Sharif University of Technology

Physics Department

Special Topics in Statistical Physics

Dynamical Analyze of GRN for AML

Yasaman Safayi Konjin

Sepehr Salmani Yeganeh

Sara Akbari Khorram

spring of 2025

Contents

Abstract

This report presents the outcomes of a collaborative project focused on analyzing the dynamics of the gene regulatory network (GRN) associated with acute myeloid leukemia (AML). The primary objective was to develop a computational model to reconstruct AMLs GRN, enabling the study of the disease through the activities of its most influential transcription factors. Initially, comprehensive data on AML were collected from relevant literature to inform the modeling process. The innovative aspect of this project lies in analyzing network dynamics using a physics-based approachspecifically, leveraging the Jacobian matrix, a method not previously applied in this context. A Python-based model was developed to generate a dataset capturing potential irregular dynamics that may contribute to cancer development. This dataset was analyzed using the k-means clustering method to identify patterns in transcription factor activity. The results demonstrated a promising alignment with documented transcription factor behaviors in AML cases reported in the literature. Furthermore, a Random Forest machine learning algorithm was employed to detect irregularities in transcription factor activities and diagnose underlying dynamics. These findings offer valuable insights into AMLs genetic dynamics and pave the way for future research in targeted diagnostics and therapeutic strategies.

1 Introduction

1.1 Problem Statement

Acute Myeloid Leukemia (AML) is a complex hematologic malignancy characterized by dysregulated gene expression, driven by aberrant dynamics in its *gene regulatory network* (GRN). Despite advances in understanding AMLs molecular basis, current computational models often fail to capture the intricate dynamics of transcription factor interactions that drive disease progression. Traditional approaches to GRN analysissuch as differential expression studieslack the capacity to model the nonlinear and dynamic behavior of gene interactions, limiting their effectiveness in identifying critical regulatory patterns.

Moreover, there remains a gap in applying physics-based methodologies, such as Jacobian matrix analysis, to the study of AMLs GRN - an approach that could uncover novel insights into its genetic dynamics. This project addresses the need for an innovative computational framework to reconstruct and analyze AMLs GRN, with a focus on identifying irregular transcription factor activities. By leveraging a physics-based approach, this study aims to deepen our understanding of AMLs regulatory mechanisms and support the development of targeted diagnostic and therapeutic strategies.

1.2 Objectives

The objectives of this project are as follows:

1. To develop a computational model for reconstructing the gene regulatory network (GRN) of acute myeloid leukemia (AML) using a physics-based approach that leverages the Jacobian matrix.

2. To generate a dataset that captures irregular dynamics in transcription factor activities potentially contributing to AML progression.

3. To analyze the generated dataset using the k-means clustering method in order to identify patterns in transcription factor behavior.

4. To implement a Random Forest machine learning algorithm for detecting and diagnosing irregularities in transcription factor activities within the AML GRN.

5. To validate the models findings against existing literature, assessing its accuracy and potential to advance targeted diagnostic and therapeutic strategies.

2 Methodology

This section outlines the computational methods used to reconstruct and analyze the gene regulatory network (GRN) of acute myeloid leukemia (AML), focusing on the dynamics of six transcription factors: CEBPA, SPI1, MYB, RUNX1, HOXA9, and MEIS1.
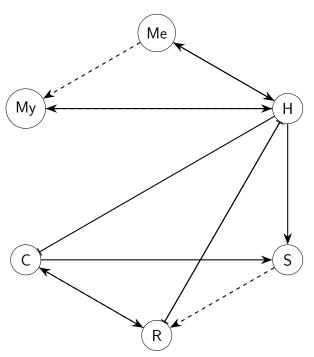

Figure 1: Gene regulatory network showing activation (solid/dashed arrows) and suppression (flat-head lines) among Me, My, H, R, C, and S.

2.1 Data Collection and Parameter Setup

A simulated dataset was generated to represent the dynamics of AMLs GRN, focusing on six transcription factors. Interaction parameters were defined as a symbolic matrix A with 19 parameters (e.g., cc, rc, hc), representing regulatory interactions. Mean and standard deviation values for scaling fixed points were assigned based on hypothetical RNA-seq data.

2.2 Model Development

A Python-based computational model was developed to reconstruct AMLs GRN using the Jacobian matrix. The matrix $\mathbf{A}$, representing GRN interactions, was constructed using `SymPy` for symbolic computations and converted to numerical form with `NumPy` and `SciPy`. The condition for non-zero fixed points was derived by solving $\det(\mathbf{A}) = 0$ for the parameter cc.

To ensure a robust representation of the GRN and mitigate matrix sparsity, hypothetical regulatory links were incorporated into the symbolic matrix. A total of 42,208 valid parameter sets were generated by randomly sampling 18 parameters within predefined ranges (e.g., `activation constants` $\in [0.01, 2]$, `suppression constants` $\in [-2, -0.01]$, `self-activation and suppression rate` $\in [-2, 2]$, `hypothetical links` $\in [0, 1]$), while ensuring $\det(\mathbf{A}) \approx 0$. Fixed points were computed using the null space of $\mathbf{A}$ and scaled to RNA-seq ranges using predefined means and standard deviations. Data generation continued iteratively until exactly `number_of_stable_fixed_points` $= 512$ stable fixed points were obtained, ensuring sufficient representation of stable system dynamics for downstream analyses.

2.3 Dynamic Analysis

The dynamics of the GRN were analyzed by computing the eigenvalues of the numerical Jacobian matrix $\mathbf{A}$. The real parts of the eigenvalues were used to classify the systems behavior into Stable, Unstable, Saddle, or Non-zero Fixed Point dynamics, with a threshold of ± 0.0001.

2.4 Clustering Analysis

The fixed points dataset, comprising expression levels of the six transcription factors, was normalized using `StandardScaler` from `scikit-learn`. The k-means clustering algorithm was applied to identify patterns in transcription factor dynamics, with the optimal number of clusters determined using the silhouette score over a range of $k = 2$ to 9. Clustering results were visualized in 2D and 3D using UMAP (`n_neighbors=30`, `min_dist=0.5`) and PCA, implemented with `umap-learn` and `scikit-learn` libraries,

respectively. Visualizations were generated using `Matplotlib` and `Plotly` to analyze cluster distributions.

2.5 Diagnostic Modeling with Random Forest

A Random Forest classifier, implemented via `scikit-learn`, was trained to classify GRN dynamics (Stable, Unstable, Saddle, Non-zero Fixed Points) based on the dataset of fixed points and interaction parameters. The dataset was split into training (80%) and testing (20%) sets using a random seed of 42 for reproducibility. Features were normalized using `StandardScaler`, and the model was trained on the scaled training data. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics, as reported in a classification report.

2.6 Validation

The clustering and classification results were validated by analyzing the distribution of dynamics within each cluster and computing statistical measures, such as the mean and standard deviation of fixed points and interaction parameters. The silhouette score assessed clustering quality, while the Random Forest models classification report ensured robust detection of dynamics. Findings were compared with documented AML transcription factor behaviors to evaluate their biological relevance.

3   Results

.

3.1   Key Findings

.

4 Discussion

.

5 Conclusion

.