbations (Yiming Kang, 2022) View paper
    - [9] Transfer learning reveals sequence determinants of the quantitative response to transcription factor dosage (Sahin Naqvi, 2024) View paper
    - [45] A quantitative model of transcription factorâ activated gene expression (Harold D. Kim, 2008) View paper
  - Expression-Based Prediction Models (3 papers)
    - [31] Inferring TF activities and activity regulators from gene expression data with constraints from TF perturbation data (Cynthia Z Ma, 2021) View paper
    - [48] Computational discovery of transcription factors associated with drug response (Casey Hanson, 2016) View paper
    - [49] TIGERi: modeling and visualizing the responses to perturbation of a transcription factor network (Han, 2017) View paper
  - Foundation Models and Generative Approaches
  - Transformer and Foundation Model Architectures (2 papers)
    - [3] A foundation model of transcription across human cell types (Xi Fu, 2025) View paper
    - [7] PertEval-scFM: Benchmarking Single-Cell Foundation Models for Perturbation Effect Prediction (Aaron Wenteler, 2025) View paper
  - Generative Modeling Approaches (2 papers)
    - [5] Predicting the unseen: a diffusion-based debiasing framework for transcriptional response prediction at single-cell resolution (Yu-Ting Wei, 2025) View paper
    - [24] The manatee variational autoencoder model for predicting gene expression alterations caused by transcription factor perturbations (Ying Yang, 2024) View paper
  - Out-of-Distribution and Transfer Learning Methods (2 papers)
  - [8] TxPert: Leveraging Biochemical Relationships for Out-of-Distribution Transcriptomic Perturbation Prediction (Wenkel, 2025) View paper
  - [25] A systematic comparison of single-cell perturbation response prediction models (Lanxiang Li, 2024) View paper
- Regulatory Network Inference and Causal Discovery
  - Dynamic and Temporal Network Models (3 papers)

  - [4] Interpretable neural odes for gene regulatory network discovery under perturbations (Lin, 2025) View paper
  - [11] Generative modeling of single-cell time series with PRESCIENT enables prediction of cell trajectories with interventions (G. Yeo, 2021) View paper
  - [42] RegVelo: gene-regulatory-informed dynamics of single cells (Weixu Wang, 2024) View paper
  - Static Regulatory Network Inference (3 papers)
  - [16] Learning causal networks using inducible transcription factors and transcriptomeâwide time series (Sean R. Hackett, 2020) View paper
  - [23] Genome-wide promoter responses to CRISPR perturbations of regulators reveal regulatory networks in Escherichia coli (Yichao Han, 2023) View paper
  - [30] Logical modeling of lymphoid and myeloid cell specification and transdifferentiation (Samuel Collombet, 2017) View paper
  - Transcription Factor Activity Inference (4 papers)
  - [20] Flexible modeling of regulatory networks improves transcription factor activity estimation (Chen Chen, 2024) View paper
  - [29] Transcription factor enrichment analysis (TFEA) quantifies the activity of multiple transcription factors from a single experiment (Jonathan D. Rubin, 2021) View paper
  - [41] STAN, a computational framework for inferring spatially informed transcription factor activity across cellular contexts (Li-nan, 2025) View paper
  - [43] Uncertainty-aware transcription factor activity and perturbation inference without additional training (Hikaru Sugimoto, 2025) View paper
- Experimental Perturbation Platforms and Resources
  - CRISPR-Based Perturbation Libraries (2 papers)
  - [2] A comprehensive library of human transcription factors for cell fate engineering (Alex H. M. Ng, 2021) View paper
  - [17] Comprehensive transcription factor perturbations recapitulate fibroblast transcriptional states (Kaden M. Southard, 2025) View paper
  - Multimodal Perturbation Datasets ★ (2 papers)
  - [0] HYPED: A Multimodal HYbrid Perturbation Gene Expression and Imaging Dataset (Anon et al., 2026) View paper
  - [10] Toward a foundation model of causal cell and tissue biology with a Perturbation Cell and Tissue Atlas (Jennifer E. Rood, 2024) View paper
- Benchmarking and Evaluation Frameworks (2 papers)
  - [32] Machine learning for perturbational single-cell omics (Yuge Ji, 2021) View paper
  - [34] A mini-review on perturbation modelling across single-cell omic modalities (George I. Gavriilidis, 2024) View paper
- Mechanistic and Systems Biology Models
  - Signaling Dynamics and Pathway Models (3 papers)
  - [21] Signal-dependent dynamics of transcription factor translocation controls gene expression (Nan Hao, 2012) View paper
  - [39] Identifying the combinatorial control of signal-dependent transcription factors (Ning Wang, 2021) View paper
  - [47] Encoding and decoding cellular information through signaling dynamics (Galit Lahav, 2013) View paper
  - Transcriptional Regulation and Gene Expression Models (3 papers)
  - [22] Modeling and designing enhancers by introducing and harnessing transcription factor binding units. (Jiaqi Li, 2025) View paper
  - [28] Cell-adhesion molecule expression prediction from transcription factor expression patterns (Kieran James Lynch, 2025) View paper
  - [33] SpecLoop predicts cell type-specific chromatin loop via transcription factor cooperation. (Lixin Ren, 2024) View paper
  - Cell Fate and Differentiation Models (3 papers)
  - [14] Dynamic modelling of pathways to cellular senescence reveals strategies for targeted interventions (Piero Dalle Pezze, 2014) View paper
  - [27] Defining an essential transcription factor program for naive pluripotency (SJ Dunn, 2014) View paper
  - [44] MAP kinase signaling specificity mediated by the LIN-1 Ets/LIN-31 WH transcription factor complex during C. elegans vulval induction (Patrick B Tan, 1998) View paper
- Domain-Specific Applications and Case Studies
  - Disease and Therapeutic Applications (5 papers)
  - [6] An integrated approach to predict activators of NRF2-the transcription factor for oxidative stress response (Y. Chushak, 2024) View paper
  - [13] NOTCH1 signaling during CD4+ T-cell activation alters transcription factor networks and enhances antigen responsiveness (Alec B. Wilkens, 2022) View paper
  - [19] A novel investigation into an E2F transcription factorârelated prognostic model with seven signatures for colon cancer patients (Xiao-Yong Shen, 2023) View paper
  - [35] Identification of novel activators of the metal responsive transcription factor (MTF-1) using a gene expression biomarker in a microarray compendium (Abigail C Jackson, 2020) View paper
  - [50] The interaction of TEA domain transcription factor 4 (TEAD4) and Yes-associated protein 1 (YAP1) promoted the malignant process mediated by serum/glucocorticoid regulated kinase 1 (SGK1) (Songlin He, 2021) View paper
  - Stress Response and Environmental Perturbations (2 papers)
  - [18] Elucidating genes and transcription factors of human peripheral blood lymphocytes involved in the cellular response upon exposure to ionizing radiation for â¦ (P Gollapalli, 2024) View paper
  - [46] Zinc sensing by metal-responsive transcription factor 1 (MTF1) controls metallothionein and ZnT1 expression to buffer the sensitivity of the transcriptome response to â¦ (JEJ Hardyman, 2016) View paper
  - Specific Transcription Factor Families and Networks (4 papers)
  - [12] Decoding the Central Dogma: Quantitative Insights into Transcription and Translation Dynamics in the p53-Mediated DNA Damage Response (Joshua FranÃ§ois, 2025) View paper
  - [26] The MAX-interacting transcription factor network (P. D. Maker, 2006) View paper
  - [38] The multiple mechanisms that regulate p53 activity and cell fate (Antonina Hafner, 2019) View paper
  - [40] Altered transcription factor targeting is associated with differential peripheral blood mononuclear cell proportions in sarcoidosis (Christian Ascoli, 2022) View paper
  - Developmental and Tissue-Specific Contexts (3 papers)
  - [15] Vascular smooth muscle cell atherosclerosis trajectories characterized at single cell resolution identify causal transcriptomic and epigenomic mechanisms of disease â¦ (DY Li, 2025) View paper

- [36] Mechanistic Modeling the Role of MicroRNAs and Transcription Factors in Disease Progression. (Abdul Fayaz Shaik Mohammad, 2024) View paper
- [37] Transcription factors as important regulators of changes in behavior through domestication of gray rats: Quantitative data from RNA sequencing (D. Oshchepkov, 2022) View paper

Narrative

Core task: predicting cellular responses to transcription factor perturbations. The field has organized itself around several complementary branches. Computational Prediction Methods develop machine learning and deep learning frameworks—ranging from foundation models like Transcription Foundation Model[3] to specialized architectures such as Neural ODEs GRN[4]—that forecast gene expression changes following TF interventions. Regulatory Network Inference and Causal Discovery focuses on reconstructing the underlying wiring diagrams and causal relationships that govern TF activity. Experimental Perturbation Platforms and Resources provide the empirical substrate: large-scale libraries (e.g., Human TF Library[2]), systematic perturbation screens, and multimodal datasets that capture diverse cellular contexts. Benchmarking and Evaluation Frameworks, exemplified by efforts like PertEval scFM[7], establish standardized metrics and test beds to compare predictive models. Mechanistic and Systems Biology Models integrate biophysical principles and dynamical systems to explain how TF dynamics encode information. Finally, Domain-Specific Applications and Case Studies translate these tools into concrete biological questions, from cancer pathways (p53 DNA Damage[12]) to vascular disease (Vascular Cell Atherosclerosis[15]).

Within the Experimental Perturbation Platforms branch, a dense cluster of works has emerged around multimodal perturbation datasets that combine high-throughput screening with rich molecular readouts. The HYPED Dataset[0] exemplifies this trend by integrating multiple data modalities to capture TF perturbation effects across diverse conditions, positioning itself alongside resources like the Perturbation Atlas[10], which similarly aggregates large-scale perturbation profiles. These datasets address a critical bottleneck: computational models require extensive training data that span varied cell types, stimuli, and genetic backgrounds. In contrast, earlier platforms often focused on single modalities or narrower experimental designs. By providing a more comprehensive empirical foundation, HYPED Dataset[0] and related efforts enable more robust benchmarking of predictive algorithms and facilitate transfer learning approaches (Transfer Learning TF[9]) that generalize across contexts. The interplay between such resource-building initiatives and computational method development remains a central theme, as richer datasets continually push the frontier of what models can learn and predict.

Related Works in Same Category

The following **1 sibling papers** share the same taxonomy leaf node with the original paper:

1. Toward a foundation model of causal cell and tissue biology with a Perturbation Cell and Tissue Atlas

Authors: Jennifer E. Rood, Anna Hupalowska, Aviv Regev, A. Hupalowska | **Year/Venue**: 2024 | **URL**: View paper

Abstract

â¦ to accurately predict the outcome of perturbations and combinations â¦ molecular events across multiple cell contexts and at â¦ atlas 4); with transcription factor (TF) overexpression (for cell â¦

Relationship Analysis

Both papers belong to the Multimodal Perturbation Datasets category, focusing on integrating multiple measurement modalities for studying transcription factor perturbations. The original paper (HYPED) presents a specific experimental dataset combining live-cell imaging with long-read scRNA-seq from human fibroblasts perturbed with MYOD1 and PRRX1 using transient RNA-based methods, while the candidate paper describes a broader vision for a Perturbation Cell and Tissue Atlas (PCTA) that aims to systematically compile diverse perturbation data across multiple cell types and contexts to enable causal modeling of cell and tissue biology.

Contributions Analysis

Overall novelty summary. The paper introduces HYPED, a multimodal benchmark dataset combining time-series live cell imaging with fluorescent cell cycle reporters and long-read single-cell RNA sequencing from human fibroblasts subjected to transient transcription factor perturbations. Within the taxonomy, it resides in the 'Multimodal Perturbation Datasets' leaf under 'Experimental Perturbation Platforms and Resources'. This leaf contains only two papers total, indicating a relatively sparse research direction compared to more crowded computational branches like 'Machine Learning Models for Gene Expression Prediction' or 'Foundation Models and Generative Approaches'.

The neighboring 'CRISPR-Based Perturbation Libraries' leaf focuses on genome-scale CRISPR screens with transcriptomic readouts, while HYPED employs transient RNA-based perturbations with multimodal measurements. The broader 'Experimental Perturbation Platforms and Resources' branch sits alongside computational prediction methods and regulatory network inference, serving as the empirical substrate for model development. The taxonomy's scope note explicitly distinguishes multimodal datasets from single-modality transcriptomic platforms, positioning HYPED's integration of imaging and sequencing as a defining characteristic within this sparse experimental niche.

Among 26 candidates examined through limited semantic search, none clearly refuted any of the three contributions. The 'HYPED multimodal benchmark dataset' contribution examined 6 candidates with no refutations. The 'first perturbation dataset using transient RNA-based methods' claim examined 10 candidates without finding prior work demonstrating this specific combination. The 'processed dataset with preprocessing pipelines and benchmarking code' contribution similarly examined 10 candidates with no clear overlaps. These statistics suggest novelty within the examined scope, though the search was not exhaustive.

Given the limited search scale and the sparse population of the 'Multimodal Perturbation Datasets' leaf, the work appears to occupy a relatively underexplored niche combining transient perturbations with multimodal temporal measurements. The absence of refutations among 26 candidates supports novelty claims, though a broader literature search might reveal additional context. The dataset's emphasis on benchmarking code and preprocessing pipelines addresses practical reproducibility concerns in this emerging experimental domain.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: HYPED multimodal benchmark dataset

Description: The authors created a new dataset combining time-series live cell imaging with fluorescent cell cycle reporters and long-read single-cell RNA sequencing from the same population of cells undergoing transient transcription factor perturbations. This dataset includes approximately 20,000 cells and 203 imaging timepoints across four experimental conditions.

This contribution was assessed against **6 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. A mini-review on perturbation modelling across single-cell omic modalities

URL: View paper

Brief Assessment

Perturbation Modelling Review[34] is a review paper that surveys existing perturbation modelling methods and datasets across single-cell omic modalities. It does not present a specific multimodal benchmark dataset combining time-series imaging with single-cell RNA sequencing from transcription factor perturbations, and therefore does not challenge the novelty of the HYPED dataset.

2. MultiMAP: dimensionality reduction and integration of multimodal data

URL: View paper

Brief Assessment

MultiMAP[62] is a computational method for dimensionality reduction and integration of existing multimodal datasets, not a benchmark dataset. It does not create or provide experimental data combining time-series imaging with sequencing from transcription factor perturbations.

3. Building, benchmarking, and exploring perturbative maps of transcriptional and morphological data.

URL: View paper

Brief Assessment

Perturbative Maps[65] cannot be evaluated as a refutation candidate because no full text context was provided for this paper. Without access to the candidate paper's content, it is impossible to assess whether it demonstrates prior work on multimodal datasets combining time-series imaging and single-cell RNA sequencing for transcription factor perturbations.

4. MIRA: joint regulatory modeling of multimodal expression and chromatin accessibility in single cells

URL: View paper

Brief Assessment

MIRA[63] focuses on joint modeling of chromatin accessibility and gene expression in single cells, not on creating multimodal benchmark datasets combining time-series imaging with transcription factor perturbations.

5. Nonlinear transcriptional responses to gradual modulation of transcription factor dosage

URL: View paper

Brief Assessment

Nonlinear TF Responses[61] focuses on CRISPR-based dosage modulation of transcription factors with targeted single-cell multimodal sequencing, not on transient RNA-based perturbations with time-series imaging and long-read sequencing as in the original paper.

6. Biologically-Aware Multimodal Representation Learning Deciphers Single-Cell Functions and Dynamics

URL: View paper

Brief Assessment

Multimodal Representation Learning[64] appears to focus on general multimodal representation learning methods rather than providing a specific benchmark dataset for transcription factor perturbations with combined imaging and sequencing modalities.

Contribution 2: First perturbation dataset using transient RNA-based methods

Description: The authors provide the first multimodal cell perturbation dataset generated using non-integrating transient RNA delivery methods (modified mRNA and siRNA) rather than permanent genome modification approaches like viral vectors or CRISPR, offering safer experimental conditions that better reflect clinical translation potential.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Efficient genetic perturbation of murine sensory neurons in vivo using CRISPR/Cas9

URL: View paper

Brief Assessment

Murine Neuron CRISPR[59] focuses on CRISPR/Cas9-based permanent genome editing in mouse sensory neurons for pain research, not transient RNA delivery methods for cell perturbation datasets.

2. Expanding horizons of CRISPR applications beyond genome editing

URL: View paper

Brief Assessment

CRISPR Beyond Editing[51] focuses on CRISPR applications in diagnostics, biosensing, gene regulation, and imaging. It does not present or discuss cell perturbation datasets using transient RNA delivery methods for machine learning applications.

3. Epigenome editing technologies for discovery and medicine

URL: View paper

Brief Assessment

Epigenome Editing[53] discusses transient epigenome editing technologies but does not present a multimodal cell perturbation dataset. The candidate focuses on epigenome editing methods rather than dataset generation for machine learning applications.

4. Gene and RNA Editing: Revolutionary Approaches to Treating Diseases

URL: View paper

Brief Assessment

Gene RNA Editing[54] is a review paper discussing gene and RNA editing technologies for disease treatment, not a dataset paper. It does not present any cell perturbation datasets or experimental data collection methods.

5. High-content CRISPR screening

URL: View paper

Brief Assessment

High Content CRISPR[52] focuses on CRISPR-based perturbations and mentions transient plasmid introduction, but does not describe a perturbation dataset using transient RNA delivery methods (modified mRNA/siRNA) as the original paper does.

6. Massively parallel in vivo Perturb-seq reveals cell-type-specific transcriptional networks in cortical development

URL: View paper

Brief Assessment

Perturb Seq Cortex[55] uses CRISPR-based in vivo perturbations, not transient RNA delivery methods. The candidate focuses on cortical development with permanent genetic modifications, representing a different experimental approach.

7. The transience of transient overexpression

URL: View paper

Brief Assessment

Transient Overexpression[58] discusses genome engineering methods and cellular perturbations but does not present a perturbation dataset or specifically address transient RNA delivery methods versus permanent genome modification.

8. Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens

URL: View paper

Brief Assessment

Perturb Seq[60] uses CRISPR-based perturbations for permanent genome modification, not transient RNA delivery methods like modified mRNA or siRNA as claimed in the original paper's contribution.

9. Programmable RNA tracking in live cells with CRISPR/Cas9

URL: View paper

Brief Assessment

CRISPR RNA Tracking[57] focuses on RNA visualization and tracking methodology using CRISPR/Cas9, not on creating cell perturbation datasets with transient RNA delivery methods for machine learning applications.

10. Manipulating and studying gene function in human pluripotent stem cell models

URL: View paper

Brief Assessment

Pluripotent Stem Models[56] focuses on gene manipulation methods in pluripotent stem cells, including transient RNA delivery (modified mRNA and siRNA) for reprogramming applications. However, it does not present a multimodal perturbation dataset combining imaging and sequencing data, which is the core novelty claim of the original paper.

Contribution 3: Processed dataset with preprocessing pipelines and benchmarking code

Description: The authors provide not only the raw and processed multimodal data but also complete preprocessing pipelines and benchmarking code, enabling machine learning researchers to evaluate and develop models for cell perturbation prediction with standardized evaluation protocols.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Benchmarking foundation cell models for post-perturbation RNA-seq prediction

URL: View paper

Brief Assessment

Benchmarking Foundation Models[68] focuses on evaluating existing foundation models (scGPT, scFoundation) on post-perturbation RNA-seq prediction using established perturb-seq datasets, rather than providing new multimodal experimental data with preprocessing pipelines for cell perturbation prediction as the original paper does.

2. Multimodal Benchmarking of Foundation Model Representations for Cellular Perturbation Response Prediction

URL: View paper

Brief Assessment

Multimodal Foundation Benchmarking[73] focuses on benchmarking foundation model representations for perturbation prediction, not on providing processed datasets with preprocessing pipelines and benchmarking code for cell perturbation prediction.

3. Fast and scalable Wasserstein-1 neural optimal transport solver for single-cell perturbation prediction

URL: View paper

Brief Assessment

Wasserstein Neural Transport[71] focuses on developing a novel W1 optimal transport solver for single-cell perturbation prediction, not on providing datasets with preprocessing pipelines and benchmarking code.

4. PerturBase: a comprehensive database for single-cell perturbation data analysis and visualization

URL: View paper

Brief Assessment

PerturBase[69] is a database that consolidates existing scperturbation datasets with analysis results, but does not provide preprocessing pipelines or benchmarking code for model development as described in the original paper's contribution.

5. A benchmark for prediction of transcriptomic responses to chemical perturbations across cell types

URL: View paper

Brief Assessment

[Final Audit Failure] The model insisted on a refutation claim but failed to provide verifiable evidence after multiple retries. Marked as cannot_refute for safety. Please manually verify the candidate text.

6. Predicting cellular responses to perturbation across diverse contexts with State

URL: View paper

Brief Assessment

State[70] focuses on developing a transformer model for predicting perturbation effects and introduces an evaluation framework (cell-eval), but does not describe providing processed datasets with preprocessing pipelines and benchmarking code as a contribution.

7. Benchmarking AI Models for In Silico Gene Perturbation of Cells

URL: View paper

Brief Assessment

Gene Perturbation Benchmark[75] focuses on benchmarking AI models for in silico gene perturbation using existing scPerturb datasets, not on creating new experimental multimodal data with preprocessing pipelines for cell perturbation prediction as in the original paper.

---

8. Predicting cellular responses to complex perturbations in highâ□□throughput screens

URL: View paper

Brief Assessment

Complex Perturbation Prediction[66] focuses on developing a computational model (CPA) for predicting cellular responses to perturbations, not on providing processed datasets with preprocessing pipelines and benchmarking code for community use.

---

9. Learning single-cell perturbation responses using neural optimal transport

URL: View paper

Brief Assessment

Neural Optimal Transport[67] focuses on learning perturbation response maps using optimal transport theory for drug response prediction, not on providing standardized preprocessing pipelines and benchmarking infrastructure for dataset evaluation.

---

10. Perturbench: Benchmarking machine learning models for cellular perturbation analysis

URL: View paper

Brief Assessment

Perturbench[72] focuses on benchmarking machine learning models for perturbation prediction using existing datasets, not on providing new experimental data with preprocessing pipelines. The original paper provides novel multimodal experimental data (imaging + sequencing) with complete preprocessing pipelines, while Perturbench[72] curates existing datasets for model evaluation purposes.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] HYPED: A Multimodal HYbrid Perturbation Gene Expression and Imaging Dataset View paper
- [1] Predicting which genes will respond to transcription factor perturbations View paper
- [2] A comprehensive library of human transcription factors for cell fate engineering View paper
- [3] A foundation model of transcription across human cell types View paper
- [4] Interpretable neural odes for gene regulatory network discovery under perturbations View paper
- [5] Predicting the unseen: a diffusion-based debiasing framework for transcriptional response prediction at single-cell resolution View paper
- [6] An integrated approach to predict activators of NRF2-the transcription factor for oxidative stress response View paper
- [7] PertEval-scFM: Benchmarking Single-Cell Foundation Models for Perturbation Effect Prediction View paper
- [8] TxPert: Leveraging Biochemical Relationships for Out-of-Distribution Transcriptomic Perturbation Prediction View paper
- [9] Transfer learning reveals sequence determinants of the quantitative response to transcription factor dosage View paper
- [10] Toward a foundation model of causal cell and tissue biology with a Perturbation Cell and Tissue Atlas View paper
- [11] Generative modeling of single-cell time series with PRESCIENT enables prediction of cell trajectories with interventions View paper
- [12] Decoding the Central Dogma: Quantitative Insights into Transcription and Translation Dynamics in the p53-Mediated DNA Damage Response View paper
- [13] NOTCH1 signaling during CD4+ T-cell activation alters transcription factor networks and enhances antigen responsiveness View paper
- [14] Dynamic modelling of pathways to cellular senescence reveals strategies for targeted interventions View paper
- [15] Vascular smooth muscle cell atherosclerosis trajectories characterized at single cell resolution identify causal transcriptomic and epigenomic mechanisms of disease â□¦ View paper
- [16] Learning causal networks using inducible transcription factors and transcriptomeâ□□wide time series View paper
- [17] Comprehensive transcription factor perturbations recapitulate fibroblast transcriptional states View paper
- [18] Elucidating genes and transcription factors of human peripheral blood lymphocytes involved in the cellular response upon exposure to ionizing radiation for â□¦ View paper
- [19] A novel investigation into an E2F transcription factorâ□□related prognostic model with seven signatures for colon cancer patients View paper
- [20] Flexible modeling of regulatory networks improves transcription factor activity estimation View paper
- [21] Signal-dependent dynamics of transcription factor translocation controls gene expression View paper
- [22] Modeling and designing enhancers by introducing and harnessing transcription factor binding units. View paper
- [23] Genome-wide promoter responses to CRISPR perturbations of regulators reveal regulatory networks in Escherichia coli View paper
- [24] The manatee variational autoencoder model for predicting gene expression alterations caused by transcription factor perturbations View paper
- [25] A systematic comparison of single-cell perturbation response prediction models View paper
- [26] The MAX-interacting transcription factor network View paper
- [27] Defining an essential transcription factor program for naive pluripotency View paper
- [28] Cell-adhesion molecule expression prediction from transcription factor expression patterns View paper
- [29] Transcription factor enrichment analysis (TFEA) quantifies the activity of multiple transcription factors from a single experiment View paper
- [30] Logical modeling of lymphoid and myeloid cell specification and transdifferentiation View paper
- [31] Inferring TF activities and activity regulators from gene expression data with constraints from TF perturbation data View paper
- [32] Machine learning for perturbational single-cell omics View paper
- [33] SpecLoop predicts cell type-specific chromatin loop via transcription factor cooperation. View paper
- [34] A mini-review on perturbation modelling across single-cell omic modalities View paper
- [35] Identification of novel activators of the metal responsive transcription factor (MTF-1) using a gene expression biomarker in a microarray compendium View paper
- [36] Mechanistic Modeling the Role of MicroRNAs and Transcription Factors in Disease Progression. View paper

- [37] Transcription factors as important regulators of changes in behavior through domestication of gray rats: Quantitative data from RNA sequencing View paper
- [38] The multiple mechanisms that regulate p53 activity and cell fate View paper
- [39] Identifying the combinatorial control of signal-dependent transcription factors View paper
- [40] Altered transcription factor targeting is associated with differential peripheral blood mononuclear cell proportions in sarcoidosis View paper
- [41] STAN, a computational framework for inferring spatially informed transcription factor activity across cellular contexts View paper
- [42] RegVelo: gene-regulatory-informed dynamics of single cells View paper
- [43] Uncertainty-aware transcription factor activity and perturbation inference without additional training View paper
- [44] MAP kinase signaling specificity mediated by the LIN-1 Ets/LIN-31 WH transcription factor complex during C. elegans vulval induction View paper
- [45] A quantitative model of transcription factorâ□□activated gene expression View paper
- [46] Zinc sensing by metal-responsive transcription factor 1 (MTF1) controls metallothionein and ZnT1 expression to buffer the sensitivity of the transcriptome response to â□¦ View paper
- [47] Encoding and decoding cellular information through signaling dynamics View paper
- [48] Computational discovery of transcription factors associated with drug response View paper
- [49] TIGERi: modeling and visualizing the responses to perturbation of a transcription factor network View paper
- [50] The interaction of TEA domain transcription factor 4 (TEAD4) and Yes-associated protein 1 (YAP1) promoted the malignant process mediated by serum/glucocorticoid regulated kinase 1 (SGK1) View paper
- [51] Expanding horizons of CRISPR applications beyond genome editing View paper
- [52] High-content CRISPR screening View paper
- [53] Epigenome editing technologies for discovery and medicine View paper
- [54] Gene and RNA Editing: Revolutionary Approaches to Treating Diseases View paper
- [55] Massively parallel in vivo Perturb-seq reveals cell-type-specific transcriptional networks in cortical development View paper
- [56] Manipulating and studying gene function in human pluripotent stem cell models View paper
- [57] Programmable RNA tracking in live cells with CRISPR/Cas9 View paper
- [58] The transience of transient overexpression View paper
- [59] Efficient genetic perturbation of murine sensory neurons in vivo using CRISPR/Cas9 View paper
- [60] Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens View paper
- [61] Nonlinear transcriptional responses to gradual modulation of transcription factor dosage View paper
- [62] MultiMAP: dimensionality reduction and integration of multimodal data View paper
- [63] MIRA: joint regulatory modeling of multimodal expression and chromatin accessibility in single cells View paper
- [64] Biologically-Aware Multimodal Representation Learning Deciphers Single-Cell Functions and Dynamics View paper
- [65] Building, benchmarking, and exploring perturbative maps of transcriptional and morphological data. View paper
- [66] Predicting cellular responses to complex perturbations in highâ□□throughput screens View paper
- [67] Learning single-cell perturbation responses using neural optimal transport View paper
- [68] Benchmarking foundation cell models for post-perturbation RNA-seq prediction View paper
- [69] PerturBase: a comprehensive database for single-cell perturbation data analysis and visualization View paper
- [70] Predicting cellular responses to perturbation across diverse contexts with State View paper
- [71] Fast and scalable Wasserstein-1 neural optimal transport solver for single-cell perturbation prediction View paper
- [72] Perturbench: Benchmarking machine learning models for cellular perturbation analysis View paper
- [73] Multimodal Benchmarking of Foundation Model Representations for Cellular Perturbation Response Prediction View paper
- [74] A benchmark for prediction of transcriptomic responses to chemical perturbations across cell types View paper
- [75] Benchmarking AI Models for In Silico Gene Perturbation of Cells View paper