An intelligent system for the analysis of Complex Gene Regulatory Networks (GRN): The Chronic Lymphocytic Leukemia GRN as study case

Pedro Pineda-Salvador[1], Katya Fuentes-Sánchez[2], Oscar Morales-Toledo[1], Ian Machuca-Islas[1], Emiliano Tizapán-Camacho[1], Misael Ávila-López[1], Dalia Domínguez-Jiménez[3]

[1] BSc Program on Biomedical Physics, Physics Department, Universidad Nacional Autónoma de México, Mexico City, Mexico
katya.fuentes23@ciencias.unam.mx
[2] Physics Department, Biomedical Physics, Universidad Nacional Autónoma de México, Mexico City, Mexico
[3] PhD Program on BioRobotics, BioRobotics Institute, Scuola Superiore Sant'Anna, Pisa, Italy

Abstract. Chronic Lymphocytic Leukemia (CLL) is a biological process in which macrophages derived from monocytes are expressed as M1, M2, and NLC phenotypes, promoting the proliferation of cancerous cells with the possibility of transitioning from one phenotype to another depending on the incoming stimulus, whether from other cells or from the macrophages themselves. Dynamic transitions between healthy and diseased states in such biological systems have been studied through Gene Regulatory Networks (GRNs). Our previous results, based on differential equation models and machine learning algorithms, have been applied to the analysis of the GRN underlying macrophage polarization in CLL, enabling the classification of genotypes and the identification of genes such as interleukins 4, 10, 13, signal transducers, and transcriptional activators 3 and 5, which are prone to induce transitions between genotypes and/or phenotypes, as well as the manipulation times and relative gene expression required to trigger these transitions. However, the results of these analyses are not intuitive for specialists, so we integrated each of the algorithms used into a single graphic user interface developed in MATLAB, which automatically performs dynamic exploration, state classification, and the identification of genes likely to transition, as well as their manipulation time in a simple manner.

Keywords: Gene Regulatory Networks, supervised machine learning algorithms, state cell transitions, graphic user interface

1 Introduction

Genetic regulatory networks (GRNs) are fundamental structures in systems biology as they allow for the modeling and prediction of gene interactions within complex dynamic systems. The study of these networks is essential for understanding cellular processes, identifying biomarkers, and developing advanced biomedical applications [2].

One of the most widely used models for studying GRNs is the Boolean model, which relies on discrete approaches to understand the system's qualitative behavior through the dynamics of two states, on and off [1, 4]. These states can evolve over time, reaching stable configurations known as attractors, which represent the system's invariant behavioral patterns [4].

In previous works, these qualitative studies have been extended to quantitative approaches using ordinary differential equations to model the network's dynamics and assess the tendency of individual genes to qualitatively change an attractor in response to modifications in their degradation rate [3]. A key challenge in this analysis is managing networks of varying dimensionality, as low-dimensional networks allow for detailed characterization of attractors, while high-dimensional networks require greater computational processing. In this context, artificial intelligence (AI) and machine learning algorithms have been widely used to infer gene interactions and model cellular dynamics from large volumes of biological data, aiming to make accurate predictions about the behavior of these networks under different conditions.

2 Dynamic analysis of gene regulatory network: TAMCLL as a study case

In our previous study, we analyzed the GRN underlying macrophage polarization in the context of CLL (TAMCLL), consisting of 40 genes (30 intracellular and 10 extracellular) with three outputs corresponding to the M1, M2, and NLC phenotypes (see Fig.1a.). Based on the Boolean model proposed by Marku Malvina, 1384 fixed-point stable states were retrieved through synchronous transition simulations from each initial condition. Given the high dimensionality of the network, supervised learning algorithms were employed to classify these attractors into the M1, M2, NLC, and M0 phenotypes (macrophages exhibiting both M1 and M2 characteristics), revealing that the KNN model trained with the genotype and the degree of each attractor achieved the highest accuracy (~96%) (see Fig. 1b).

To explore the transition dynamics between phenotypes, a bifurcation analysis was conducted by approximating the Boolean equation system to ordinary differential equations using the Glass dichotomous model [4]. A control input u and decay rates k were introduced for each gene, promoting transitions between stable states by modifying their expression levels (see Fig. 1c.). In this way, key genes and trajectories for transitions of interest (M1, M2, and M0 to NLC and vice versa) were identified. The results indicated that the IL10 gene, with a tendency greater than 0.8 (see Fig. 1d.), is the main driver of these transitions, being associated with M2 phenotype signaling and treatment resistance in CLL [5]. Similarly, the MCSF gene (with a tendency between 0.7 and 0.8) is associated with NLCs and TAMs (tumor-associated macrophages). In general, the genes that facilitate the most transitions correspond to extracellular signals and the group of transcription factors and interleukins.

To achieve this comprehensive analysis, various computational tools and different programming languages were used, which can be confusing for users switching between programs. Therefore, an interface was developed to integrate each of the

algorithms for classification and dynamic analysis of the TAMCLL-GRN. This interface includes functions that are accessible to beginners while still offering all (or most) of the options that an experienced specialist would require.

3 Methods: algorithms implementation

To integrate classification algorithms into the interface, attractors and the transition table were retrieved from the R BoolNet package. The first algorithm loads .txt files with Boolean functions, extracts attractors, and converts their genetic profiles from binary to decimal. It also processes the transition table to calculate interaction degrees and steps to the attractor, storing this in a feature table for classification. The trained KNN model then classifies the attractors into phenotypes, and then an algorithm counts their distribution, resulting in 79 attractors of M1, 50 of NLC, 337 of M2, and 918 of M0.

Dynamic analysis was performed using Glass's dichotomic model, calculating gene propensity to induce transitions between phenotypes via bifurcation analysis. Two algorithms were used for this: the first rewrites Boolean functions for the Glass model, and the second performs bifurcation analysis by varying the decay rate (parameter k). A fourth algorithm retrieves classified attractors and stores them as initial conditions for solving a system of ODEs, allowing gene analysis and perturbation selection. Results, including phenotypic transitions, are stored in a dataframe. Finally, the last algorithm summarizes and visualizes phenotypic transition distributions, calculating average manipulation time per gene and generating ASTD matrices [5] to estimate each gene's propensity to induce transitions between M1, M2, M0, and NLC.

4 Data visualization scheme and integration

The graphical interface consists of a window with three main processes: attractor classification, bifurcation analysis, and the analysis of the transition distribution between phenotypes (see Fig.2a.). Upon launching the application, the user is prompted to load the necessary information for network implementation, specifically the .txt file containing the Boolean functions of the network and the files generated in R, which include the 1384 network attractors and the transition table with "attractor assignment" and "transitions to attractor." After loading these files, a characteristics table is generated, which includes the size of the attraction basin, the longest number of steps to the attractor, the number of interactions, and its decimal equivalent.

Next, the trained machine learning model is loaded. Once this stage is completed, the application allows for the classification of attractors and the counting of their distribution by phenotype. Subsequently, for dynamic analysis, the continuous approximation of the Boolean functions describing the network is transcribed. Then, the user selects which genes will be turned on or off during the analysis. Based on this selection, a gene vector is generated, and a bifurcation analysis is performed. This analysis consists of a transition table between phenotypes, the transition time between each manipulation, and the gene that promotes each transition, with the option to download the results in .csv format. Next, a button is available to perform the transition distribution

analysis (ASTD) using matrices. This analysis displays information on how many transitions each gene produces, the average manipulation times, and the tendency of genes to promote transitions between different phenotypes. Additionally, it generates the corresponding graphs for the ASTD matrices and the analyzed gene tendencies, with the option to save them as image files. It is worth noting that during these two processes, a pop-up window appears indicating the progress of the analyses.

Once the analysis is completed, a button is enabled with the option to save the information generated in the attractor classification and dynamic analysis in .txt format. Finally, an option is available for the user to exit the interface, closing the program (see Fig, 2b.).

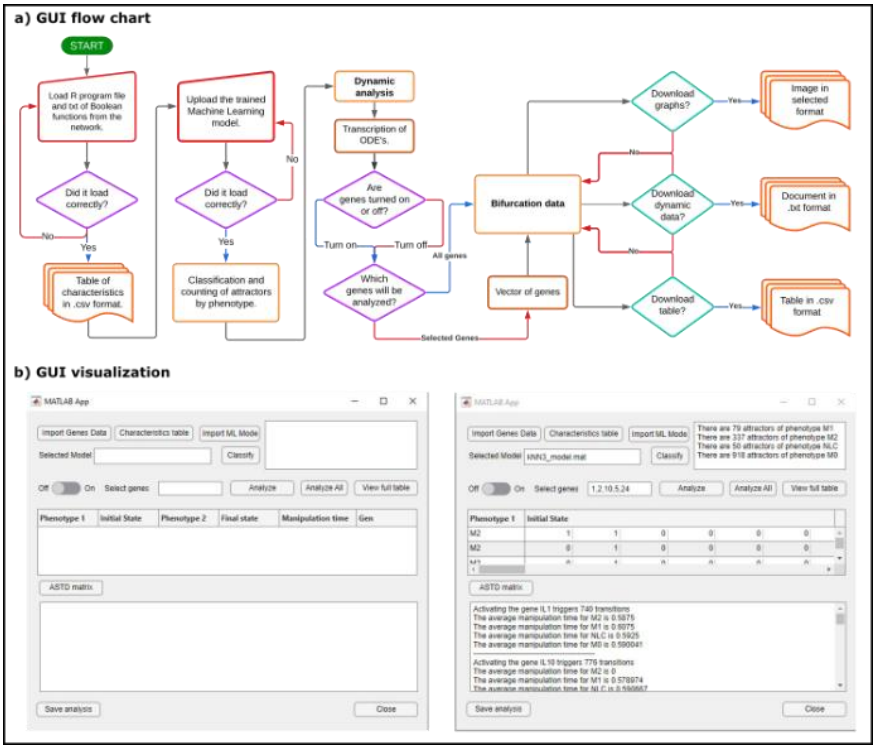

Fig.2. GUI model building interface: a) flow chart describing the interface algorithms, b) GUI visualization analyzing the selected genes.

5 Discussion

This interface integrates all the algorithms described in section 3, making possible a more user-friendly interaction with the user. It was also possible to create a section within the interface, which serves to select genes that the user wants to analyze, reducing the processing time and visualization of results, to finally perform a statistical analysis that can deliver the tendency of each gene to generate transitions between

phenotypes. On the other hand, the execution time to extract the features and analyze all the genes with all the initial conditions is extensive, so it is proposed to reduce this time by optimizing the algorithms to obtain the features and the solution of the ODE system when performing the complete analysis. Also, being an interface that performs a complete analysis of the transitions between phenotypes, we propose to generalize the implemented algorithms to extend this type of analysis to not only the TAMCLL network but also to other GRNs of different dimensions that describe biological processes of interest. When extending the analysis to other GRNs it will be necessary to implement within the interface a training stage of classification models according to the network to be studied or supervised learning models to eliminate the feature extraction time.

6 Conclusion

This GUI developed in MATLAB provides a novel, easy-to-use solution capable of automatically performing dynamic exploration, state classification and identification of potential genes to be transitioned as well as their manipulation time in a simple way. Its workflow guides the user through each step of the TANNCLL-GRN analysis, making it accessible to any type of specialist user.

References

1. Avila-Ponce de León U, Vázquez-Jiménez A, Matadamas-Guzman M, Pelayo R, Resendis-Antonio O. Transcriptional and Microenvironmental Landscape of Macrophage Transition in Cancer: A Boolean Analysis. Front Immunol. 2021 Jun 10;12:642842. doi: 10.3389/fimmu.2021.642842. PMID: 34177892; PMCID: PMC8222808.
2. Linde J, Schulze S, Henkel SG, Guthke R. Data- and knowledge-based modeling of gene regulatory networks: an update. EXCLI J. 2015 Mar 2;14:346-78. doi: 10.17179/excli2015-168. PMID: 27047314; PMCID: PMC4817425.
3. Palma A, Jarrah AS, Tieri P, Cesareni G, Castiglione F. Gene Regulatory Network Modeling of Macrophage Differentiation Corroborates the Continuum Hypothesis of Polarization States. Front Physiol. 2018 Nov 27;9:1659. doi: 10.3389/fphys.2018.01659. PMID: 30546316; PMCID: PMC6278720.
4. Veliz-Cuba, A., & Murrugarra, D. (2015). Steady state analysis of Boolean models: A dimension reduction approach. En R. S. Robeva (Ed.), *Algebraic and discrete mathematical methods for modern biology* (pp. 121-139). Academic Press. https://doi.org/10.1016/B978-0-12-801213-0.00006-X
5. Fuentes-Sánchez, K. et al. (2025). Understanding the Role that Genes Play on the Underlying Genetic Regulatory Network (GRN) of Macrophage Polarization in the Context of Chronic Lymphocytic Leukemia. In: Flores Cuautle, J.d.J.A., et al. XLVII Mexican Conference on Biomedical Engineering. CNIB 2024. IFMBE Proceedings, vol 116. Springer, Cham. https://doi.org/10.1007/978-3-031-82123-3_15