UE20CS390A - Capstone Project Approval

Project Title : Exploring single cell dynamics in response to chemical perturbations
Project ID : PW_GS_04
Project Guide : Dr Gowri Srinivasa
Project Team : 030_091_113_201

Ankush HV	PES1UG21CS091
Ganya Umesh	PES1UG21CS201
Arya Vinayak R	PES1UG21CS113
Adithi S Murthy	PES1UG21CS030

PES UNIVERSITY

Outline

- Problem Statement
- Scope and Feasibility study
- Applications/Use cases
- Expected Deliverables
- Capstone (Phase-I & Phase-II) Project Timeline
- Any other information

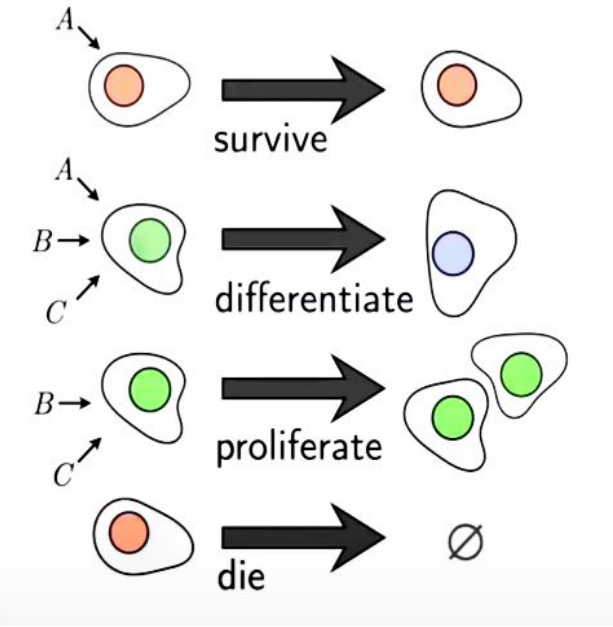
PES UNIVERSITY

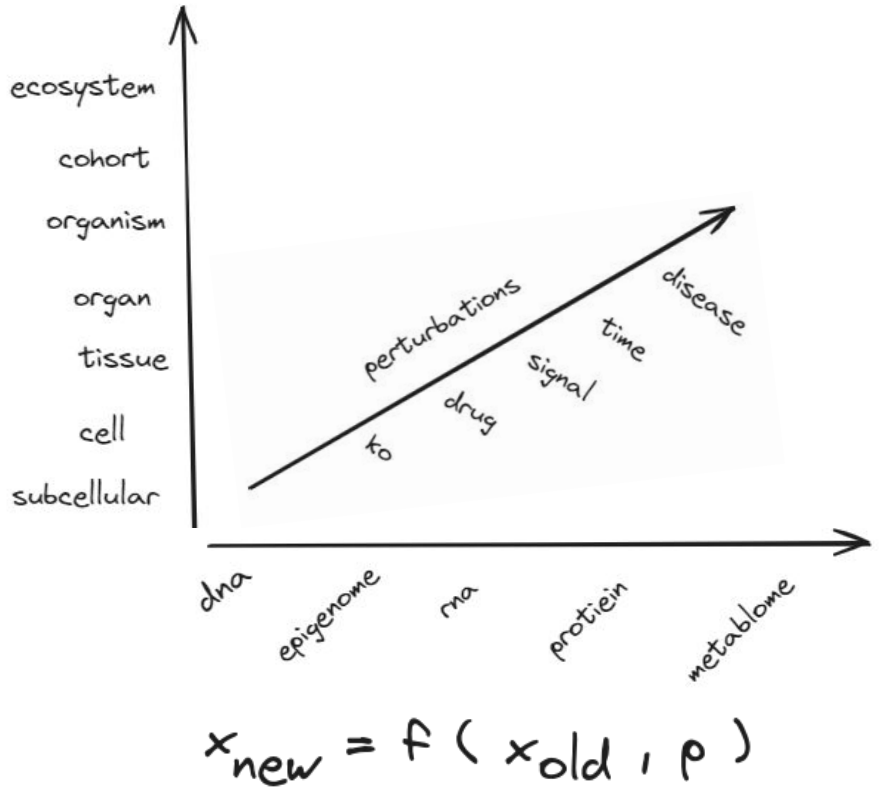
Problem Statement

- The complexity of human biology stems from the intricate roles and interactions of approximately 37 trillion cells that make up our tissues, organs, and bodily systems.
- Recent breakthroughs in single-cell technology have granted us detailed insights into cellular functions at the genetic and protein levels.
- To harness single-cell techniques for drug development, it is necessary to establish clear connections between chemical changes and their effects on cells.
- Despite their value, these experiments are resource-intensive and expensive, prompting the need for machine learning models.
- Single-cell perturbational datasets capture the variations in gene expression when cells encounter various stimuli, including drugs, chemicals, or environmental shifts.

Problem Statement

Understanding the Cellular Underpinning of health and disease



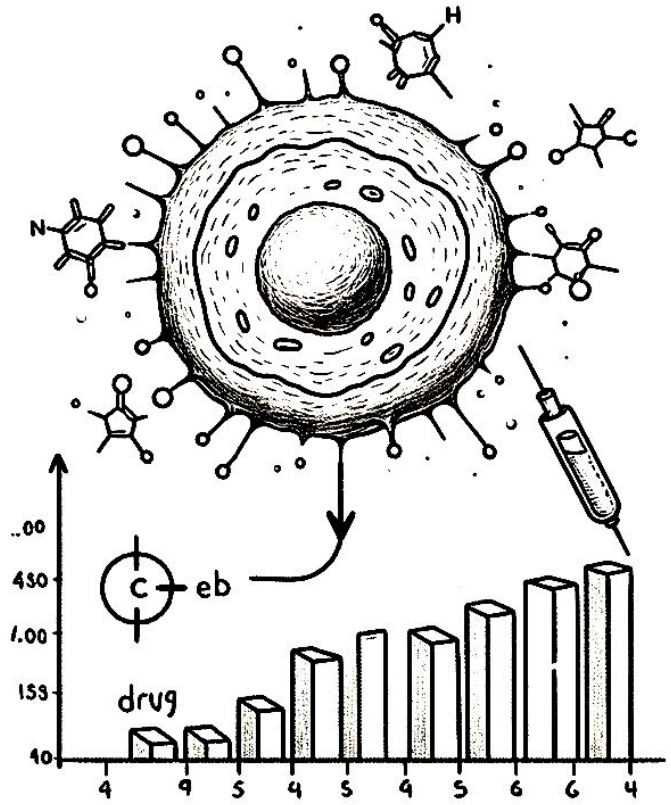


Can we understand a system
if we can predict its behavior?

Problem Statement

Problem statement:

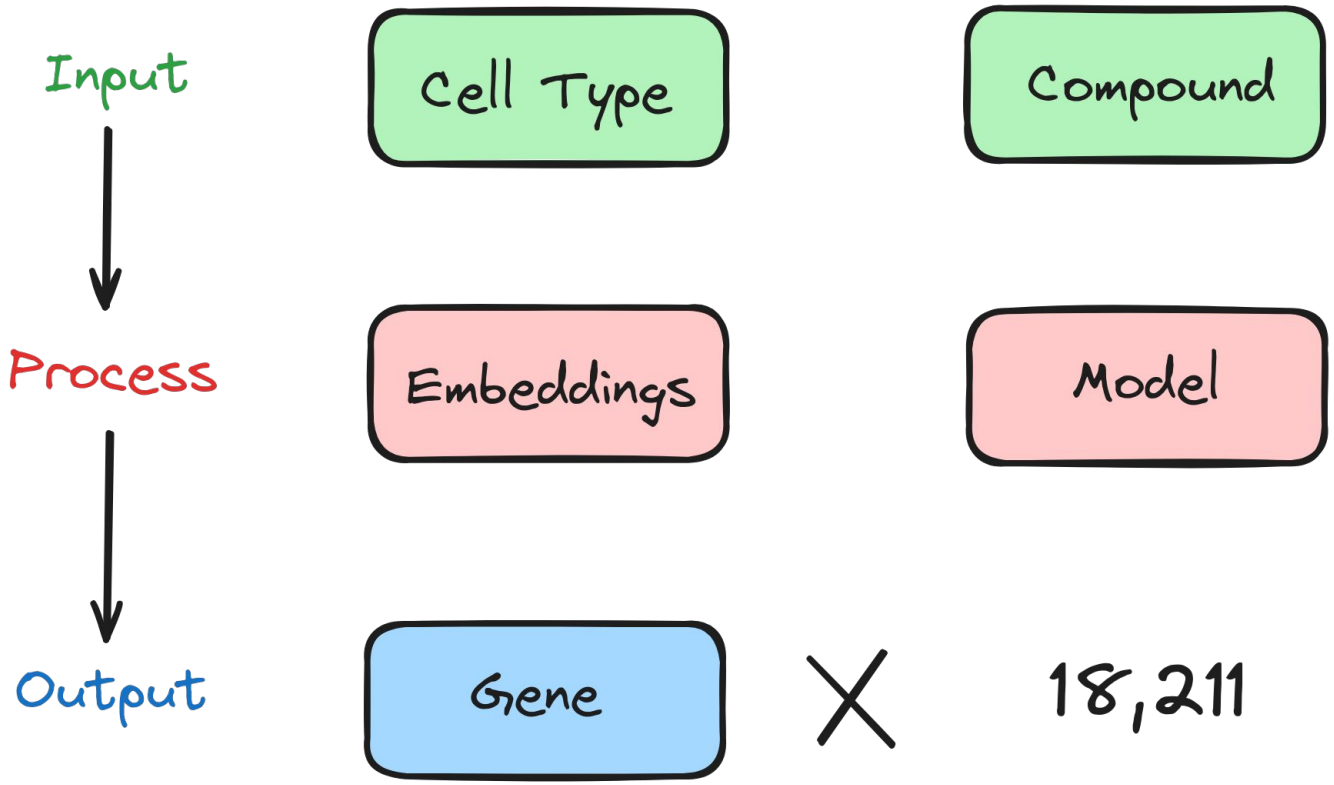
Single-cell studies unveil the intricate cellular diversity within biological samples, a crucial aspect often overlooked by traditional uniform approaches. This methodological shift is transforming disease research and personalized medicine, allowing for the development of targeted, more effective treatments by understanding the unique characteristics at the cellular level



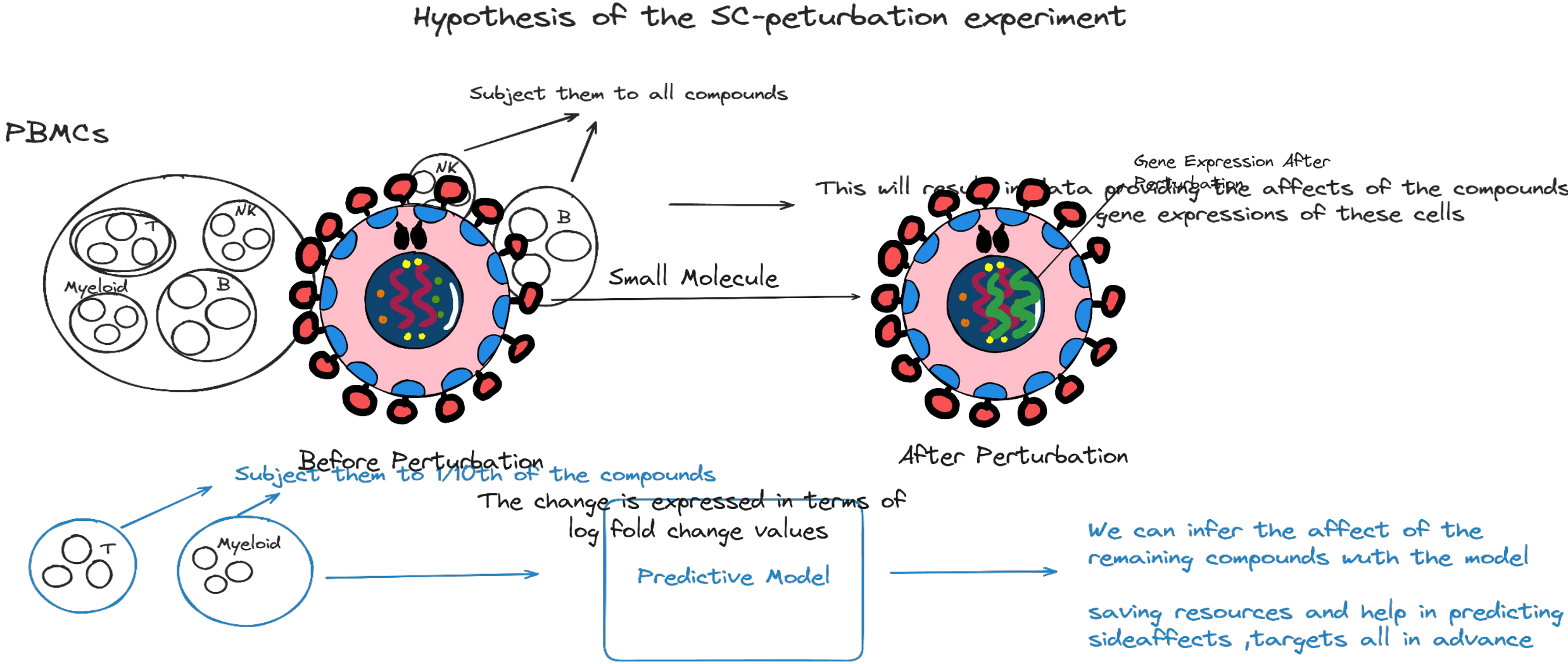
PES UNIVERSITY

Scope and Feasibility study

What data are we dealing with



Scope and Feasibility study



PES UNIVERSITY

Scope and Feasibility study

Datasets:

Dataset from Open Problems on single cell analysis:

- Gene Expression: Includes data for 144 compounds from the LINCS Connectivity Map dataset.
- Cell Types: Focuses on Human PBMCs comprising T-cells, B-cells, myeloid cells, and NK cells.
- Scope: Allows estimation of the impact of experimental perturbations on the expression level of each gene, covering 18,211 genes.

cell_type	sm_name	sm_lincs_id	SMILES	control	A1BG	A1BG-AS1	A2M	A2M-AS1	A2MP1
NK cells	Clotrimazole	LSM-5341	Clc1ccccc1C(c1ccccc1)(c1ccccc1)n1ccnc1	False	0.10472	-0.077524	-1.625596	-0.144545	0.143555

id	cell_type	sm_name
0	B cells	5-(9-Isopropyl-8-methyl-2-morpholino-9H-purin-...

PES UNIVERSITY

Scope and Feasibility study

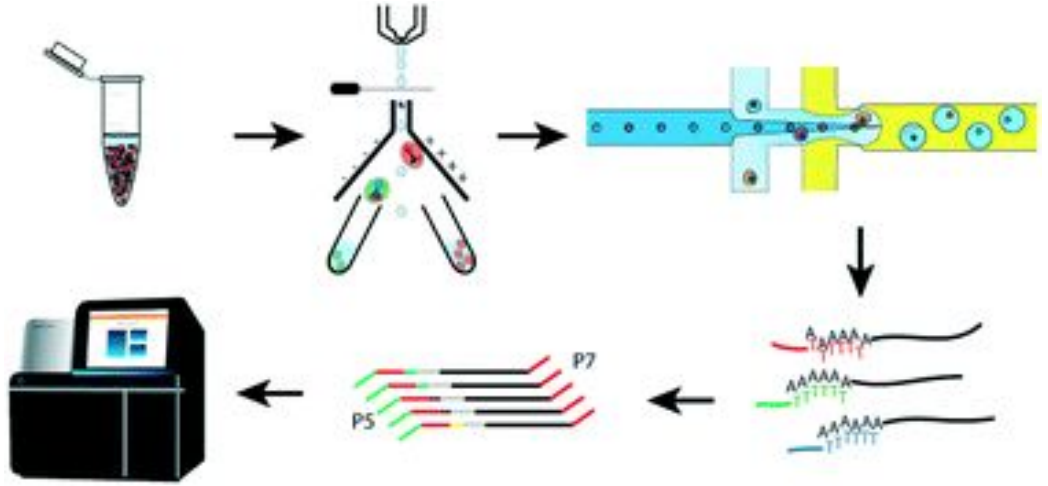
Challenges :

- Data Diversity: The vast number of genes present creates complexity in analysing and drawing meaningful insights due to the high-dimensional nature of the data.
- Heterogeneity of Single-Cell Responses: Single-cell responses to perturbations are highly variable due to factors like pre-existing variability in mRNA and protein abundance, cellular states, and the microenvironment. This heterogeneity makes it challenging to accurately predict and understand the mechanisms underlying these responses.

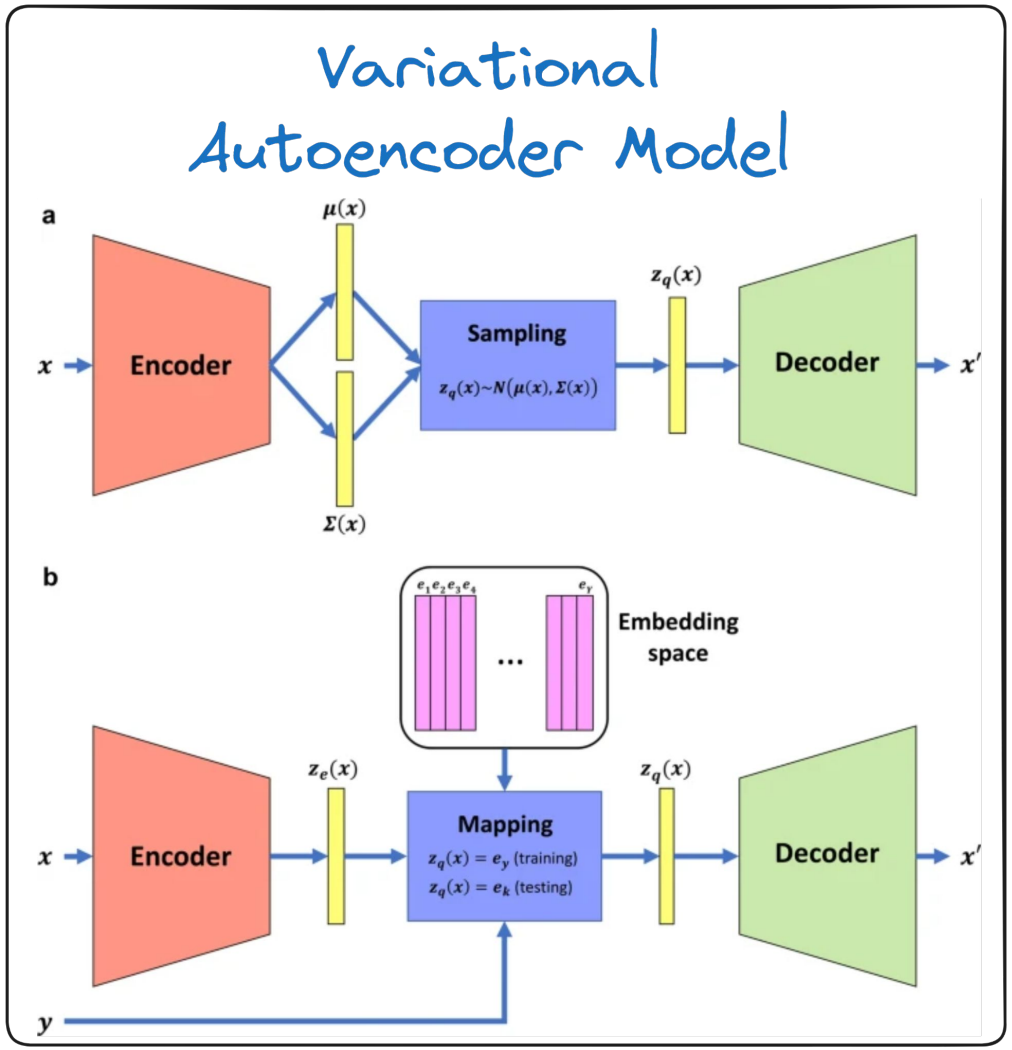
Scope and Feasibility study

Challenges:

- Interpretability: The complexity of deep learning models poses a challenge in understanding and explaining the reasoning behind their predictions and identified patterns.
- Technical Limitations: Single-cell sequencing technologies can have limitations such as high costs, limited throughput, and technical biases.
- Cell Viability: The process of obtaining single-cell measurements often involves destroying the cells, which limits the ability to conduct dynamic or longitudinal studies.



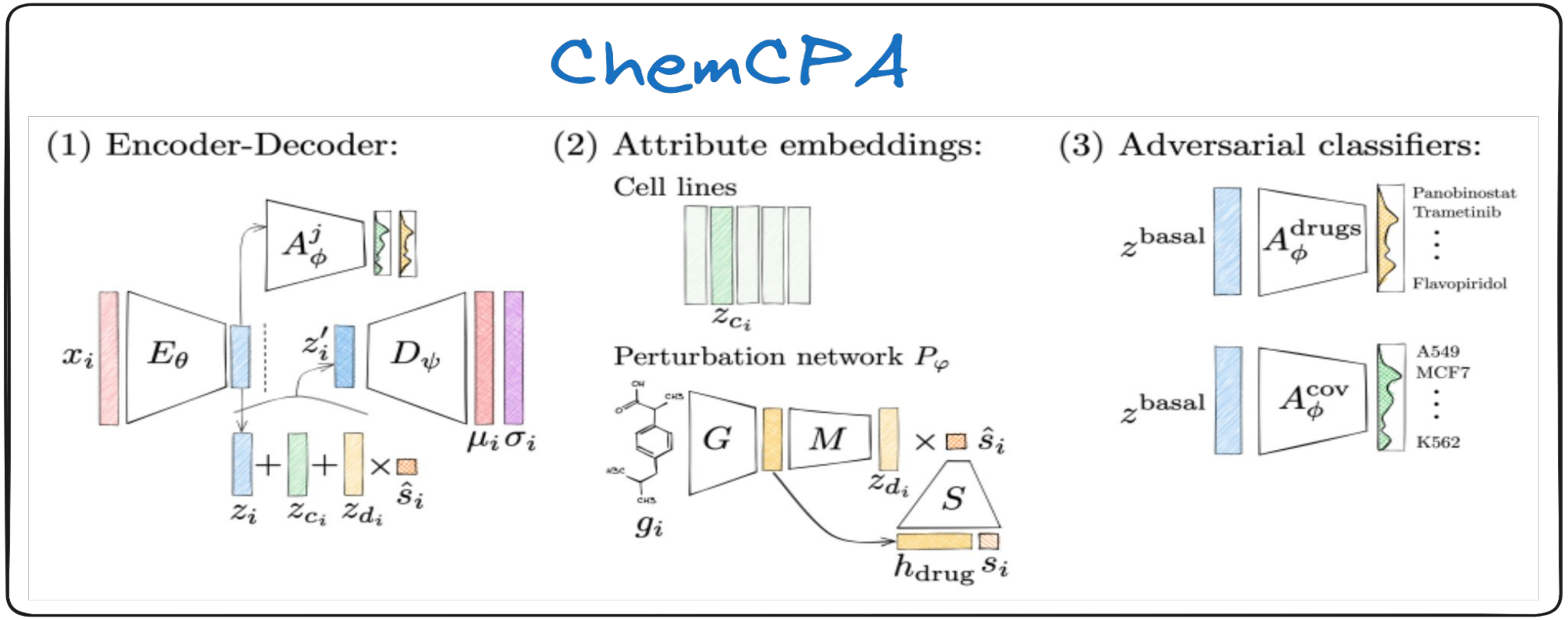
Scope and Feasibility study



Approach I: Variational Autoencoder Model

VAEs work directly with raw count data, eliminating the need for preprocessing steps like normalization and transformation.
They use deep generative models to robustly estimate gene expression levels and latent cell representations, capturing the variability across different cell populations.

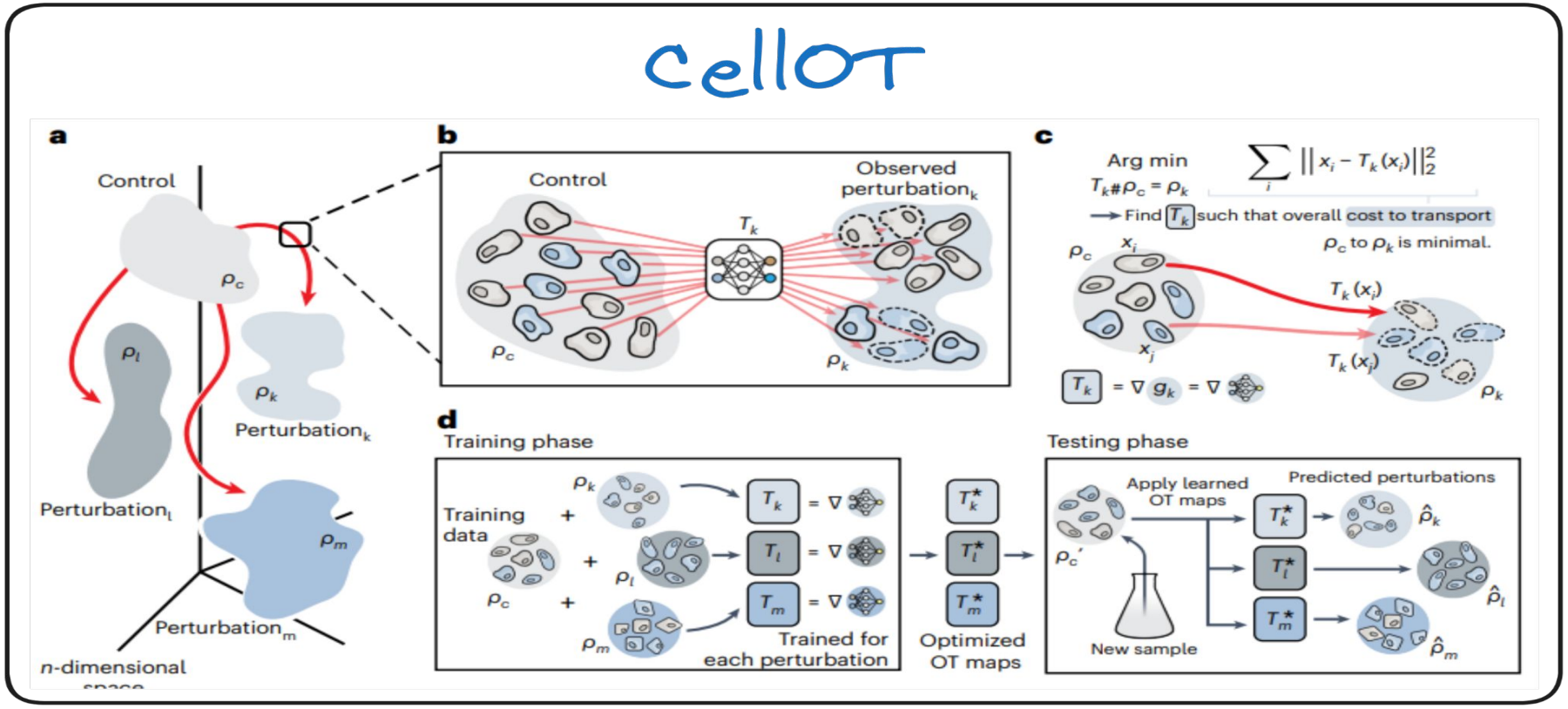
Scope and Feasibility study



Approach II: ChemCPA

The proposed model, chemCPA (Compositional Perturbation Autoencoder), employs an encoder-decoder architecture combined with adversarial training to handle dataset comprising gene expression data and attributes related to single-cell RNA sequencing (scRNA-seq) perturbation. It utilizes separate latent spaces for drug and cell-line attributes, employing novel embedding networks for molecular representations and dosage information.

Scope and Feasibility study



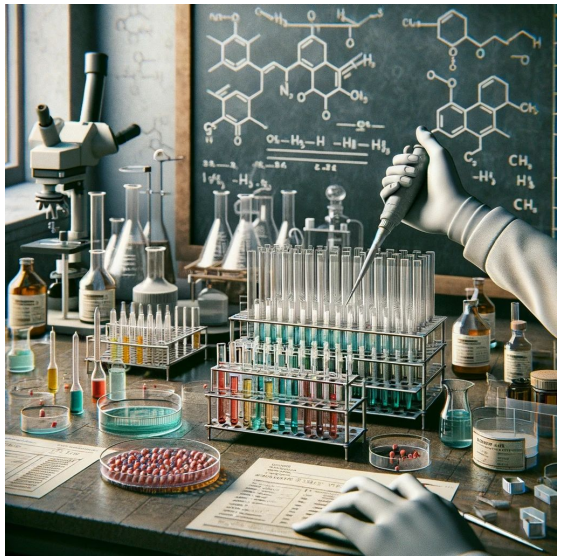
Approach III: CellOT

CellOT employs an innovative approach using optimal transport (OT) to analyze single-cell perturbation data. This method focuses on aligning distributions of perturbed and non-perturbed cells to identify the most probable states of cells post-perturbation.

Applications/Use cases

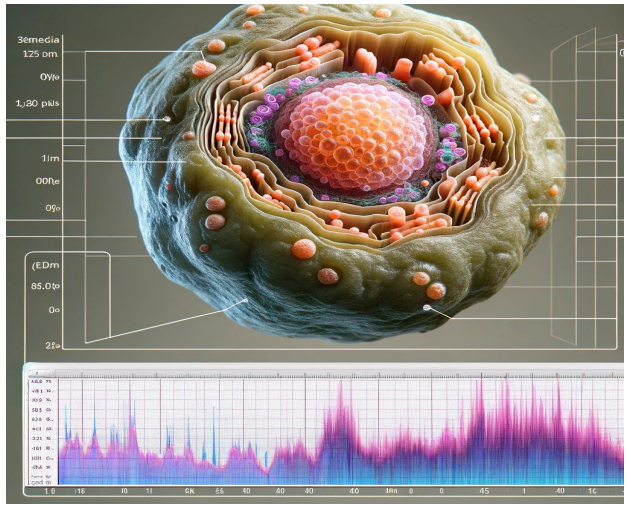
1. Drug Screening

Traditional way



- High-throughput screening
- Time-consuming and costly

Our way :

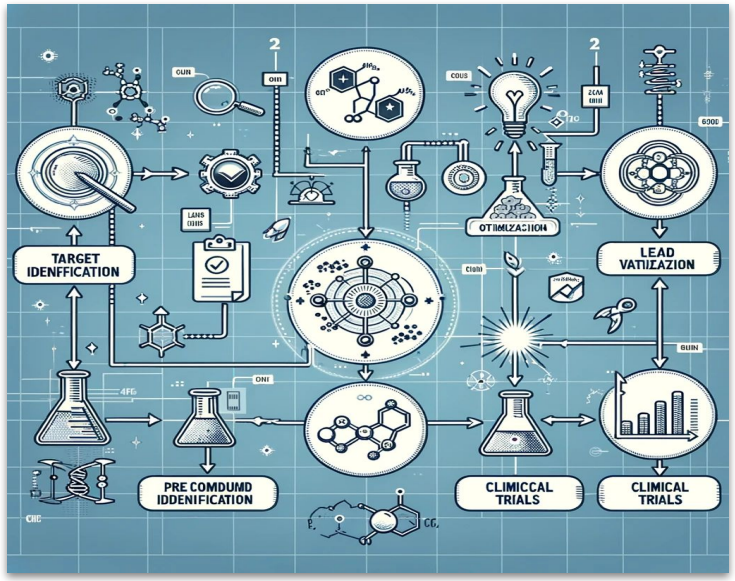


- Precision at single cell level
- Rapid candidate identification

Applications/Use cases

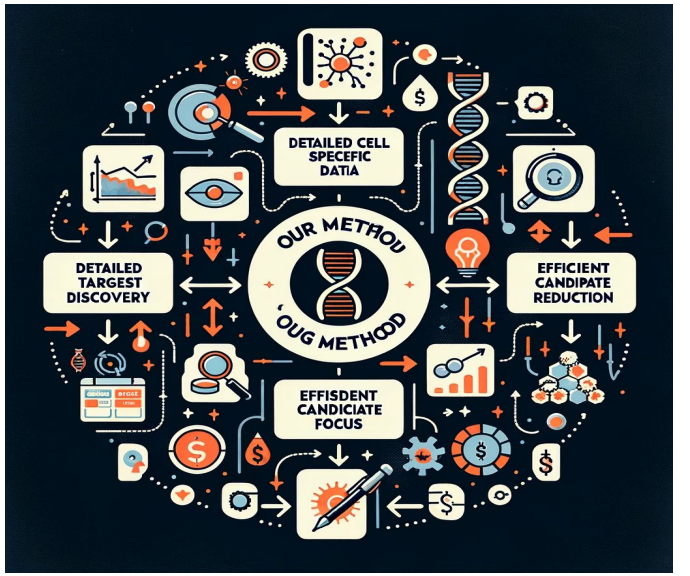
2. Drug Discovery

Traditional way



- Involves target identification, validation, lead compound identification, optimization, preclinical and clinical trials..
- High failure rate.

Our way :

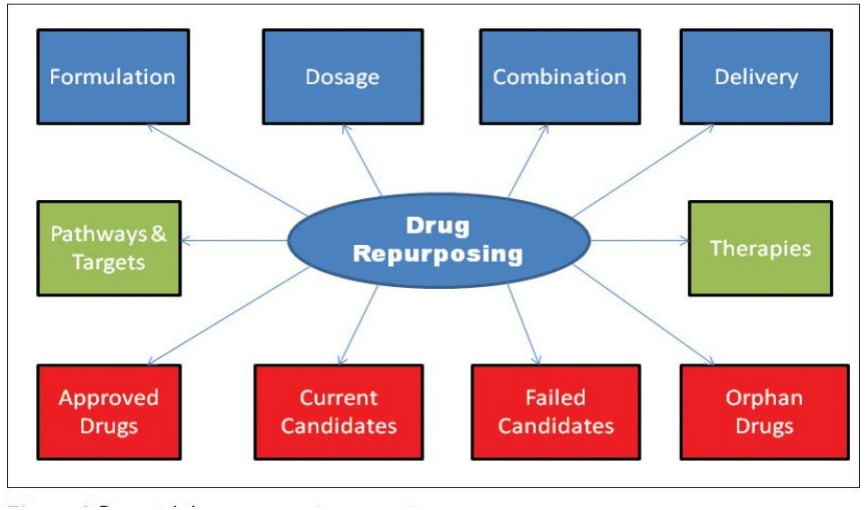


- Detailed, cell-specific data.
- Novel target discovery.
- Efficient candidate focus.

Applications/Use cases

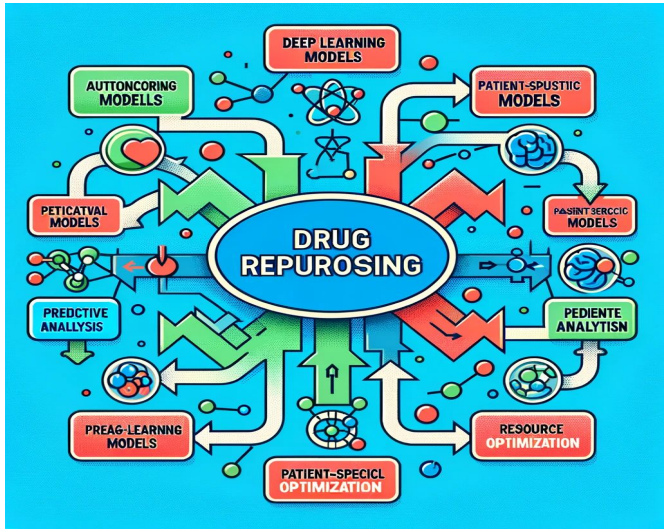
3. Drug Repurposing

Traditional way



- Often based on serendipity.
- Trial and error approach.

Our way :



- Specific cell type analysis.
- Gene expression data utilization.

PES UNIVERSITY

Expected Deliverables

- Capstone-I deliverables
  - High level design(HLD)
  - Literature Survey
  - Project Requirements
  - Baseline model
  - Phase 1

- Capstone-II deliverables
  - Low level design (LLD)
  - Implementation and Testing
  - Final project and report

Timeline

ID	Name
1	Problem Statement
2	Project Requirements Specification
3	Literature Survey
4	High-Level Design Document (HLD)
5	Preliminary Data Preprocessing
6	Implementation of Baseline Models
8	Further Literature Review
9	Low-Level Design Document (LLD)
10	Further Implementation and Testing
11	Final Project Report Generation

PES UNIVERSITY

References

[1] Lotfollahi, M., Wolf, F.A. and Theis, F.J., 2018. Generative modeling and latent space arithmetics predict single-cell perturbation response across cell types, studies and species. *bioRxiv*, p.478503.

[2] Bunne, C., Stark, S.G., Gut, G., Del Castillo, J.S., Levesque, M., Lehmann, K.V., Pelkmans, L., Krause, A. and Rätsch, G., 2023. Learning single-cell perturbation responses using neural optimal transport. *Nature Methods*, *20*(11), pp.1759-1768.

[3] Hetzel, L., Boehm, S., Kilbertus, N., Günnemann, S. and Theis, F., 2022. Predicting cellular responses to novel drug perturbations at a single-cell resolution. *Advances in Neural Information Processing Systems*, *35*, pp.26711-26722.

[4] Peidli, S., Green, T.D., Shen, C., Gross, T., Min, J., Garda, S., Yuan, B., Schumacher, L.J., Taylor-King, J.P., Marks, D.S. and Luna, A., 2024. scPerturb: harmonized single-cell perturbation data. *Nature Methods*, pp.1-10.

[5] Ji, Y., Lotfollahi, M., Wolf, F.A. and Theis, F.J., 2021. Machine learning for perturbational single-cell omics. *Cell Systems*, *12*(6), pp.522-537.

PES UNIVERSITY

Thank
You