National Institute for Theory and Mathematics in Biology
Funded by the U.S. National Science Foundation and the Simons Foundation

A mathematical framework for classification in cell state dynamics

Carlos Floyd, Ben Kuznets-Speck, Suriyanarayanan Vaikuntanathan, Yogesh Goyal

The classification of biological observations into meaningful categories is a fundamental challenge across disciplines. This proposal builds on advancements in single-cell technologies and increasing appreciation of non-genetic plasticity to tackle the dynamic classification of cell states and fates. Historically, tools such as microscopy shaped our understanding of cellular morphology. Today, high-dimensional data from single-cell profiling reveals vast heterogeneity, necessitating new frameworks for defining and identifying cell states.

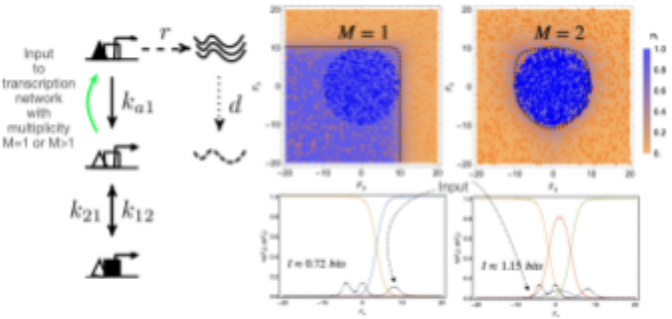

Fig 1: Recent work from the Vaikuntanathan group shows how multiplicity of input (M) systematically determines the classification and signal processing abilities of biochemical networks. Our proposed work seeks to expand this idea to complex high dimensional gene transcription networks and cell state dynamics. We seek to explore how this new mathematics can help provide insight into how transcriptionally identical cells respond to drug perturbations.

Cell states, described as a cell's molecular characteristics at a given moment, often transition under perturbations, leading to distinct fates. However, the thresholds distinguishing states and fates remain poorly defined. This research proposes a novel integration of mathematical modeling and experimental biology to address these challenges. By leveraging concepts from dynamical systems and information theory, the study aims to develop a mathematically consistent framework for classifying cellular perturbations. Inspired by recent insights into biochemical networks as classifiers (FIG 1, arXiv:2409.05827), the project explores how coupled gene transcription dynamics can predict state-fate transitions and how their classification capacity can be tuned. Experimental systems, including drug resistance in melanoma and pancreatic cancer, will provide rich datasets to test and refine these models.

The integration of biology with mathematics is central to this research. Biological phenomena such as plasticity and heterogeneity are reimagined through the lens of statistical mechanics, new ways to define and explore cellular states. This interdisciplinary approach promises to uncover universal principles governing cell fate decisions, paving the way for deeper insights into cancer resistance and beyond.

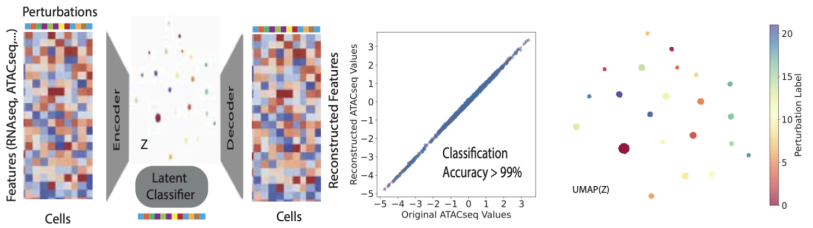

FIG 2. ExPert: Extracting perturbations from gene expression and chromatin accessibility. Left: a multimodal autoencoder projects high dimensional RNAseq and ATACseq data onto a latent space designed so that points are separated by gene perturbation. Center: ATACseq profiles are accurately reconstructed from the autoencoder, which classifies perturbations with near perfect accuracy. Right: UMAP of the autoencoder latent space.

In general, we are working to classify gene perturbations (knock-ups-and-downs, FIG 2. above) in order to infer which perturbations cause a given change in gene expression. The goal of the project is to infer causal perturbations leading to and mitigating drug resistance fates in cancer.

This research was supported by the NSF (DMS-2235451) and Simons Foundation (MP-TMPS-00005320).