Gene Regulatory Networks behind simple patterning phenotypes: circuit analysis

Sebastian Espejo Agudelo

August 24, 2022

Contents

1 Introduction

During development, animal embryos rely on differential expression of some genes to create early patterns that define cell fate. This process of early pattering ends up giving rise to differentiated tissues and body structure in the adult. One of the most studied mechanisms of differentiation is the gene regulatory network (GRN) model (Fig. 1A) that interpret the concentration of molecules in the cellular environment, producing a clear pattern of gene expression (Fig. 1B) [1].

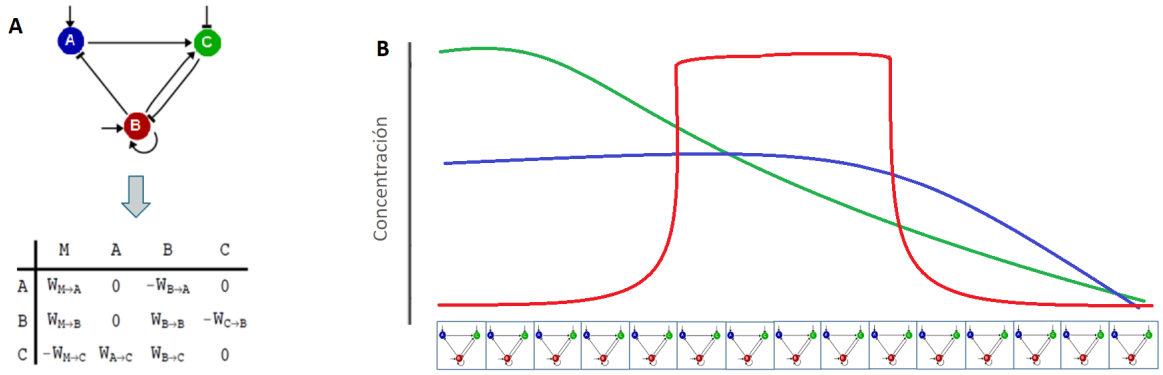

Figure 1. A) Exemplar of a gene regulatory network and its representation as a matrix of interactions, where negative interactions represent inhibitions, positive interactions represent activations and zeros, no interactions. Arrows coming from outside represent the interpretation of the external molecule. Every W represent a specific value between 0 and 10. B) Representation of the model, every W represent a specific value between 0 and 10. The array of GRN represent a tissue and one of the genes (B:red) only expresses in the middle of it as a result of the gene interactions.

All the 2061 GRNs were obtained by Arboleda et al [2]. with the following conditions: they are composed in 3 genes, they are embedded in an array of identical circuits representing an embryonic tissue (Fig 1B.), at least one of the genes is able to receive an external signal and, at least one of the genes is going to express only in the middle of the tissue. They used an evolutionary algorithm that recreates events of mutation and replication to obtain a population of 2061 GRNs able to complete this phenotypical task. The evolutionary algorithm also selected parameters of diffusion and degradation of this gene products that are disregarded in the current study.

However, these parameters are necessary to create the dynamical system of gene expression that leads to the stationary levels of expression of the three genes represented in Fig 1B.

The purpose of this project is to analyze this large population of GRNS in terms of diversity and complexity. All the analyses were done in R, and GRN representation was done in Mathematica.

2 Data set up

Data was obtained from the author Github repository. Every GRN is defined by 18 parameters, 12 parameters of interaction between elements of the circuit (see Fig 1A) and, 6 parameters of diffusion and degradation. This diffusion and degradation parameters were disregarded for the current study.

To perform some of the following analysis, I generated the set of all the possible circuit topologies, being topology the type of interaction (inhibition or activation) of the GRNS without considering specific values of these interactions. There is a total of 881 topologies.

3 Analysis

3.1 Topological considerations

Every topology is represented several times in the original population not only because they appear with different interaction values, but because every topology can have six equal configurations or permutations with exactly the same connectivity but interchanged positions of the nodes (see one example in Fig. 2)

Fig 2. Example of two GRNs with identical topology. The combinatorics of nodes and interactions allow having up to 6 identical topologies per circuit.

This important porperty was taken into account. Since this study base the analysis on the circuitry of the GRNs, it was necessary to consider a circuit and its permutations as the same topology. The amount of topologies without considering its permutations is 714, it means that 167 topologies are permutations of another one.

3.2 Abundance analysis

To understand how well represented are every type of topology in the original sample of GRNs, I created an abundance histogram as they are done for species abundance analysis.

The histogram shows (Fig 3.) that most of the circuits have a non-repeating or permutated topology, but few topologies have a lot of ocurrences, suggesting that some circuit topologies may be very robust at producing the phenotype regardless of the specific values of the interaction. Biologically, it may suggest that some mechanism of interaction rely more on the type of interaction than on its strength.

The Shannon diversity index of the original population of GRNs is 5.91. Although the interpretation of this index is usually relative, to have a sense of its magnitude, the if it is compared to natural ecological communities, the index for the population of GRNs is high. The Renyi profile was also suggested to compare different measurements of diversity (see scripts)

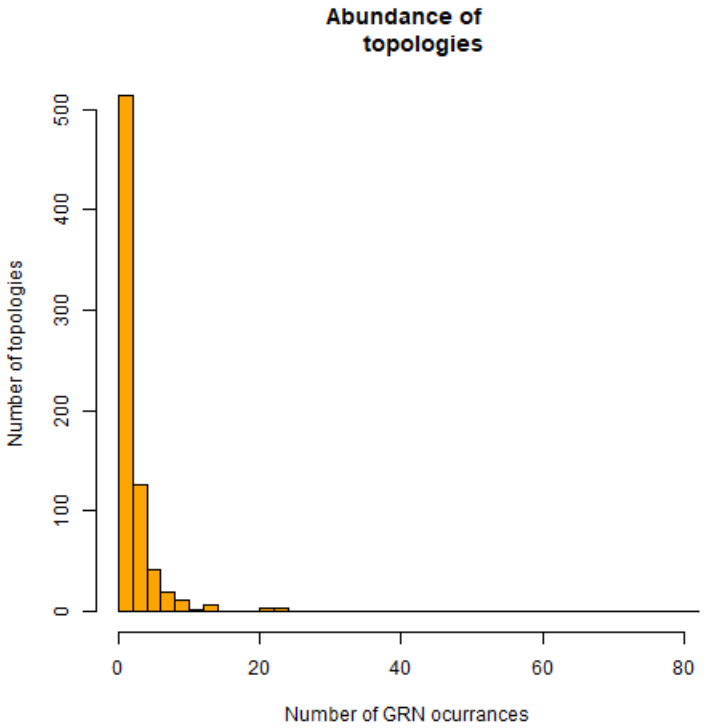

Figure 3. Histogram of abundance of GRNs in terms of the topology they have.

3.3 Complexity analysis

Another way to get a general overview of the total population of GRNs is counting the amount of interactions they have. This number can be understood as a rudimentary proxy of circuit complexity. In Fig 4 a histogram of complexity is shown. If we assume that the evolutionary algorithm is unbiased at exploring the possible set of GRNs, what Fig. 4 tells us is more probable to have GRNs of 8 to 10 interactions able to perform the phenotype. Few GRNs with 5 or 12 interactions are able to attain the desired phenotype.

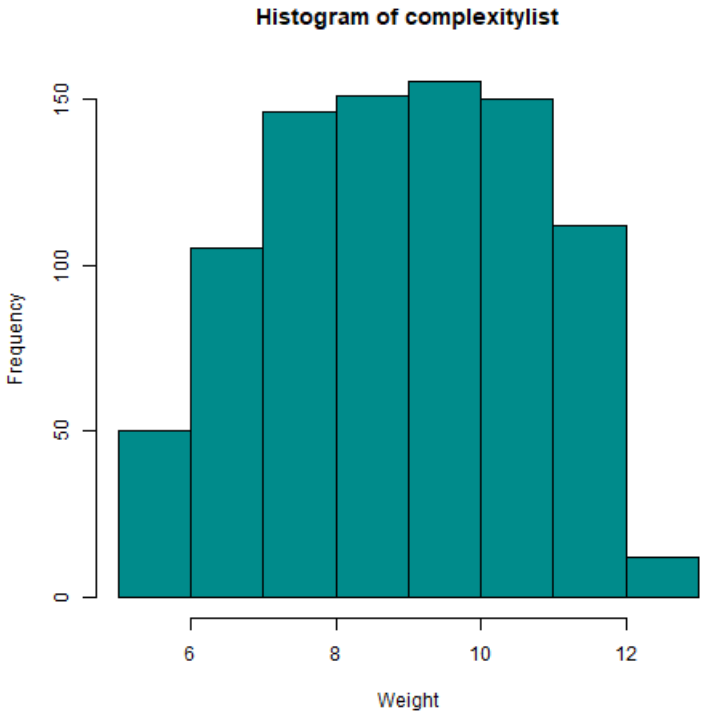

Figure 4. Frequency of amount of interactions for the 714 obtained unique topologies.

3.3.1 Conclusions

With this preliminary analysis, it is possible to identify potential mechanisms behind pattern formation. Although they are very diverse, some of them share similar topological traits. Most of the GRNs have an intermediate number of itneractions. In natural systems not all the GRNs are possible. Gene transcription and translation is a process that can have fluctuations due to chance, if GRNs able to cope with fluctuations appear in natural systems, their probability to be selected is high. In the virtual population of GRNs, such circuits are those highly represented. There must be specific type of interactions behind this properties that may be the next step in this analyses.

References

[1] Y. Schaerli, A. Munteanu, M. Gili, J. Cotterell, J. Sharpe, M. Isalan, *Nature Communications* **Sept. 2014**, *5*, DOI 10.1038/ncomms5905.

[2] J. C. Arboleda-Rivera, G. Machado-Rodríguez, B. A. Rodríguez, J. Gutiérrez, *PLOS Computational Biology* **Feb. 2022**, *18*, (Ed.: D. M. Umulis), e1009704.