Cell Mechanics with AI

TFM and Single Cell RNASeq datasets

Amina Ghandouz

Contents

List of Figures

1 Abstract

Cell mechanics play a fundamental role in processes such as growth, differentiation, migration, and disease progression. Traditional methods for studying cellular mechanics rely heavily on experimental biophysics, which, while powerful, are often time-consuming and limited in scalability. In this project, we propose an interdisciplinary approach that combines **physics**, **biology**, and **computer science** to model cell mechanics using machine learning applied to genomics data.

By leveraging **single-cell RNA sequencing** and **traction force microscopy datasets**, we aim to uncover statistical relationships between gene expression patterns and the mechanical behavior of cells. Our methodology integrates data preprocessing, feature extraction, and the development of predictive models capable of linking genomic signatures with physical properties such as traction forces and stiffness. The expected outcome is a computational framework that not only enhances our understanding of how genetics drive cellular mechanics but also enables predictive insights into pathological conditions where mechanical regulation is disrupted, such as cancer and fibrosis.

This work demonstrates the potential of **artificial intelligence** to bridge **molecular data** with **physical cell behavior**, paving the way toward data-driven discoveries in **mechanobiology**.

2 Introduction

2.1 Background

Cell mechanics represent a fundamental aspect of **biology**, governing processes such as cell growth, migration, differentiation, and tissue morphogenesis. Alterations in mechanical properties have been closely linked to pathological conditions, including cancer metastasis and fibrosis, highlighting the need for quantitative approaches to study cellular mechanics. Traditionally, methods such as traction force microscopy **(TFM)**, atomic force microscopy **(AFM)**, and optical tweezers have been used to probe the physical properties of cells. While these experimental techniques provide valuable insights, they are often labor-intensive, low-throughput, and difficult to scale to large datasets.

Recent advances in genomics, particularly **single-cell RNA sequencing (scRNA-seq)**, have generated high-dimensional molecular data that capture the transcriptional state of individual cells. This opens new opportunities to integrate molecular signatures with physical phenotypes, bridging biology and physics at the cellular level. However, extracting meaningful relationships between gene expression and mechanical properties remains a challenging problem, due to the complexity and heterogeneity of both data types.

Machine learning and **artificial intelligence** offer powerful tools to address these challenges by uncovering hidden patterns and predictive relationships in high-dimensional, multimodal datasets. The convergence of **biology**, **physics**, and **AI** thus presents a promising pathway for developing computational frameworks that can predict and model cellular mechanics in a scalable and data-driven manner.

2.2 Problem Statement

Despite the critical role of **cell mechanics** in fundamental biological processes and disease progression, current approaches to studying these properties face important limitations. Experimental techniques such as traction force microscopy, atomic force microscopy, and optical tweezers provide precise measurements of mechanical forces, but they are typically slow, costly, and restricted to small sample sizes. This makes them unsuitable for large-scale or high-throughput studies. Furthermore, these methods often capture mechanical properties in isolated contexts and may not fully account for the molecular complexity underlying cell behavior.

On the other hand, genomics technologies such as **single-cell RNA sequencing** provide comprehensive molecular profiles of thousands of cells at once, but these data are rarely integrated with mechanical measurements. As a result, there is a missing link between gene expression patterns and physical cell states. This gap prevents the development of predictive, scalable models that can connect molecular signatures with biophysical properties.

The absence of robust computational frameworks that unify genomics data with cell mechanics creates a significant barrier to mechanobiology research. There is a pressing need for approaches that leverage artificial intelligence to bridge this divide, enabling predictive insights into how molecular programs govern cellular mechanical behavior.

2.3 Objectives

The main objective of this project is to develop a **computational framework** that **models cell mechanics** using **machine learning applied to genomics data**. More specifically, the project aims to:

- Establish a data pipeline that integrates **single-cell RNA sequencing data** with cell **mechanics datasets** such as **traction force microscopy**.
- Apply **machine learning techniques** to identify statistical relationships between gene expression patterns and mechanical phenotypes (e.g., stiffness, deformability and traction forces).
- Build predictive models capable of inferring mechanical behavior directly from molecular data.
- Evaluate the performance and interpretability of the models to identify key genomic features most correlated with mechanical states.
- Provide a proof-of-concept framework that demonstrates the potential of artificial intelligence to bridge molecular and physical domains in cell biology.

2.4 Scope

This project is designed as a proof-of-concept study that explores **the integration of genomics** and **machine learning** to **model cell mechanics**. The scope is intentionally focused and limited to ensure feasibility within the timeframe and resources available. Specifically, the project will:

- Use publicly available datasets, primarily **single-cell RNA sequencing (scRNA-seq)** and **traction force microscopy (TFM)** as the main sources of molecular and mechanical data.
- Develop preprocessing and machine learning pipelines to identify and model relationships between gene expression and cellular mechanical properties.
- Emphasize methodological exploration and proof-of-concept modeling rather than clinical application or large-scale deployment.
- Focus on predictive modeling and feature interpretation at the research level, without extending to experimental validation in wet-lab settings.
- Limit the study to a computational investigation of selected datasets, without addressing broader biological variability across tissues or species.

2.5 Significance

Understanding the relationship between molecular programs and mechanical behavior of cells is a key challenge in modern biology and medicine. By proposing a data-driven approach to model cell mechanics from genomics, this project contributes to bridging the gap between physics, biology, and artificial intelligence. The significance of this work can be outlined as follows:

- **Scientific contribution:** Provides a proof-of-concept framework for linking gene expression profiles to cellular mechanics, opening new avenues for mechanobiology research.
- **Methodological innovation:** Demonstrates how machine learning can integrate multi-modal datasets (genomics + mechanics) and extract predictive, interpretable insights.

- **Biomedical relevance:** Offers potential applications in understanding diseases where mechanics play a central role, such as cancer progression, tissue fibrosis, and metastasis.
- **Practical impact:** Suggests scalable, computational alternatives to experimental biophysics methods, reducing cost and enabling larger-scale analyses.
- **Interdisciplinary value:** Illustrates how combining biology, physics and computer science can generate novel perspectives and tools for complex biological problems.

3 Datasets

3.1 Single-cell RNA Sequencing (scRNA-seq) Data

3.1.1 Source

The dataset used in this project was obtained from Kaggle, originally derived from a study on U2OS osteosarcoma cells. It is provided in the h5ad format, commonly used for single-cell RNA sequencing (scRNA-seq) data, and processed through the AnnData framework.

3.1.2 Structure and Dimensions

- Number of cells (observations): 1,102
- Number of genes (variables): 10,000
- Data type: scRNA-seq counts with annotations
- Dimensionality reduction precomputed: PCA (obsm[$'X'_{pca}$])
- Additional layers: raw matrix, spliced/unspliced counts (useful for RNA velocity analysis).

3.1.3 Cell-Level Metadata (adata.obs)

Each cell is annotated with:

- Cell cycle phase: G2M, G1, S-phase, or undefined which will be removed later
- n_genes_by_counts: number of expressed genes per cell
- pct_counts_mt: percentage of mitochondrial gene expression (cell quality indicator)
- total_counts: total RNA counts per cell
- total_counts_mt: raw mitochondrial counts

3.1.4 Gene-Level Metadata (adata.var)

Each gene is annotated with:

- Accession ID (Ensembl ID)
- Chromosome location (Start, End, Strand)
- Mean counts per gene

- Dropout percentage (fraction of cells where the gene is not detected)
- mt flag indicating mitochondrial genes

3.1.5 Pre-processing:

Due the small number of undefined cell cycle phase which was only 84 from 1102 total cells ($\sim 7.6\%$) we did remove them.

3.1.6 Exploratory Data Analysis and Visualization:

To better understand the structure and quality of the dataset, several exploratory plots were generated.

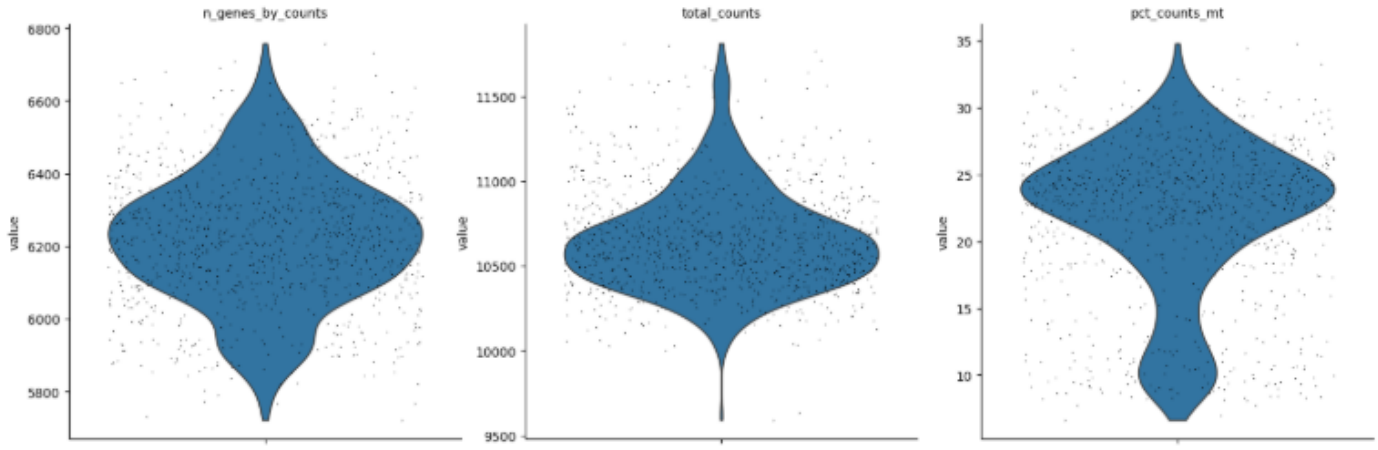

Figure 1: QC metric distributions for scRNA-seq dataset

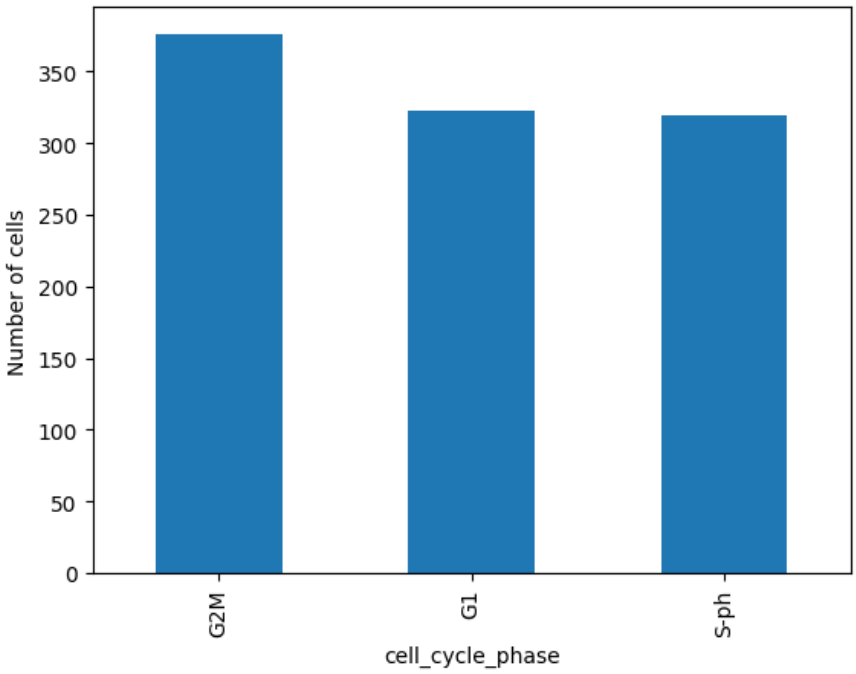

Figure 2: Cell distribution across cell cycle phases for scRNA-seq dataset

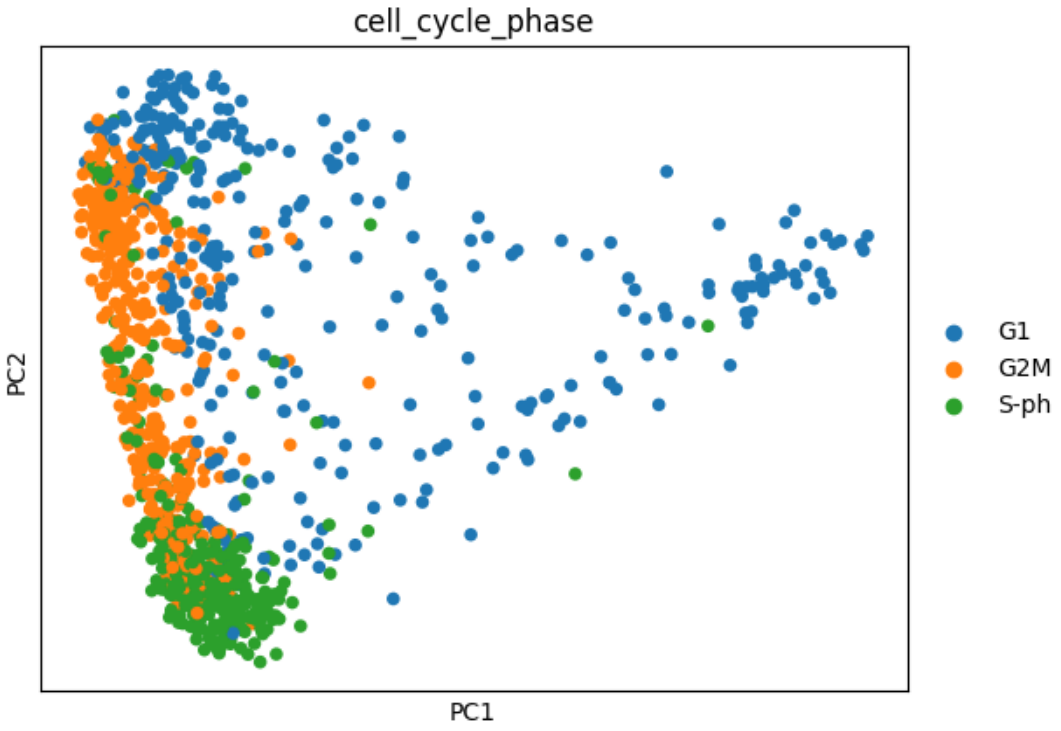

Figure 3: PCA projection of cells for scRNA-seq dataset

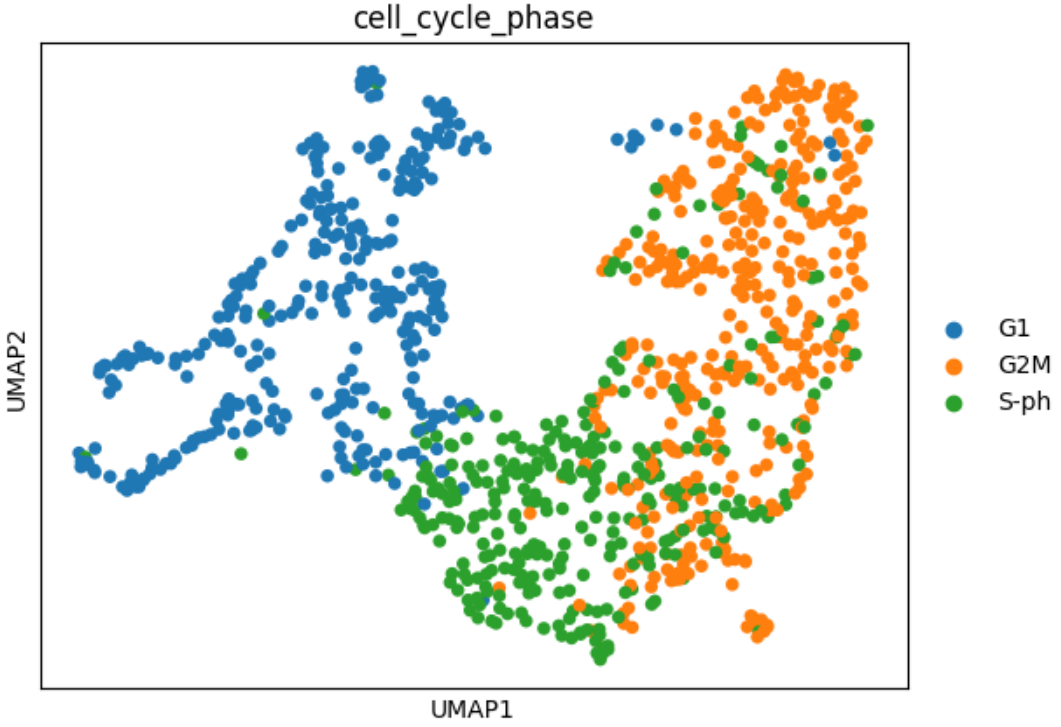

Figure 4: UMAP embedding of cells for scRNA-seq dataset

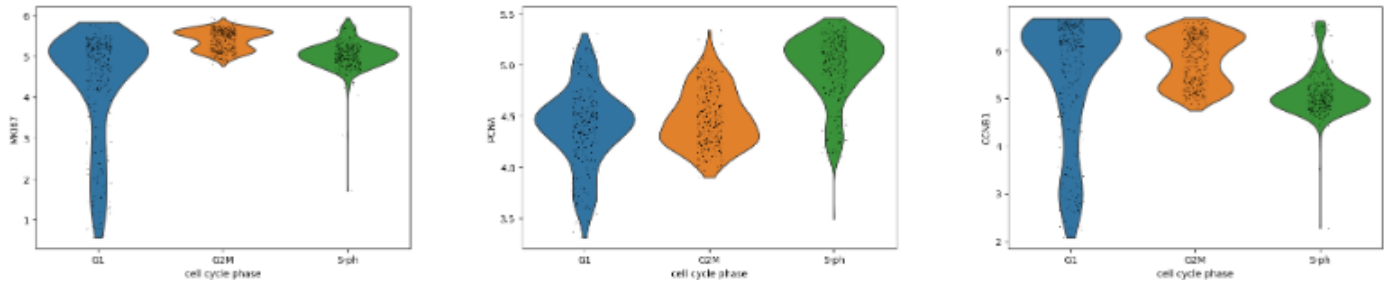

Figure 5: Expression of selected genes for scRNA-seq dataset

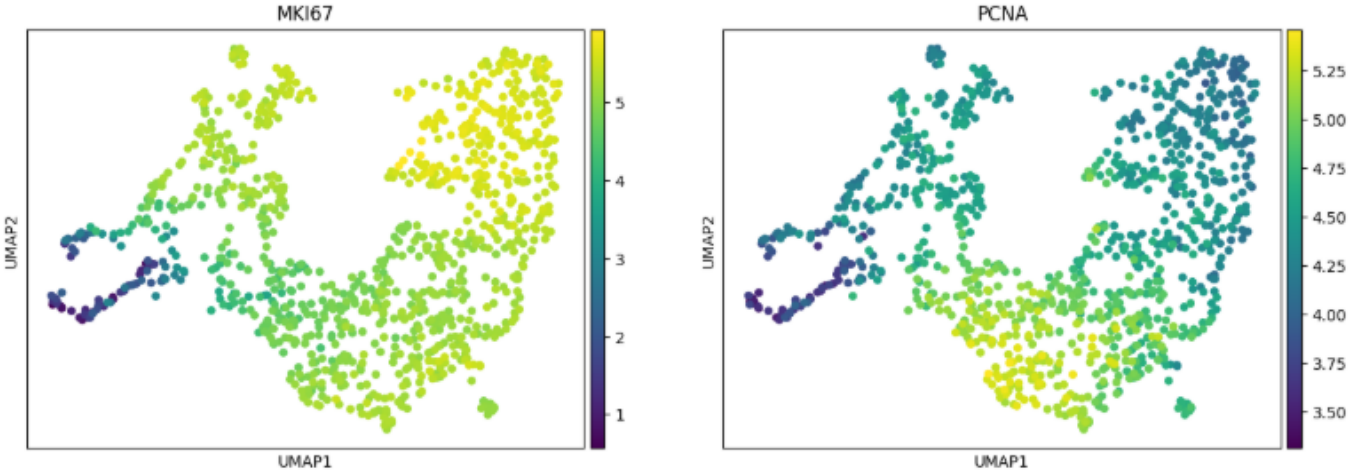

Figure 6: UMAP visualization colored by gene expression for scRNA-seq dataset

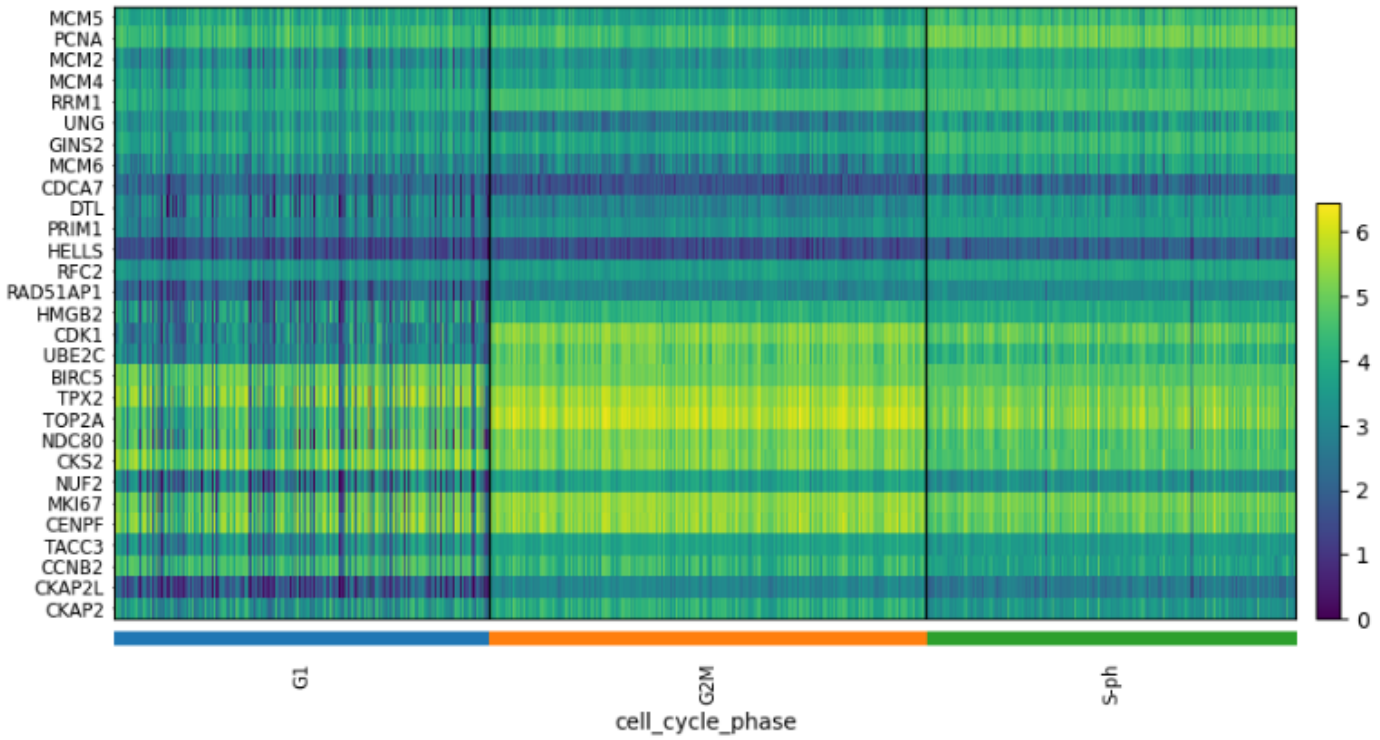

Figure 7: Heatmap of cell cycle gene expression for scRNA-seq dataset

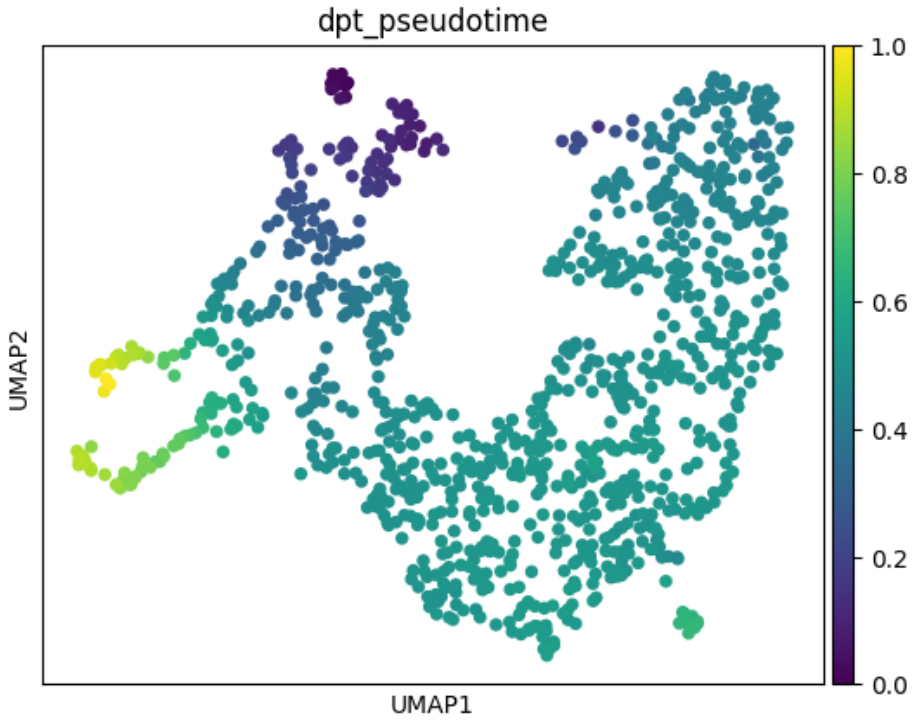

Figure 8: Trajectory-like cell cycle progression for scRNA-seq dataset