BRIDGE:A MULTIMODAL DATASET FOR BIOLOGICAL REPRESENTATION LEARNING ACROSS IMAGING AND GENE EXPRESSION

Anonymous authors
Paper under double-blind review

ABSTRACT

Cellular responses to perturbation include molecular and phenotypic changes that cannot be captured in a single assay. Multimodal approaches use two or more assays to capture complementary data on these changes from the same experimental setting. For example, single-cell RNA sequencing (scRNA-seq) provides detailed but destructive measurements of transcriptomic state, while live-cell imaging non-destructively measures phenotype changes. A multimodal dataset that uses both assays captures more features of the system. Here, we introduce a multimodal dataset combining time-series live-cell imaging at 20-min resolution with endpoint long-read scRNA-seq comprising 16,296 cells from experimentally matched populations of human fibroblasts. We transiently perturb cell-cycle synchronized fibroblasts using modified mRNA (mmRNA)-mediated MYOD1 overexpression, RNA interference (RNAi)-mediated PRRX1 suppression, and their combination. Since the imaging and sequencing measurements were obtained from different cells, direct cell-to-cell supervision is unavailable. We therefore formulate the integration as a cross-modal prediction problem, using cell-cycle state as a shared biological anchor across modalities. The resulting model predicts population-level gene programs from imaging-derived features, an improvement over prediction based on cell-cycle phase alone. Together, this dataset enables the development and benchmarking of methods that integrate time-resolved cellular phenotypes with transcriptomic states, while our results demonstrate the potential of non-destructive imaging for inferring transcriptional responses to cellular perturbations.

1 INTRODUCTION

Predicting how cell states change in response to perturbations is a central challenge in computational biology. This problem has motivated the development of machine-learning models that predict responses and model cell behavior more broadly (Cui et al., 2024; Bunne et al., 2023; Lotfollahi et al., 2019; Roohani et al., 2024; Rosen et al., 2026; Adduri et al., 2025), with potential applications in drug discovery and personalized therapeutics. While the detailed molecular state of a cell can be measured via single-cell RNA sequencing (scRNA-seq), the data are static and the assay destructive, precluding direct observation of how any individual cell evolves over time.

CRISPR-based screens coupled with scRNA-seq enable the profiling of thousands of genetic perturbations at single-cell resolution (Dixit et al., 2016; Datlinger et al., 2017; Przybyla & Gilbert, 2022; Replogle et al., 2022; Xiao et al., 2015). These datasets have enabled an alternative computational strategy to predict perturbation response by pretraining foundation models on large collections of single-cell transcriptomic data and applying these in zero shot settings or through fine-tuning on perturbation specific data (Theodoris et al., 2023; Cui et al., 2024; Hao et al., 2024). In zero shot applications, pretrained representations can be used to estimate the shift in state under in-silico perturbation conditions, whereas fine-tuned models directly predict the post-perturbation gene-expression. However, current benchmarks show that these representations provide limited and inconsistent improvements (Wenteler et al., 2024; Ahlmann-Eltze et al., 2025). More importantly, existing evaluations are restricted to static transcriptomic endpoints and do not test whether the representations

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

capture perturbation-induced changes, cell-cycle progression, time-resolved trajectories, or phenotypic heterogeneity.

Live-cell imaging offers a complementary view to offset this loss of temporal information. While time-series imaging captures nuclear morphology, proliferation, and cell-cycle dynamics from the same single cells over time, scRNA-seq provides detailed cell state measurements at individual time points. Connecting these modalities would enable the inference of molecular responses from observable cellular phenotypes. Although datasets combining morphological and transcriptomic measurements have begun to emerge, they generally rely on fixed cells or static imaging (Chen et al., 2022; Haghighi et al., 2022; Saunders et al., 2025). This integration remains difficult because imaging and sequencing typically differ in temporal resolution and cannot be paired at the single-cell level: imaging provides repeated measurements of the same living cells, while sequencing provides endpoint measurements from a separate set of cells sampled from the same experimental condition.

To address this data and evaluation gap, we introduce BRIDGE: Biological Representation Learning across Imaging and Gene Expression, a multi-modal perturbation dataset that combines time-series live-cell imaging of nuclear morphology and cell-cycle dynamics at 20-minute resolution with endpoint long-read scRNA-seq from experimentally matched populations of human fibroblasts. The dataset captures the cellular responses to overexpression of MYOD1, suppression of PRRX1, and their combination in the setting of cell reprogramming.

Because the imaging and sequencing measurements were obtained from different cells sampled from matched conditions, direct cell-level pairing was not possible. To integrate the two populations on the same experimental axis, we used cell-cycle state as the shared coordinate measured independently from the FUCCI reporter (Grant et al., 2018) in imaging data and cell-cycle associated gene expression in scRNA-seq data (Tirosh et al., 2016). Using this alignment, we formulated a benchmark for predicting predefined gene programs from endpoint imaging features and show that imaging-derived features improve prediction relative to cell-cycle phase alone. BRIDGE supports the development and evaluation of models that relate observable cellular phenotypes to molecular responses under perturbations.

1.1 Biological Model System

Cellular reprogramming is the experimental conversion of cells from one identity to another. The first successful demonstration of direct cell reprogramming was reported in the 1980s, when Harold Weintraub showed that the addition of the transcription factor (TF) MYOD1 converts fibroblast cells into skeletal muscle cells (Weintraub et al., 1989). In 2007, Shinya Yamanaka demonstrated that introducing the TFs OCT4, SOX2, KLF4, and MYC into human fibroblasts generated embryonic-like pluripotent stem cells that could be differentiated into many target cell types (Takahashi & Yamanaka, 2006). These studies established direct cell reprogramming and induced pluripotency as two major forms of TF-mediated cell reprogramming.

Both approaches show that altering a small number of TFs can substantially change cellular identity. MYOD1 is a canonical example of a master regulator, or a TF that controls a broad gene network. When MYOD1 is added to cells, it activates skeletal muscle genes and induces expression of the endogenous MYOD1 gene through autoregulation (Thayer et al., 1989).

Cellular reprogramming is well-suited to studying perturbation responses, as it produces heterogeneous responses, with cells following different trajectories that require paired phenotypic and molecular measurements to characterize (Cacchiarelli et al., 2018). mmRNA provides a transient, non-integrating method for expressing TFs while RNAi suppresses the transcripts of specific TFs. Unlike genome-integrating methods, these approaches enable gene activity to be perturbed without permanent modifications, making them well-suited for studying the dynamics of cellular responses. As a model system for studying cellular responses to transient TF perturbations, we use the well-characterized myogenic reprogramming system induced by MYOD1 activation in human fibroblasts (Liu et al., 2018). MYOD1 mmRNA has previously been used to convert fibroblasts toward a myoblast-like state (Preskey et al., 2016; Qabrati et al., 2023). To explore opposing regulatory perturbations, we extend this system with RNAi-mediated suppression of PRRX1, a master TF associated with fibroblast identity (Lee et al., 2022; Tomaru et al., 2014), alongside mmRNA-mediated MYOD1 activation and a combination of MYOD1 activation with PRRX1 suppression. This design lets us assess whether transient suppression of fibroblast-associated gene programs, together with

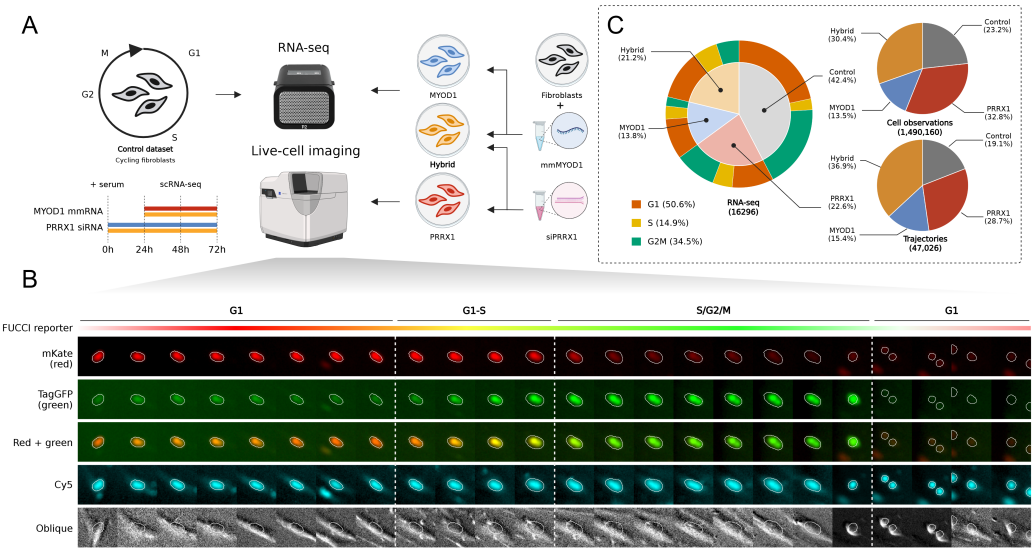

Figure 1: The BRIDGE Dataset: (A) Experimental design: Proliferating control fibroblasts were separated by Fluorescence Activated Cell Sorting (FACS) into G1, S, and G2/M populations for RNA sequencing. For the perturbation experiment, fibroblasts were synchronized by serum starvation for 24 hours, released and assigned to one of three conditions: MYOD overexpression using modified mRNA, PRRX1 suppression using siRNA, and the combined condition (hybrid). Cells were imaged by live-cell imaging for for 72 hours and cells from a subset of the wells were collected for single-cell RNA-seq (destructive) 48 hours after synchronization release. (B) Representative images of a tracked nucleus progressing through the cell-cycle and dividing. The FUCCI channels are mKate(red) and TagGFP(green). The Cy5 (blue) channel labels the nuclei and is used for nuclear segmentation and subsequently for cell tracking. The oblique-illumination channel captures cell morphology. (C) The dataset composition: The RNA-seq dataset consists of 16,296 cells, while the imaging dataset consists of 47,026 track segments from 1,503,069 segmented cell-frame observations

activation of the myogenic program, alters the efficiency and heterogeneity of reprogramming responses.

2 Related Work

2.1 Multimodal imaging and transcriptome datasets

Pooled CRISPR-based screens with transcriptome-wide readouts using scRNA-seq measure responses to a large number of perturbations and cells (Dixit et al., 2016; Datlinger et al., 2017; Replogle et al., 2022). Since conventional RNA-seq destroys the cell, these datasets provide only a static measurement of the cells at the time of measurement, imaging provides complementary phenotypic measurements. Datasets combining morphology with gene expression have enabled the comparison across conditions (Haghighi et al., 2022; Saunders et al., 2025).

Techniques, such as Live-seq, enable the temporal measurement of the transcriptomic response while keeping the cell alive (Chen et al., 2022). Spatial transcriptomic methods, instead, capture the gene expression at the exact position on the tissue (Ståhl et al., 2016). These approaches answer different questions from aligning live imaging trajectories with transcriptomic endpoints.

2.2 Perturbation prediction and foundation models

The growth of single-cell perturbation data has enabled a range of computational models that predict the transcriptional response of molecular or genetic perturbations. Approaches include differential expression and regulatory-network based methods, latent variable and generative models, and

Table 1: **Overview of related datasets and methods.** Checkmarks indicate available measurements, not same-cell pairing.

Dataset or method	RNA-seq	Perturbation	Imaging	Cell-cycle dynamics
Only imaging				
LIVECell (Edlund et al., 2021)	–	–	✓	–
Only RNA readout				
cellxgene (Program et al., 2025)	✓	–	–	–
Perturbation + RNA readout				
Perturb-seq (Dixit et al., 2016)	✓	✓	–	–
Replogle et al. (Replogle et al., 2022)	✓	✓	–	–
Imaging + RNA readout + perturbation				
Haghighi et al. (Haghighi et al., 2022)	✓	✓	✓	–
Perturb-Multimodal (Saunders et al., 2025)	✓	✓	✓	–
Live-seq (Chen et al., 2022)	✓	✓	✓	–
BRIDGE (ours)	✓	✓	✓	✓

perturbation-specific architectures (Lotfollahi et al., 2019; Bunne et al., 2023; Roohani et al., 2024). A complementary approach is to train foundation models on large amounts of transcriptomic collections and then later fine-tuning and linear-probing on perturbation-specific data (Theodoris et al., 2023; Cui et al., 2024; Hao et al., 2024).

Generalization to unseen perturbations and cellular contexts also remains challenging while recent evaluations have also found that complex models do not improve upon simple baselines (Ahlmann-Eltze et al., 2025; Viñas Torné et al., 2025; Wenteler et al., 2024). Benchmark conclusions also depend on datasets, splits, and metrics used. Most evaluations focus on transcriptomic endpoints, leaving out the ability of representations to describe temporally observed phenotypes.

2.3 Cross-modal representation learning

Learning representations and matching cellular images with gene expression is straightforward when both measurements are directly observed from the same cell (Xie et al., 2023). But, this matching is non-trivial when the assays sample different cells. Existing approaches include shared latent spaces (Zhang et al., 2026), contrastive learning (Xie et al., 2023), and optimal transport (Samaran et al., 2024) for measurements without direct pairs. The choice of method depends on the shared variable across assays. In BRIDGE, imaging and sequencing measure separate cells from matched experimental conditions. We use cell-cycle state, estimated separately from FUCCI reporters, and the transcriptomic expression as a shared biological coordinate between modalities. Our initial benchmark uses endpoint imaging features to predict population-level gene program scores.

3 Dataset Collection

3.1 Experimental overview

The BRIDGE dataset was designed to characterize the dynamical responses to transient perturbations using complementary measurement modalities. The experiment involves two data subsets. First, human neonatal fibroblasts were allowed to proliferate and were classified using Fluorescence Activated Cell Sorting (FACS) into G1, S or G2/M condition. Second, synchronized human neonatal fibroblasts were plated and assigned to control, MYOD1 overexpression, PRRX1 suppression, or

combined-perturbation conditions. These experimentally matched populations were then monitored using live-cell imaging and were barcoded for endpoint long-read single-cell RNA sequencing.

Here, Fibroblasts were cultured under standard conditions and synchronized under serum starvation for 24 hours at the G0/G1 phase. Cells were released from synchronization before applying the perturbation conditions. This approach was adopted to reduce the differences in initial cell-cycle composition of the cells and provide a coordinated reference between the imaging and sequencing populations.

3.2 MYOD1 AND PRRX1 PERTURBATIONS

The three different perturbation conditions were chosen to characterize the effects of reprogramming on cellular dynamics. MYOD1 and PRRX1 are both TFs that can regulate gene expression and have been shown to promote conversion efficiency. We use the individual and combinatorial effects of these two TFs to determine how cells react to such perturbations.

The MYOD1 TF is a well characterized master regulator in reprogramming of fibroblasts into the myogenic lineage, facilitating the switch from proliferation to differentiation and prompting proliferating myoblasts to exit the cell cycle and differentiate into mature muscle fibers. PRRX1 is a central gene in maintaining the fibroblast identity and promotes the myofibroblastic phenotype during wound healing via TGF-β signaling (Tomaru et al., 2014; Lee et al., 2022). For PRRX1 suppression, siRNA was introduced immediately after synchronization release, and mmRNA for MYOD1 (mmMYOD1) was added 24 hours after synchronization release Li et al. (2019); Steichen et al. (2014).

3.3 LIVE-CELL IMAGING AND ENDPOINT LONG-READ SCRNA-SEQ

Perturbed cells were imaged using a Zeiss Celldiscoverer 7 at 20-minute intervals for a total of 72 hours. The cells are transfected with the Incucyte Cell Cycle Reporter System. The imaging data is captured across 4 different channels: TagGFP, mKate, Cy5, and oblique. TagGFP and mKate serve as channels for the FUCCI reporters, and the Cy5 channel represents the cell nuclei, while the oblique channel provides the morphology and physical structure of the cell Grant et al. (2018).

The cell populations for sequencing were captured 48 hours after serum release. Expression measurements were obtained for cells from all 4 conditions. The resulting data provides the molecular endpoint measurements, complementing the dynamical phenotypes observed through live-cell imaging. The processed dataset consists of 16,296 sequenced cells together with time-series imaging data collected across four conditions.

4 DATA PROCESSING

4.1 IMAGE PROCESSING

The collected imaging data was acquired in two consecutive CZI files covering the same wells with a total of 204 frames per well. Each well corresponds to one experimental replicate and condition. The overlapping mosaic tiles were placed using stage coordinates and combined using cosine feathering. The Cy5 channel is the nuclear reporter and was used as segmentation channel for StarDist(Weigert & Schmidt, 2022). For each segmented cell, the corresponding 32 x 32 multichannel crops were retained for further processing.

The segmented nuclei were independently tracked using bayesian track (Ulicna et al., 2021) with motion and visual updates to recover the cell trajectories. This tracking used the nuclear area and the mean intensities of Cy5, TagGFP, and MKate channels after population-level normalization within each well. The procedure produced 47,026 track segments from 1,503,069 segmented cell-frame observations. Static and dynamic features were captured for every cell in every frame.

4.2 RNA-SEQ PROCESSING

The raw long-read single cell transcriptomic data was processed using the EPI2ME wf-single-cell (epi2melabs, 2026) pipeline with the human GRCh38 reference genome. The processed matrices

from the control experiment and the multiplexed perturbation experiments were combined into a common AnnData object containing cells as rows and genes as columns. Perturbation conditions were assigned to individual well barcodes using a custom demultiplexing pipeline. Cells were subjected to gene-count, library-size, mitochondrial-content and doublet quality control procedures. The normalized log-expression values were stored for downstream analyses. After quality control, the dataset consists of 16,296 cells, split into 6,902 control cells, 3,683 siPRRX1 cells, 2,255 mmMYOD1 cells, 3,456 combined-perturbation cells, and 25,126 genes Table 4.

The shared transcriptomic state was determined by using the top 2,000 most variable genes from control cells only. A 15-dimensional PCA model was fitted to normalized control expression and project the cells from all four conditions, resulting in an unbiased gene selection and a common PCA basis.

4.3 Cell-cycle and progression inference

Cell-cycle phases were assigned using the Scanpy cell-cycle workflow function `score_gene_cell_cycle` which uses the canonical cell-cycle markers from (Tirosh et al., 2016) yielding the G1, S and G2/M labels. A median trajectory was calculated across the control trajectories and its orientation was compared with the end group trajectories. Every sequenced cell was assigned to its nearest point on the 100-position trajectory. This trajectory progress index was used as the inferred cell-cycle coordinate A.7.

The cell-cycle phase for imaging data was inferred from the tracked TagGFP and MKate reporter signals using a custom phase-calling pipeline. The reporter signals were normalized and smoothed within each track before phase calling. The FUCCI marker distinguishes G1, G1-S transition, S, G2/M, and mitotic states, with unresolved frames excluded from further downstream tasks. For further cross-modal analysis, imaging phases were mapped down to the three RNA categories of G1, S, and G2/M. Each imaging phase was also associated with a RNA-progress interval.

4.4 Cross-modal anchoring and data splits

Since the imaging and single-cell RNA-sequencing were performed on different cells without one-to-one matched samples, cross-modal alignment was performed at the population level using the shared biological variables: perturbation conditions, cell-cycle phase, and inferred cell-cycle progress between the two datasets. The RNA cells were assigned to G1, S, or G2/M phase and split into ten progress bins of equal width sections within the normalized TrajectoryNet (Tong et al., 2020) coordinate. Imaging-derived FUCCI labels were mapped to the same phases. The midpoints for the imaging progress were then represented by the same ten progress bins. This coordinate is interpreted as the phase and progression associated anchor value rather than a substitute for experimental time.

The RNA cells were divided into reference and evaluation subsets, with a $70:30$ split across each condition, phase, and progress bin. This split was fixed for all subsequent model experiments.

5 Benchmarking tasks

We benchmarked two complementary model classes. First, we evaluated whether pretrained single-cell foundation models and conventional dimension reduction baselines produced RNA representations that retained perturbation and cell-cycle information. Next, we evaluated models that directly predicted post-perturbation gene expression, including generalization to the combined perturbation condition which was kept unobserved. The tasks are reported separately as informative embedding does not guarantee an efficient perturbation-response predictor.

5.1 Benchmark design and evaluation

The data contains four measured conditions: control, siPRRX1, mmMYOD1, and the combined siPRRX1 and mmMYOD1. To maintain identical and class-balanced inputs across models, we construct a cohort of 1,000 cells, sampled per condition without replacement, and all models were evaluated on the same cell identities. We evaluated PCA, scVI (Lopez et al., 2018), Geneformer (Theodoris et al., 2023), UCE (Rosen et al., 2026), scGPT (Cui et al., 2024), scFoundation (Hao

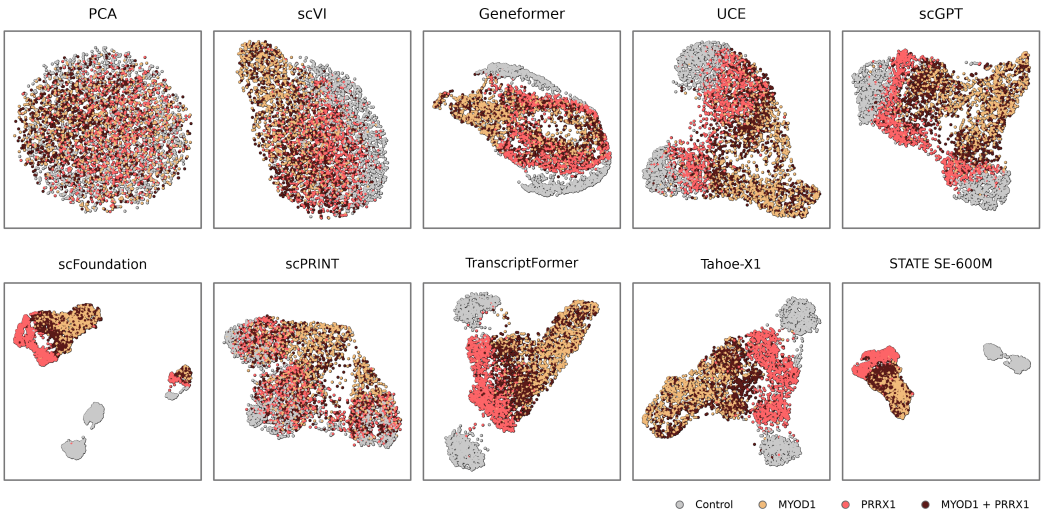

Figure 2: UMAP visualization of the shared, condition-balanced 4,000 cell cohort embedding using PCA, scVI, Geneformer, UCE, scGPT, scFoundation, scPRINT, TranscriptFormer, Tahoe-X1 and STATE SE 600M. Each representation was standardized and reduced to at most 50 Principal Components and projected with identical UMAP parameters. The cells are colored based on perturbation condition. Separation by condition indicates that the representation maintains the condition specific structure. UMAP is used as a qualitative visualization and not as a ranking metric. The quantitative assessment was performed using a linear probe balanced accuracy and additivity metrics. Note: UMAP coordinates, cluster sizes and distances cannot be compared across models directly.

et al., 2024), scPRINT (Kalfon et al., 2025), TranscriptFormer (Pearce et al., 2026), Tahoe-X1 (Gandhi et al., 2025) and STATE SE-600M (Adduri et al., 2025). Each emebedding was standardized and reduced to, at most, 50 components before evaluation. We used five-fold stratified sampling with a fixed seed. Table 2 reports the aggregate out-of-fold accuracy together with the sample standard deviation of balanced accuracy across five folds. The folds were chosen on cells rather than on the replicates and the SD is the model-validation variability.

The perturbation benchmark compared CellOT (Bunne et al., 2023), CPA (Lotfollahi et al., 2023), GEARS (Roohani et al., 2024), scGen (Lotfollahi et al., 2019), scGPT (Cui et al., 2024) perturbation models, and scFoundation (Hao et al., 2024) + GEARS. Predictions were evaluated for PRRX1 and MYOD1 conditions and the combined perturbation using a shared output schema of predicted expression, observed expression, condition labels, and gene identifiers. Each condition contained 300 predicted and 300 observed cells. The scFoundation + GEARS was evaluated on 19,264 genes, while the other models were evaluated on a common 4000 highly variable genes (HVG) as the gene space. This space included both the MYOD1 and the PRRX1 genes within the rank list. All model runs used one fixed training seed. The mean and SD in Table 3 are calculated across the three conditions.

5.2 Perturbation state prediction

We quantified perturbation information in each RNA representation using a four class linear probe for control, PRRX1, MYOD1 and the combined perturbation. Balanced accuracy averages recall across the four classes with a four class expected chance of approximately 0.25. Macro-F1 was retained as a secondary metric. Here, a high score indicates that the measured condition is linearly separable in the representation, but it does not show that the model can predict counterfactual perturbation response.

STATE SE-600M achieved the highest perturbation accuracy of (0.915 ± 0.013), followed by scFoundation (0.88 ± 0.010), TranscriptFormer (0.865 ± 0.013) and Tahoe-X1 (0.855 ± 0.009). The conventional baselines scVI and PCA achieved (0.653 ± 0.021) and (0.593 ± 0.013) respectively.

Sequencing source and condition are structurally confounded, these values measure condition separability and cannot be interpreted as batch-independent perturbation generalization. Additionally, we measure the geometry of the combined perturbation as outlined in A.8

5.3 CELL-CYCLE STATE PREDICTION

To measure the cell-state information independent from the perturbation labels, we train a second linear probe to classify the three annotated RNA cell-cycle states as G1, S and G2/M. The probe uses the same representation and five-fold startified protocol as perturbation-state prediction. Balanced accuracy has a three-class expected chance of approximately 0.333.

scFoundation achieved the highest cell-cycle balanced accuracy (BA) (0.930 ± 0.004), followed by Geneformer (0.921 ± 0.007), STATE SE-600M (0.918 ± 0.007), and TranscriptFormer (0.907 ± 0.010). The baselines scVI and PCA achieved (0.671 ± 0.029), (0.605 ± 0.021), respectively. High cell-cycle decoding means that a model preserves the biological cell cycle coordinate. Depending on the downstream task, the same signal may contain a useful state or correspond to a different biological axis.

5.4 PERTURBATION-RESPONSE PREDICTION

We evaluated the predicted expression changes after subtracting the measured control mean from all conditions to remove any baseline cell activity. For each condition, the Δ Pearson is the Pearson correlation between the predicted and observed control-subtracted mean expression vectors. Δ RMSE measures the corresponding magnitude error. Since baseline expression in the uncentered conditions could dominate the correlations, we use the control subtracted metrics as the primary mean.

To evaluate population structures, we jointly projected 30 components of the predicted and observed cells and computed energy distance. Lower values indicate closer predicted and observed distributions. We also recorded the predicted entries outside the observed expression range as a stability criteria to penalize the metrics for failing to recover cellular heterogeneity. This was done to restrict the model from predicting a single mean vector to obtain mean-response score. CellOT obtained the best aggregate rank, with mean ΔPearson 0.950 ± 0.070, mean ΔRMSE 0.081 ± 0.097, and energy distance 2.184 ± 2.395. scGen ranked second (Δ Pearson 0.923 ± 0.062; ΔRMSE 0.080 ± 0.048). GEARS achieved mean Δ Pearson 0.866 ± 0.054 but had larger magnitude and distributional errors.

The combined PRRX1 and MYOD1 condition was absent from model training and served as the held-out composition test. On this condition, scGPT had the smallest held-out magnitude error, whereas CellOT and scGen had the highest direction agreement. The scFoundation + GEARS RMSE is reported but should not be directly compared as it was calculated over 19,264 genes compared to 4,000 genes for the other models.

6 BRIDGE MODEL

BRIDGE was designed to estimate the population average transcriptional program activity from the time-resolved imaging measurements. Imaging features were reduced to the 16 highest ranked features for both current frame and previous frame dynamics (details in supplement **??**) These selected features were passed to a multilayer perceptron consisting of input layer norm, a 128-unit GELU layer, 15% dropout and a four unit output layer.

The target for the BRIDGE model is the averaged scores for prespecified gene programs following the UCell (Andreatta & Carmona, 2021) approach and a custom standardized mean scoring criteria (Details in Supplement A.11).

6.1 COMPARISON WITH BASELINES

We compared BRIDGE with several prespecified baselines. The phase/progress-only baseline received the cell-cycle covariates but no imaging features. The current-frame model received selected instantaneous imaging features, whereas the dynamic model only received causal temporal summaries. A combined model received both current and dynamic features and an all-current model

Table 2: Foundation model representation benchmark

Model	Perturbation BA (± fold SD) ↑	Cell-state BA (± fold SD) ↑	Double additivity cosine ↑	Double relative error ↓
STATE SE-600M	**0.915 ± 0.013**	0.918 ± 0.007	0.957	0.946
scFoundation	0.888 ± 0.010	**0.930 ± 0.004**	0.939	1.048
UCE	0.788 ± 0.008	0.854 ± 0.013	0.950	0.874
TranscriptFormer	0.865 ± 0.013	0.907 ± 0.010	0.947	1.074
Geneformer	0.847 ± 0.011	0.921 ± 0.007	0.942	1.122
scGPT	0.826 ± 0.016	0.904 ± 0.015	0.923	0.945
Tahoe-X1	0.855 ± 0.009	0.906 ± 0.019	0.908	0.978
scVI	0.653 ± 0.021	0.671 ± 0.029	0.947	1.044
scPRINT	0.666 ± 0.020	0.592 ± 0.020	**0.976**	1.245
PCA	0.593 ± 0.013	0.605 ± 0.021	0.870	**0.665**

Table 3: Perturbation prediction benchmark scores

Model	Δ Pearson (mean ± SD) ↑	Δ RMSE (mean ± SD) ↓	Energy distance (mean ± SD) ↓	Unseen-double Δ Pearson ↑	Unseen-double Δ RMSE ↓
CellOT	$\mathbf{0.950 \pm 0.070}$	0.081 ± 0.097	$\mathbf{2.184 \pm 2.395}$	**0.870**	0.193
scGen	0.923 ± 0.062	$\mathbf{0.080 \pm 0.048}$	4.569 ± 3.436	0.853	0.135
scGPT	0.885 ± 0.065	0.090 ± 0.029	8.482 ± 2.149	0.810	0.121
CPA	0.645 ± 0.381	0.130 ± 0.068	7.642 ± 4.512	0.206	0.202
GEARS	0.866 ± 0.054	0.141 ± 0.071	14.189 ± 5.958	0.805	0.200
scFoundation +GEARS	0.774 ± 0.123	0.084 ± 0.025	17.850 ± 4.917	0.640	**0.090**

received all 235 current frame measurements. As negative control, imaging features were shuffled among observations with the same phase and bin, separately within the training and evaluation sets. This preserves cell-cycle composition and the distribution of every feature while shuffling its association with perturbation identity. Finally, a condition-aware oracle with the known perturbation identity was used in place of the probabilities. This quantified the performance available from the RNA reference atlas when condition was perfectly known.

Across five seeds on held-out wells, BRIDGE's 16-feature dynamic-morphology MLP achieved macro (R^2=0.364) (MSE =0.0464) for standardized-mean scores and (R^2 = 0.407) (MSE = 0.00281) for UCell scores, reducing error by 20.5% and 21.2%, respectively, relative to phase and progress-only baselines. Detailed model performance and ablation study results can be found in the supplement A.14

6.2 Limitations

The BRIDGE dataset includes effects of three unique transcription factor perturbations on skin fibroblasts. Although this describes the activity of a small subset in detail, and introduces a new paradigm to understand perturbations, the vast combinatorial space of transcription factors and cell types remains largely unexplored.

The BRIDGE model provides a population-level gene-program estimate and does not infer the transcriptome of an individual imaged cell. The connection between modalities depends on predefined conditions, phase and the progress along their trajectories. The model also cannot predict RNA response to unseen out-of-distribution perturbations without an appropriate RNA reference or additional generalization mechanism.

AI USE STATEMENT

We used generative AI tools for implementation of methods, cleaning datasets, and writing code. Additionally, AI was used to improve the readability of the paper. AI was not used to design the experiments or in the interpretation of results. We have reviewed all AI-assisted work by manual checks at every stage of implementation and execution of code and in validation. We take responsibility for the final content of the work, including text, claims and artifacts with the aid of generative AI.

ETHICS STATEMENT

All resources provided as part of this cell reprogramming study are strictly for research purposes only and should not be used in clinical settings and diagnostic procedures. No sensitive information is included in the dataset. With the aforementioned restrictions, we have not identified any potential adverse impacts from the HYPED dataset.

REPRODUCIBILITY STATEMENT

The full experimental setup and the processing methodologies have been provided in A.1 and A.2. The data includes replicates across both imaging and sequencing modality ensuring experimental reproducibility. The raw sequencing data and the imaging data along with the full processing pipelines have been generated and will be released as part of the full dataset. This includes the processed and segmented images and tracks along with a toy dataset that will allow researchers to explore and analyze the work without the need to download the full dataset.

REFERENCES

Abhinav K Adduri, Dhruv Gautam, Beatrice Bevilacqua, Mohsen Naghipourfar, Alishba Imran, Rohan Shah, Noam Teyssier, Rishi Verma, Christopher Carpenter, Basak Eraslan, et al. Predicting cellular responses to perturbation across diverse contexts with state. *Cell*, 2025.

Constantin Ahlmann-Eltze, Wolfgang Huber, and Simon Anders. Deep-learning-based gene perturbation effect prediction does not yet outperform simple linear baselines. *Nature Methods*, 22(8): 1657–1661, 2025.

Massimo Andreatta and Santiago J Carmona. Ucell: Robust and scalable single-cell gene signature scoring. 2021.

Volker Bergen, Marius Lange, Stefan Peidli, F Alexander Wolf, and Fabian J Theis. Generalizing rna velocity to transient cell states through dynamical modeling. *Nature biotechnology*, 38(12): 1408–1414, 2020.

Charlotte Bunne, Stefan G Stark, Gabriele Gut, Jacobo Sarabia Del Castillo, Mitch Levesque, Kjong-Van Lehmann, Lucas Pelkmans, Andreas Krause, and Gunnar Rätsch. Learning single-cell perturbation responses using neural optimal transport. *Nature methods*, 20(11):1759–1768, 2023.

Davide Cacchiarelli, Xiaojie Qiu, Sanjay Srivatsan, Anna Manfredi, Michael Ziller, Eliah Overbey, Antonio Grimaldi, Jonna Grimsby, Prapti Pokharel, Kenneth J Livak, et al. Aligning single-cell developmental and reprogramming trajectories identifies molecular determinants of myogenic reprogramming outcome. *Cell systems*, 7(3):258–268, 2018.

Yi Cao, Zizhen Yao, Deepayan Sarkar, Michael Lawrence, Gilson J Sanchez, Maura H Parker, Kyle L MacQuarrie, Jerry Davison, Martin T Morgan, Walter L Ruzzo, et al. Genome-wide myod binding in skeletal muscle cells: a potential for broad cellular reprogramming. *Developmental cell*, 18(4):662–674, 2010.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *2021 IEEE/CVF international conference on computer vision (ICCV)*, pp. 9630–9640. IEEE, 2021.

Wanze Chen, Orane Guillaume-Gentil, Pernille Yde Rainer, Christoph G Gäbelein, Wouter Saelens, Vincent Gardeux, Amanda Klaeger, Riccardo Dainese, Magda Zachara, Tomaso Zambelli, et al. Live-seq enables temporal transcriptomic recording of single cells. *Nature*, 608(7924):733–740, 2022.

J Choi, ML Costa, CS Mermelstein, C Chagas, S Holtzer, and H Holtzer. Myod converts primary dermal fibroblasts, chondroblasts, smooth muscle, and retinal pigmented epithelial cells into striated mononucleated myoblasts and multinucleated myotubes. *Proceedings of the National Academy of Sciences*, 87(20):7988–7992, 1990.

Stephen T Crews and Joseph C Pearson. Transcriptional autoregulation in development. *Current Biology*, 19(6):R241–R246, 2009.

David Croft, Gavin O'kelly, Guanming Wu, Robin Haw, Marc Gillespie, Lisa Matthews, Michael Caudy, Phani Garapati, Gopal Gopinath, Bijay Jassal, et al. Reactome: a database of reactions, pathways and biological processes. *Nucleic acids research*, 39(suppl_1):D691–D697, 2010.

Haotian Cui, Chloe Wang, Hassaan Maan, Kuan Pang, Fengning Luo, Nan Duan, and Bo Wang. scgpt: toward building a foundation model for single-cell multi-omics using generative ai. *Nature methods*, 21(8):1470–1480, 2024.

Paul Datlinger, André F Rendeiro, Christian Schmidl, Thomas Krausgruber, Peter Traxler, Johanna Klughammer, Linda C Schuster, Amelie Kuchler, Donat Alpar, and Christoph Bock. Pooled crispr screening with single-cell transcriptome readout. *Nature methods*, 14(3):297–301, 2017.

Robert L Davis, Harold Weintraub, and Andrew B Lassar. Expression of a single transfected cdna converts fibroblasts to myoblasts. *Cell*, 51(6):987–1000, 1987.

Atray Dixit, Oren Parnas, Biyu Li, Jenny Chen, Charles P Fulco, Livnat Jerby-Arnon, Nemanja D Marjanovic, Danielle Dionne, Tyler Burks, Raktima Raychowdhury, et al. Perturb-seq: dissecting molecular circuits with scalable single-cell rna profiling of pooled genetic screens. *cell*, 167(7): 1853–1866, 2016.

Christoffer Edlund, Timothy R Jackson, Nabeel Khalid, Nicola Bevan, Timothy Dale, Andreas Dengel, Sheraz Ahmed, Johan Trygg, and Rickard Sjögren. Livecell—a large-scale dataset for label-free live cell segmentation. *Nature methods*, 18(9):1038–1045, 2021.

epi2melabs. wf-single-cell, 2026. URL `https://github.com/epi2me-labs/wf-single-cell`.

Shreshth Gandhi, Farnoosh Javadi, Valentine Svensson, Umair Khan, Matthew G Jones, John Yu, Daniele Merico, Hani Goodarzi, and Nima Alidoust. Tahoe-x1: Scaling perturbation-trained single-cell foundation models to 3 billion parameters. *bioRxiv*, pp. 2025–10, 2025.

Gavin D Grant, Katarzyna M Kedziora, Juanita C Limas, Jeanette Gowen Cook, and Jeremy E Purvis. Accurate delineation of cell cycle phase transitions in living cells with pip-fucci. *Cell cycle*, 17(21-22):2496–2516, 2018.

Marzieh Haghighi, Juan C Caicedo, Beth A Cimini, Anne E Carpenter, and Shantanu Singh. High-dimensional gene expression and morphology profiles of cells across 28,000 genetic and chemical perturbations. *Nature methods*, 19(12):1550–1557, 2022.

Minsheng Hao, Jing Gong, Xin Zeng, Chiming Liu, Yucheng Guo, Xingyi Cheng, Taifeng Wang, Jianzhu Ma, Xuegong Zhang, and Le Song. Large-scale foundation model on single-cell transcriptomics. *Nature methods*, 21(8):1481–1491, 2024.

Jérémie Kalfon, Jules Samaran, Gabriel Peyré, and Laura Cantini. scprint: pre-training on 50 million cells allows robust gene network predictions. *Nature Communications*, 16(1):3607, 2025.

Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E Kastriti, Peter Lönnerberg, Alessandro Furlan, et al. Rna velocity of single cells. *Nature*, 560(7719):494–498, 2018.

Keun-Woo Lee, So-Young Yeo, Jeong-Ryeol Gong, Ok-Jae Koo, Insuk Sohn, Woo Yong Lee, Hee Cheol Kim, Seong Hyeon Yun, Yong Beom Cho, Mi-Ae Choi, et al. Prrx1 is a master transcription factor of stromal fibroblasts for myofibroblastic lineage progression. *Nature Communications*, 13(1):2793, 2022.

Hanqin Li, Houbo Jiang, Xinzhen Yin, Jonathan E Bard, Baorong Zhang, and Jian Feng. Attenuation of prrx2 and hey2 enables efficient conversion of adult human skin fibroblasts to neurons. *Biochemical and biophysical research communications*, 516(3):765–769, 2019.

Arthur Liberzon, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P Mesirov, and Pablo Tamayo. The molecular signatures database hallmark gene set collection. *Cell systems*, 1(6): 417–425, 2015.

Sijia Liu, Haiming Chen, Scott Ronquist, Laura Seaman, Nicholas Ceglia, Walter Meixner, Pin-Yu Chen, Gerald Higgins, Pierre Baldi, Steve Smale, et al. Genome architecture mediates transcriptional control of human myogenic reprogramming. *Iscience*, 6:232–246, 2018.

Romain Lopez, Jeffrey Regier, Michael B Cole, Michael I Jordan, and Nir Yosef. Deep generative modeling for single-cell transcriptomics. *Nature methods*, 15(12):1053–1058, 2018.

Mohammad Lotfollahi, F Alexander Wolf, and Fabian J Theis. scgen predicts single-cell perturbation responses. *Nature methods*, 16(8):715–721, 2019.

Mohammad Lotfollahi, Anna Klimovskaia Susmelj, Carlo De Donno, Leon Hetzel, Yuge Ji, Ignacio L Ibarra, Sanjay R Srivatsan, Mohsen Naghipourfar, Riza M Daza, Beth Martin, et al. Predicting cellular responses to complex perturbations in high-throughput screens. *Molecular systems biology*, 19(6):MSB202211517, 2023.

Douglas P Millay, Jason R O'Rourke, Lillian B Sutherland, Svetlana Bezprozvannaya, John M Shelton, Rhonda Bassel-Duby, and Eric N Olson. Myomaker is a membrane activator of myoblast fusion and muscle formation. *Nature*, 499(7458):301–305, 2013.

Nanoporetech. Dorado, 2026. URL `https://github.com/nanoporetech/dorado`.

James D Pearce, Sara E Simmonds, Gita Mahmoudabadi, Lakshmi Krishnan, Giovanni Palla, Ana-Maria Istrate, Alexander Tarashansky, Benjamin Nelson, Omar Valenzuela, Donghui Li, et al. Transcriptformer: A generative cell atlas across 1.5 billion years of evolution. *Science*, 393 (6806):aec8514, 2026.

Bennett H Penn, Donald A Bergstrom, F Jeffrey Dilworth, Eyal Bengal, and Stephen J Tapscott. A myod-generated feed-forward circuit temporally patterns gene expression during skeletal muscle differentiation. *Genes & development*, 18(19):2348–2353, 2004.

David Preskey, Thomas F Allison, Mark Jones, Kamel Mamchaoui, and Christian Unger. Synthetically modified mrna for efficient and fast human ips cell generation and direct transdifferentiation to myoblasts. *Biochemical and biophysical research communications*, 473(3):743–751, 2016.

CZI Cell Science Program, Shibla Abdulla, Brian Aevermann, Pedro Assis, Seve Badajoz, Sidney M Bell, Emanuele Bezzi, Batuhan Cakir, Jim Chaffer, Signe Chambers, et al. Cz cellxgene discover: a single-cell data platform for scalable exploration, analysis and modeling of aggregated data. *Nucleic acids research*, 53(D1):D886–D900, 2025.

Laralynne Przybyla and Luke A Gilbert. A new era in functional genomics screens. *Nature Reviews Genetics*, 23(2):89–103, 2022.

Xhem Qabrati, Inseon Kim, Adhideb Ghosh, Nicola Bundschuh, Falko Noé, Andrew S Palmer, and Ori Bar-Nur. Transgene-free direct conversion of murine fibroblasts into functional muscle stem cells. *NPJ Regenerative Medicine*, 8(1):43, 2023.

Malgorzata E Quinn, Qingnian Goh, Mitsutoshi Kurosaka, Dilani G Gamage, Michael J Petrany, Vikram Prasad, and Douglas P Millay. Myomerger induces fusion of non-fusogenic cells and is required for skeletal muscle development. *Nature communications*, 8(1):15665, 2017.

Joseph M Replogle, Reuben A Saunders, Angela N Pogson, Jeffrey A Hussmann, Alexander Lenail, Alina Guna, Lauren Mascibroda, Eric J Wagner, Karen Adelman, Gila Lithwick-Yanai, et al. Mapping information-rich genotype-phenotype landscapes with genome-scale perturb-seq. *Cell*, 185(14):2559–2575, 2022.

Yusuf Roohani, Kexin Huang, and Jure Leskovec. Predicting transcriptional outcomes of novel multigene perturbations with gears. *Nature biotechnology*, 42(6):927–935, 2024.

Yanay Rosen, Yusuf Roohani, Ayush Agrawal, Leon Samotorčan, Tabula Sapiens Consortium, Stephen R Quake, and Jure Leskovec. Universal cell embedding provides a foundation model for cell biology. *Nature*, pp. 1–9, 2026.

Jules Samaran, Gabriel Peyré, and Laura Cantini. scconfluence: single-cell diagonal integration with regularized inverse optimal transport on weakly connected features. *Nature Communications*, 15 (1):7762, 2024.

Reuben A Saunders, William E Allen, Xingjie Pan, Jaspreet Sandhu, Jiaqi Lu, Thomas K Lau, Karina Smolyar, Zuri A Sullivan, Catherine Dulac, Jonathan S Weissman, et al. Perturb-multimodal: A platform for pooled genetic screens with imaging and sequencing in intact mammalian tissue. *Cell*, 188(17):4790–4809, 2025.

Patrik L Ståhl, Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, Michaela Asp, Jakub O Westholm, Mikael Huss, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353 (6294):78–82, 2016.

Clara Steichen, Eléanor Luce, Jérôme Maluenda, Lucie Tosca, Inmaculada Moreno-Gimeno, Christophe Desterke, Noushin Dianat, Sylvie Goulinet-Mainot, Sarah Awan-Toor, Deborah Burks, et al. Messenger rna-versus retrovirus-based induced pluripotent stem cell reprogramming strategies: analysis of genomic integrity. *Stem Cells Translational Medicine*, 3(6):686–691, 2014.

Kazutoshi Takahashi and Shinya Yamanaka. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *cell*, 126(4):663–676, 2006.

Mathew J Thayer, Stephen J Tapscott, Robert L Davis, Woodring E Wright, Andrew B Lassar, and Harold Weintraub. Positive autoregulation of the myogenic determination gene myod1. *Cell*, 58 (2):241–248, 1989.

Christina V Theodoris, Ling Xiao, Anant Chopra, Mark D Chaffin, Zeina R Al Sayed, Matthew C Hill, Helene Mantineo, Elizabeth M Brydon, Zexian Zeng, X Shirley Liu, et al. Transfer learning enables predictions in network biology. *Nature*, 618(7965):616–624, 2023.

Itay Tirosh, Benjamin Izar, Sanjay M Prakadan, Marc H Wadsworth, Daniel Treacy, John J Trombetta, Asaf Rotem, Christopher Rodman, Christine Lian, George Murphy, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell rna-seq. *Science*, 352(6282):189–196, 2016.

Yasuhiro Tomaru, Ryota Hasegawa, Takahiro Suzuki, Taiji Sato, Atsutaka Kubosaki, Masanori Suzuki, Hideya Kawaji, Alistair RR Forrest, Yoshihide Hayashizaki, FANTOM Consortium, et al. A transient disruption of fibroblastic transcriptional regulatory network facilitates transdifferentiation. *Nucleic acids research*, 42(14):8905–8913, 2014.

Alexander Tong, Jessie Huang, Guy Wolf, David Van Dijk, and Smita Krishnaswamy. Trajectorynet: A dynamic optimal transport network for modeling cellular dynamics. In *International conference on machine learning*, pp. 9526–9536. PMLR, 2020.

Kristina Ulicna, Giulia Vallardi, Guillaume Charras, and Alan R Lowe. Automated deep lineage tree analysis using a bayesian single cell tracking approach. *Frontiers in Computer Science*, 3: 734559, 2021.

Ramon Viñas Torné, Maciej Wiatrak, Zoe Piran, Shuyang Fan, Liangze Jiang, Sarah A Teichmann, Mor Nitzan, and Maria Brbić. Systema: a framework for evaluating genetic perturbation response prediction beyond systematic variation. *Nature Biotechnology*, pp. 1–10, 2025.

Isaac Virshup, Sergei Rybakov, Fabian J Theis, Philipp Angerer, and F Alexander Wolf. anndata: Access and store annotated data matrices. *Journal of Open Source Software*, 9(101):4371, 2024.

Martin Weigert and Uwe Schmidt. Nuclei instance segmentation and classification in histopathology images with stardist. In *2022 IEEE International Symposium on Biomedical Imaging Challenges (ISBIC)*, pp. 1–4. IEEE, 2022.

Harold Weintraub, Stephen J Tapscott, Robert L Davis, Mathew J Thayer, Mohammed A Adam, Andrew B Lassar, and A Dusty Miller. Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of myod. *Proceedings of the National Academy of Sciences*, 86(14):5434–5438, 1989.

Aaron Wenteler, Martina Occhetta, Nikhil Branson, Magdalena Huebner, Victor Curean, WT Dee, WT Connell, Alex Hawkins-Hooker, Siu Pui Chung, Yasha Ektefaie, et al. Perteval-scfm: benchmarking single-cell foundation models for perturbation effect prediction. *BioRxiv*, pp. 2024–10, 2024.

Yun Xiao, Yonghui Gong, Yanling Lv, Yujia Lan, Jing Hu, Feng Li, Jinyuan Xu, Jing Bai, Yulan Deng, Ling Liu, et al. Gene perturbation atlas (gpa): a single-gene perturbation repository for characterizing functional mechanisms of coding and non-coding genes. *Scientific reports*, 5(1): 10889, 2015.

Ronald Xie, Kuan Pang, Sai Chung, Catia Perciani, Sonya MacParland, Bo Wang, and Gary Bader. Spatially resolved gene expression prediction from histology images via bi-modal contrastive learning. *Advances in Neural Information Processing Systems*, 36:70626–70637, 2023.

Xinyi Zhang, GV Shivashankar, and Caroline Uhler. Partially shared multi-modal embedding learns holistic representation of cell state. *Nature Computational Science*, 6(3):285–300, 2026.

A Appendix

A.1 Experimental Setup

Human neonatal foreskin fibroblasts (BJ, ATCC CRL-2522) carrying the Incucyte Cell Cycle Reporter (Sartorius 4779) were used for perturbation experiments. Cells were cultured at 37 °C in 5 % CO_2 on standard cultureware in full media (FM; Dulbecco's Modified Eagle's Medium (DMEM, Gibco 11965-092) + 10 % Fetal Bovine Serum (FBS, Corning 35-015-CV) + 1% Penicillin-Streptomycin (P/S, Gibco 15140-122)). Cells were seeded at 1.03 - 1.33 x 10^4 cells/cm^2 in 6- and 48-well plates, followed by 24 hours of serum starvation (0.2% FBS) for G0/G1 phase synchronization. At t=0h, cells were released from serum starvation, and respective cells were transfected with PRRX1 siRNA (25 nM; Dharmacon) in FM using the TransIT-X2 Dynamic Delivery System (Mirus MIR6000). At t=24h, siRNA was washed out with FM. Respective cells were then transfected with MYOD1 modified mRNA (1 ng/μl; Pseudouridine + Silica purification, Trilink WOTL39876) in FM supplemented with 2 μM 4OH-Tamoxifen (Sigma-Aldrich H7904) using the Lipofectamine MessengerMAX Transfection Reagent (Invitrogen 100026485). At t=48h, cells were harvested with TrypLE Express (Gibco 12604-013) for scRNA-seq sample preparation. For imaging plates, cells were treated the same as described above with two differences: (1) Cells were cultured in imaging media (FluoroBrite DMEM (Gibco A18967-01) + 10% FBS + 1% P/S) instead of standard FM; (2) Cells were stained with 250 nM SiR-DNA (Cytoskeleton, Inc. CY-SC007) at t=-24h to enable tracking of cell nuclei.

A.2 Single Cell RNA Sequencing with Long Reads

Cells from siPRRX1, siPRRX1/mmMYOD1, and mmMYOD1 were individually labeled with TotalSeq-B human cell hashing antibodies (BioLegend, Cat# 394631, #394633, and #394635, respectively). 20,000 cells from each condition were mixed together (viability >98%) and single cell barcoded on the 10x Genomics Chromium Controller (X) using the Next GEM Single Cell 3' Kit V4. Barcoded cDNA amplicons from transcript and hashing libraries were prepared according to the Oxford Nanopore Technologies (ONT) 3' cDNA protocol (SQK-LSK114,

Table 4: Demultiplexing statistics for the hybrid-perturbation and control samples.

Demultiplexing statistic	**Hybrid**	**Control**
Total reads evaluated for barcode assignment	182,772,743	156,280,367
Exact barcode match	76.65%	74.49%
Corrected barcode match	8.79%	8.52%
Multiple barcode hits	2.60%	2.18%
No barcode hits	11.96%	14.80%
Valid cell barcodes	10,895	8,963
High-quality reads	134,381,035	102,222,758
High-quality reads assigned to valid barcodes	126,307,883	93,257,239
Demultiplexing accuracy	**94%**	**91%**

SST_9198_v114_revJ_13Nov2024). Libraries were assessed for quality following ONT recommendations. Prepared libraries were sequenced on the ONT PromethION Solo 2 (P2) sequencer. Raw reads were base-called with Dorado v0.9.1 Nanoporetech (2026) using the High Accuracy basecalling model and stored as fastq files.

For unperturbed controls, fibroblasts were sorted by fluorescence activated cell sorting (FACS) into their respective cell cycle phases after staining with 16 μM Hoechst 33342 for 50 minutes. Cells in G1, S, and G2/M were individually labeled with TotalSeq-B hashing antibodies (BioLegend, Cat# 394631, #394603, #394605, respectively), mixed, and single cell barcoded (viability $> 97\%$). Transcript and hashing libraries were processed and sequenced as described above.

Raw sequencing files obtained from scRNA-seq of the perturbed cells and controls were processed using the EPI2ME wf-single-cell pipeline v3.1.0 epi2melabs (2026) with the human GRCh38 reference genome. The processed gene expression matrices were combined into a single AnnData object containing cells as rows (obs) and genes as columns (var) Virshup et al. (2024).

Respective perturbation conditions (siPRRX1, siPRRX1/mmMYOD1, mmMYOD1) were assigned to individual cell barcodes using a custom demultiplexing pipeline. We add the $assigned_condition$ column to AnnData.obs.

A.3 Imaging channels

The Incucyte reporter is the commercial version of the PIP-FUCCI system, which fluorescently labels two regulatory proteins: Cdt1 and Geminin, associated with different phases of the cell cycle. Each protein is fused to a distinct fluorescent protein and the rate of degradation of these proteins can be imaged with fluorescence microscopy, and the position of the cell in the cell cycle can be accurately determined based on fluorescence intensity.

A.4 Sequencing Quality Control

For both scRNA-seq libraries (Hybrid + Control), demultiplexing was performed through the EPI2ME wf-single-cell pipeline. The Hybrid dataset had 94% of high-quality reads assigned to 10,895 valid single-cell barcodes. The Control dataset had 91% of high-quality reads assigned to 8,963 valid barcodes. See additional barcode assignment statistics on Table 4

Low quality cells were identified using a median absolute deviation (MAD)-based approach applied to each dataset. For each dataset, cells were flagged as outliers if their log-transformed total counts or number of detected genes were 5 MADs from the group median. Mitochondrial outliers were defined as cells with % mitochondrial counts 20% or 5 MADs from the group median. Cells flagged by any metric did not pass QC. 87% of Hybrid cells and 78% of Control cells were of sufficient quality. Additionally, low quality genes were identified as those with less than 10 counts, with 87.3% of genes passing this QC metric.

In the present study, transcriptional changes in the BRIDGE scRNA-seq dataset support successful induction of both perturbations. Specifically, while 0% of control cells expressed MYOD1, 82.9% of mmMYOD1-treated cells showed detectable MYOD1 expression. For PRRX1 knockdown, 30.6%

of control cells did not express PRRX1 compared to 66.7% of siPRRX1-treated cells, and 86.2% of siPRRX1-treated cells exhibited PRRX1 levels below the median PRRX1 expression observed in controls. Together, these data indicate robust uptake and activity of both mmRNA delivery and siRNA-mediated suppression

A.5 Exogenous Transcription Factor Quantification

Expression of master regulators like MYOD1 are tightly auto-regulated Crews & Pearson (2009). Externally introduced mmMYOD1 is translated into the MYOD protein, which increases the natural expression of the MYOD1 gene within the cell. Quantifying the ratio of external mmMYOD1 (exogenous) transcripts to the naturally produced MYOD1 (endogenous) transcripts enables us to provide a measure of reprogramming within the cells. The exogenous mmMYOD1 transcripts contain a unique sequence which localizes the mRNA to the Endoplasmic Reticulum (ER) which is absent in the endogenous transcripts. With long-read sequencing, the full RNA transcript including the ER sequence is captured for exogenously introduced transcripts.

To measure this quantification, we constructed a custom reference genome with the unique mmMYOD ER sequence as an exon within the MYOD1 gene and aligned the raw scRNA-seq reads to the custom reference genome. We also mapped each read back to the original cell it was expressed in and quantified the exogenous vs endogenous ratio at the single cell resolution. We propose this pipeline as an early measure of reprogramming from RNA-seq data.

A.6 Imaging feature extraction

Static features were calculated for every cell in every frame separately. Morphological measurements and the shape representations were included in the features. Within each nuclear mask, intensity summaries were captured for all four channels, including the mean, maximum, and sum. Additional oblique-channel features were calculated from a local area centered on the nucleus to capture cellular appearance beyond the nuclear boundary.

Dynamical features were extracted from the tracks computed in the processing step. For each trajectory, frame-to-frame displacement and total path length were computed. Causal summaries of features were captured over rolling windows of 3, 6, and 12 frames, with mean-squared-displacement computed at lags of 1, 2, 3, and 5 frames. Each feature was only computed with the current and the preceding observations, specifically to mitigate future information leakage. For image-based representation learning experiments, the retained cell crops were additionally encoded using a frozen DINO vision transformer (Caron et al., 2021).

A.7 RNA trajectory embedding

The RNA-seq data was used to train a TrajectoryNet (Tong et al., 2020) model that explains the progression of expression across time. The spliced and unspliced count matrices were generated using Velocyto (La Manno et al., 2018), and the RNA velocities for each cell were estimated using scVelo (Bergen et al., 2020). For the velocity-matched cells, the precomputed gene-space vectors were projected in the PCA space using the precomputed PCA loadings. A TrajectoryNet model was trained in this PCA space using the original labels together with the projected RNA-velocity vectors. The flow was evaluated at 100 distinct trajectory positions.

Table 5: Predefined gene programs used for RNA program scoring. Programs were identified using primary experimental literature and curated pathway resources.

No.	Program	Genes	Definition	Reference
P01	Early myogenic regulatory program	MYOD1, MYOG, MEF2D, MEF2C, CDH15, DES	Early genes involved in transition from fibroblast toward skeletal myogenic identity following MYOD activation.	(Davis et al., 1987)

Continued on next page

Table 5 – continued from previous page

No.	Program	Genes	Definition	Reference
P02	Late skeletal-muscle differentiation	DES, ACTA1, ACTN2, MYH3, MYH8, TNNT1, TNNI1, TNNI2, TPM1, TPM2, MYL1, CKM	Acquisition of differentiated skeletal-muscle identity.	(Choi et al., 1990)
P03	Sarcomere and myofibril organization	ACTN2, TTN, NEB, MYOM1, MYOM2, TCAP, LDB3, FLNC, ACTA1, TNNT1	Assembly and organization of sarcomeric and myofibrillar structures.	(Penn et al., 2004)
P04	Myoblast fusion competence	MYMK, MYMX, CDH15, DYSF, ANXA1, ANXA5	Membrane and cellular machinery associated with acquisition of myoblast fusion.	(Cao et al., 2010)
P05	Fibroblast and extracellular-matrix identity	COL1A1, COL1A2, COL3A1, COL5A1, COL5A2, FN1, DCN, LUM, FBLN1, SPARC, PDGFRA	Canonical fibroblast matrix-producing and stromal-identity program.	(Millay et al., 2013)
P06	PRRX1-associated myofibroblast / TGF-β state	PRRX1, ACTA2, TAGLN, POSTN, THBS1, COL1A1, COL3A1, FN1, TGFBR1, TGFBR2, SMAD2, SMAD3	PRRX1-associated stromal and myofibroblastic state linked to TGF-β signaling and extracellular-matrix remodeling.	(Quinn et al., 2017)
P07	Cell-cycle exit	CDKN1A, CDKN1B, RB1, RBL2, BTG1, BTG2	Growth arrest associated with conversion from a proliferating fibroblast to a myogenic state.	(Lee et al., 2022)
P08	S-phase	MCM5, PCNA, TYMS, FEN1, MCM2, MCM4, RRM1, RRM2, CDC45, RFC2	DNA-replication and S-phase transcriptional program.	(Liberzon et al., 2015)
P09	G2/M	HMGB2, CDK1, NUSAP1, UBE2C, BIRC5, TPX2, CENPF, CCNB1, CCNB2, TOP2A	Mitotic and G2/M transcriptional program.	(Liberzon et al., 2015)
P10	Actin–cytoskeletal remodeling / cell shape	ACTB, ACTG1, ACTN1, ACTN4, MYH9, MYH10, CFL1, PFN1, FLNA, VIM	Cytoskeletal organization expected to affect cellular and nuclear shape, spreading, and mechanical state.	(Liberzon et al., 2015)e
P11	Focal adhesion / integrin mechanotransduction	ITGB1, ITGA5, PXN, VCL, TLN1, PTK2, ZYX, FERMT2, ILK, PARVA	Cell–matrix adhesion, spreading, and integrin-associated mechanotransduction.	(Tirosh et al., 2016)
P12	Type-I interferon / innate RNA sensing	IFIT1, IFIT2, IFIT3, ISG15, MX1, OAS1, OAS2, DDX58, IFIH1, STAT1, IRF7	Innate immune response to exogenous or double-stranded-RNA-like stimuli.	
(Croft et al., 2010) P13	Oxidative / mitochondrial maturation	PPARGC1A, ESRRA, TFAM, COX5A, NDUFS1, NDUFV1, SDHA, UQCRC1, ATP5F1A, CPT1B	Oxidative metabolic remodeling associated with mature skeletal-muscle energetics.	(Liberzon et al., 2015)
P14	Senescence / maladaptive growth arrest	CDKN2A, SERPINE1, IGFBP7, GDF15, GLB1, IL6, CXCL8	Durable stress-associated growth arrest distinguished from productive myogenic cell-cycle exit.	(Liberzon et al., 2015)

A.8 PERTURBATION GEOMETRY

Let $\boldsymbol{\mu}_0$, $\boldsymbol{\mu}_P$, $\boldsymbol{\mu}_M$, and $\boldsymbol{\mu}_{PM}$ denote the control, PRRX1, MYOD1, and the combined-condition centroids respectively. The additive prediction is,

$$\widehat{\boldsymbol{\delta}}_{PM} = (\boldsymbol{\mu}_P - \boldsymbol{\mu}_0) + (\boldsymbol{\mu}_M - \boldsymbol{\mu}_0), \tag{1}$$

and the observed combined effect is $\boldsymbol{\delta}_{PM} = \boldsymbol{\mu}_{PM} - \boldsymbol{\mu}_0$. We report their cosine similarity, for which one indicates perfect directional alignment, and the relative error

$$\frac{\|\boldsymbol{\delta}_{PM} - \widehat{\boldsymbol{\delta}}_{PM}\|_2}{\|\boldsymbol{\delta}_{PM}\|_2}, \tag{2}$$

for which zero is optimal. These metrics must be interpreted jointly: a high cosine with relative error near one indicates that the direction is approximately correct but the magnitude is inaccurate.

A.9 BRIDGE FORMULATION

A.10 BRIDGE MODEL TEMPORAL IMAGING ENCODER

The imaging representation was constructed from the 235 manually defined features. FUCCI reporter measurements were used to infer cell-cycle state but were not included among the features used to predict perturbation condition. Each feature was centered by its training-set median and divided by the IQR. Non finite values were replaced by zero scaled values were clipped to $[-10, 10]$. Temporal information was computed causally from the current and three preceding valid observations within each cellular track. For every imaging measurement, five representations were computed: the current values, the four-frame mean, SD, linear slope and the change from first to last frame. The observations at the beginning of the track were repeated to get the four from window. No future frames were used in the current observation to prevent information leakage to the model.

From these dynamical features, the 16 highest ranked features were selected from the representations. These features were ranked based on their association with the gene program prediction targets, and were retained only when the direction agreed across the acquisitions.

For an observation i, let $\mathbf{d}_i \in \mathbb{R}^{16}$ denote the selected causal temporal features, calculated from the mean, standard deviation, slope, and change over the current and three preceding valid frames. The model input was

$$\mathbf{a}_i = [\mathbf{d}_i;\, \mathrm{onehot}(\phi_i);\, \mathrm{onehot}(b_i);\, \sin(2\pi\rho_i);\, \cos(2\pi\rho_i)]\,, \tag{3}$$

where ϕ_i is the imaging-derived cell-cycle phase, $\rho_i \in [0, 1]$ is cell-cycle progress, and b_i is its corresponding progress bin.

A.11 TRANSCRIPTOME REPRESENTATION

The transcriptomic output was represented as 14 prespecified gene programs. Program activity was calculated using two scoring methods. The first method involved the standardized mean score for each gene where the mean and SD were computed from the RNA reference cells and the values were averaged across the detected genes in the program. For the second method we followed the UCell (Andreatta & Carmona, 2021) approach where genes were ranked independently within each cell and program was calculated using the normalized Mann-Whitney formulation. Genes absent from the dataset were omitted. No PCA was used in either of the methods to calculate program score.

Each RNA-seq cell j is represented by a vector $\mathbf{y}_j \in \mathbb{R}^{14}$ containing the activities of 14 predefined gene programs.

The full dataset was divided into a 70:30 split stratified by perturbation condition, RNA phase and progress bin. For each scoring method separate evaluation atlases were constructed based on the stratification and averaging the scores within each condition. The gene programs were not used as a direct target for regression, instead the probabilities from the classifier were converted to predictions using the stratified RNA reference atlas. At evaluation, predictions from held-out wells were averaged within each strata and compared with the corresponding program means from the atlas.

A.12 TRAINING OBJECTIVES AND INFERENCE

For the imaging split complete wells were used as the splitting criteria during training. Every perturbation condition was included in both the training and test splits and no track was included in both sets. The neural network was trained to classify the four measured perturbation conditions using cross-entropy loss. To deal with unbalanced sampling we used a class-balanced sampling approach.

An MLP f_η produced probabilities over the four observed perturbation conditions:

$$\boldsymbol{\pi}_i = \mathrm{softmax}(f_\eta(\mathbf{a}_i)) \,. \tag{4}$$

The model was trained using the observed perturbation identity and class-balanced categorical cross-entropy:

$$\mathcal{L}_{\mathrm{cls}} = -\frac{1}{N_{\mathrm{train}}} \sum_{i=1}^{N_{\mathrm{train}}} w_{c_i} \log \pi_i(c_i), \tag{5}$$

where c_i is the true condition and w_{c_i} accounts for unequal condition frequencies.

Model was optimized with the AdamW for 30 epochs and a learning rate of 3×10^{-4} and a decay of 1×10^{-4} and batch size of 2,048. Training duration was fixed and evaluation wells were not used for checkpoint selection.

A.13 CROSS-MODAL ANCHORING

For each condition c, phase p, and progress bin b, a single RNA reference vector is defined as

$$\mathbf{A}^{\mathrm{ref}}_{c,p,b} = \frac{1}{|\Omega^{\mathrm{ref}}_{c,p,b}|} \sum_{j \in \Omega^{\mathrm{ref}}_{c,p,b}} \mathbf{y}_j, \tag{6}$$

where $\Omega^{\mathrm{ref}}_{c,p,b}$ contains the RNA reference cells in that stratum.

The predicted RNA program vector for an imaging observation is the probability-weighted combination

$$\widehat{\mathbf{y}}_i = \sum_{c \in \mathcal{C}} \pi_i(c)\, \mathbf{A}^{\mathrm{ref}}_{c,\phi_i,b_i}. \tag{7}$$

This provides a cell-cycle stratified, population level anchoring.

A.14 MODEL PERFORMANCE AND ABLATION STUDIES

The performance from five training seeds are reported in the table. Bridge with dynamic features reduced MSE by 20.5% relative to the phase/progress-only baseline under standardized scoring and by 21.2 % under UCell scoring. Its macro R^2 increased from 0.2425 to 0.3645 for standard scores and from 0.2379 to 0.4069 for UCell scores. Raw MSE values cannot be compared across the different scoring systems as they follow different scoring scales.

The temporal feature ablation showed that the past measurements contained useful information compared to current frame features alone. The 16-feature dynamic model achieved lower error than the 16-feature current-frame model under both scoring systems. Combining current and dynamic did not improve performance over dynamic features alone. Including all 235 features also reduced performance. These suggest that compact changes over time were more informative than either a large instantaneous feature set or the concatenation of all representations. The matched shuffling analysis tests for whether the imaging contribution can be explained by cell-cycle composition or model capacity. The real features outperformed their phase/progress-matched shuffled version in all 90 of the 90 paired configuration and seed comparisons under UCell scoring. The selected 16-feature dynamic MLP, shuffling degraded performance to the phase/progress-only baselines.

Table 6: The prediction performance of the BRIDGE model on held-out wells. Values are mean ± SD across five training seeds. The condition aware oracle is deterministic and has no cross seed variance.

	Standardized-mean scoring		UCell scoring	
Model	Macro MSE ↓	Macro R^2 ↑	Macro MSE ↓	Macro R^2 ↑
Phase/progress only	0.05843 ± 0.00002	0.2425 ± 0.0005	0.003566 ± 0.000001	0.2379 ± 0.0004
Phase/progress-matched shuffle	0.05842 ± 0.00004	0.2429 ± 0.0007	0.003565 ± 0.000002	0.2381 ± 0.0007
16 current-frame features	0.05074 ± 0.00063	0.3360 ± 0.0049	0.003091 ± 0.000013	0.3720 ± 0.0025
16 current + 16 dynamic features	0.05136 ± 0.00041	0.3351 ± 0.0028	0.003090 ± 0.000012	0.3736 ± 0.0024
All 235 current-frame features	0.05284 ± 0.00044	0.3233 ± 0.0040	0.003240 ± 0.000027	0.3527 ± 0.0047
BRIDGE: 16 dynamic features	$\mathbf{0.04644 \pm 0.00049}$	$\mathbf{0.3645 \pm 0.0036}$	$\mathbf{0.002809 \pm 0.000017}$	$\mathbf{0.4069 \pm 0.0028}$
Condition-aware RNA oracle	0.004917	0.7165	0.000174	0.8027

We also compared using 8, 16 or 32 features and with a linear classifier vs a MLP head. The 16-feature dynamic MLP gave the lowest average gene-program MSE under both scoring systems. The scoring methods also had high agreement of cell-level program-score Spearman correlation of 0.8849, and condition associated effects had a Spearman correlation of 0.957. Agreement indicates that result was not specific to the scoring procedure.