RESYS Project

Differentiation of hematopoietic precursors in embryos

M. Jeremie Perrin
M2 BIM student
ENS Cachan

M. Corbin Hopper
M2 MPRI student

2019 – 2020

The Dataset

The Early Hematopoietic process

The biological process we are studying is one of differentiation of multipotent precursor cells *Hemangioblasts* into hematopoietic and endothelial cells alike. In the mouse embryo, there is emergence of blood cells at day 7 in the yolk sac and this is the marker of the beginning of the hematopoiesis. Both the Hematopoietic Stem Cells (HSC) and the Endothelial Stem Cells (EPC) originate from these multipotent precursor cells. The blood cells will be formed by differentiation of the HSCs while the vasculature (tissue of the blood vessels) will be formed through differentiation of the ESCs.

The Experiment

In the paper *Decoding the Regulatory Network for Blood Development from Single-Cell Gene Expression Measurements*, single-cell gene expression data was used in order to infer the regulatory networks at work during the hematopoietic differenciation process. It is known that blood development initiates at gastrulation from mesodermal cells, which initially have the potential to form blood, endothelium and smooth muscle cells. They showed that single-cell analysis of a developing organ coupled with computational approaches can reveal the transcriptional programs that control organogenesis.
In order to acquire the data necessary they sampled single-cells in *in vivo* mice embryos. The cells were sampled from the mesoderm and their potential to differentiate into blood cells was asserted thanks to expression of Flk1 and Runx1 expression. The sampling was done at four distinct time points.
Those four times points define groups of cells, which are not hemogeneous since the differentiation process is asynchronous. That is to say some cells begin their differentiation process earlier than others :

E7.00 At this time point the cells are labeled "PS"
E7.50 ———————————————— "NP"
E7.75 ———————————————— "HF"
E8.25 At this time points cells were categorized into two different set. Those cells which expressed *GFP* were labeled "4SG" and where considered as putative blood cells while those that did not where labeled "4SFG" and considered as putative endothelial cells.

At each time points gene expression of a set of genes was measured in each cell, these genes were selected by hand as they were known to play a role in the process. Forty-six genes were selected, out of those : four were housekeeping genes in order to assess the quelity of the measures. Nine were markers known to identify the different cell states and thirty-three were transcription known to play a role in the transcriptional program underlying

the differentiation process.

Categorizing Genes

To be pertinent in our analysis we first need to categorize the genes into the different cell states they belong to. We use the literature to guide us in our task. In the article, the authors underlign groups of genes as being caracteristic of the two end states :

- For Hematopoietic Cells : Hbb-bH1, Gata1, Nfe2, Gfi1b, Ikzf1 (Ikaros) and Myb
- For Endothelial Cells : Erg, Sox7, Sox17, Hoxb4, Cdh5

In order to build our own gene categories, we have proceeded in the following way. Each gene is represented by 5 values, those are the mean expression at each sampling points. Before doing unsupervised clustering we wanted to see if the genes were spatially grouping in terms of when they were expressed and how much they were expressed. We therefore ran a Principal Component Analysis (see Fig. 1).Using this two dimensional representation of the data we were able to see that the genes characteristic of the different end stages grouped together. We separated the data into three clusters using K-Means algorithm (see again Fig. 1).

We then plotted the actual mean expression values of the different genes grouped by

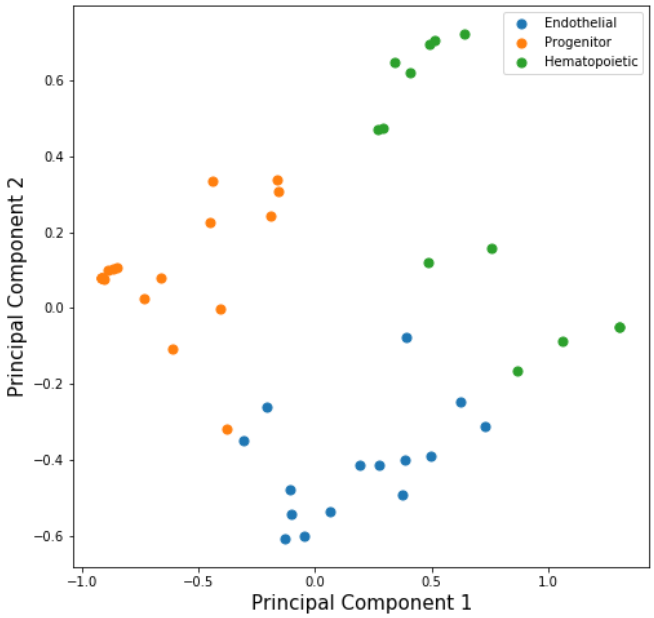

FIGURE 1 – Genes represented after PCA transformation and grouped by unsupervised clustering into three clusters.

their infered categories (see Fig). We clearly see the distinction in between groups, the clear expression of genes categorized as Hematopoietic in the 4SG stage as well as the expression of Endothelial genes in the 4SFG stage.

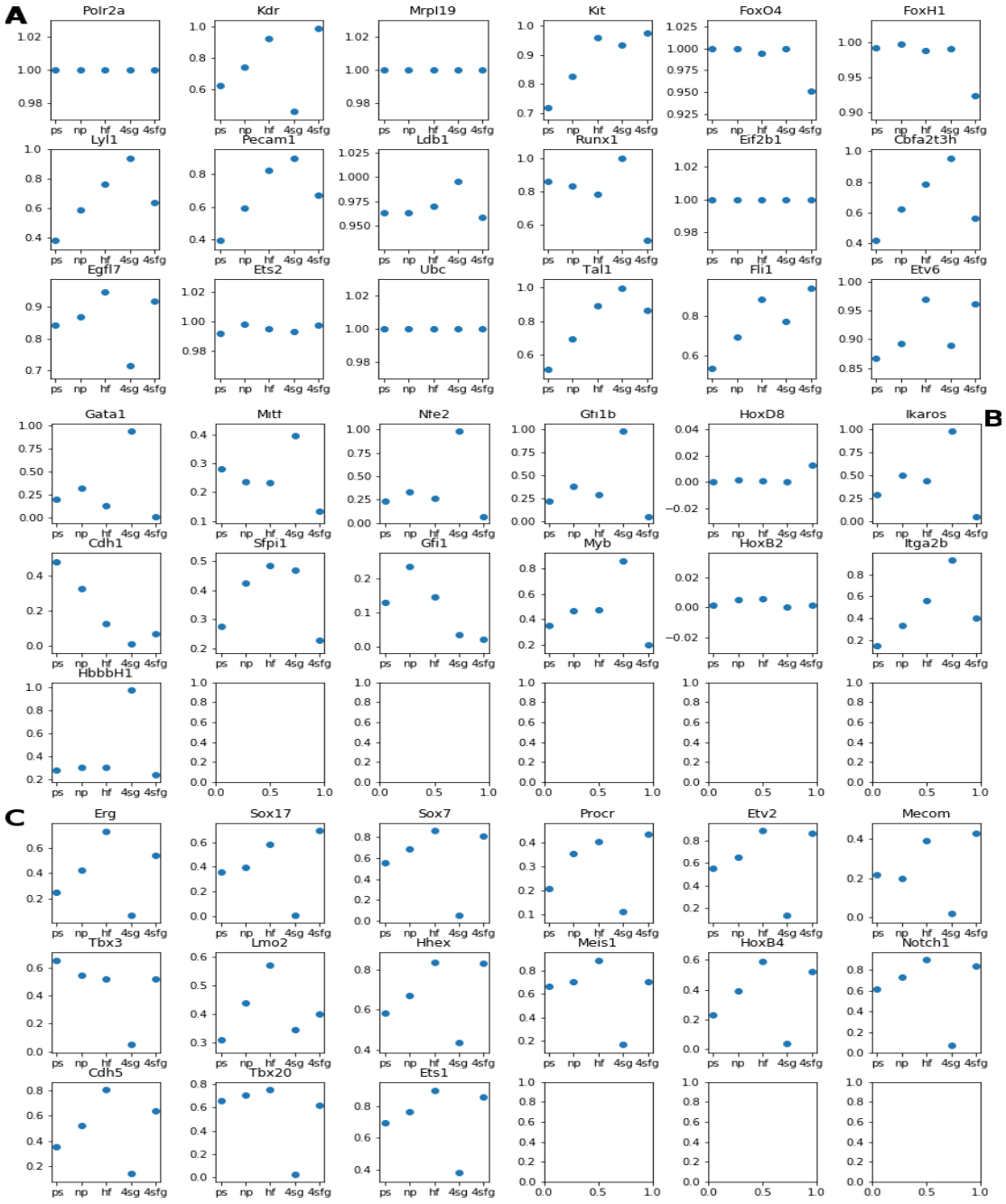

FIGURE 2 – Mean gene expression values throughout the experiment. **A** Precursor genes **B** Hematopoietic genes **C** Endothelial genes

Using MIIC

We then turned to MIIC to infer a network from the dataset provided. Now that we've categorized the genes and trimmed those that did not provide enough information. We downloaded the tool from gitHub and compiled the sources. We present the resulting graph in Fig. 3.

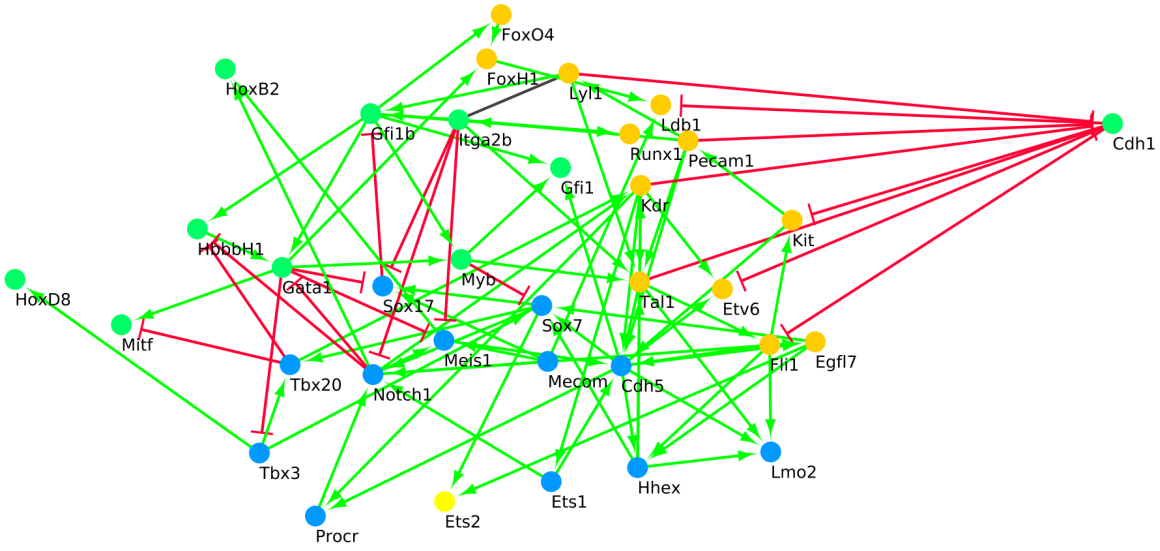

FIGURE 3 – Network Inferred by MIIC. Precursor genes, Hematopoietic genes and Endothelial genes

Discussion The network clearly represents the hematopoietic process as seen by the repression links infered between the genes classified as hematopoietic and those classified as endothelial. The Cdh1 gene behaves as a regulator (through repression) of the other precursor genes, if we look back at its gene expression levels we see that the expression drops throughout the differentiation process. It leads us to believe that it was wrongly categorized as an hematopoietic gene, that it should rather be categorized as a precursor gene.

Their Network Inference Method

The network inference tool used in the article was specially developed for their purpose. And indeed it achieves good results. Moreover it is explanatory, in the sense that it provides

the update function of each gene as a boolean fucntion. We were first tempted to implement it but we finally decided, the amount of work was excessive. Still, we found it interesting to describe their method.

State-transition graph The number of cell is great enough to consider we have all the states the cells can be in. Every pair of states that differ in the expression of exactly one gene are connected to form the state-transition graph.

Network Inference Then they search for the direction of the edges as well as the Boolean update function for each of the genes. To do so, they translate their problem into a SAT problem.

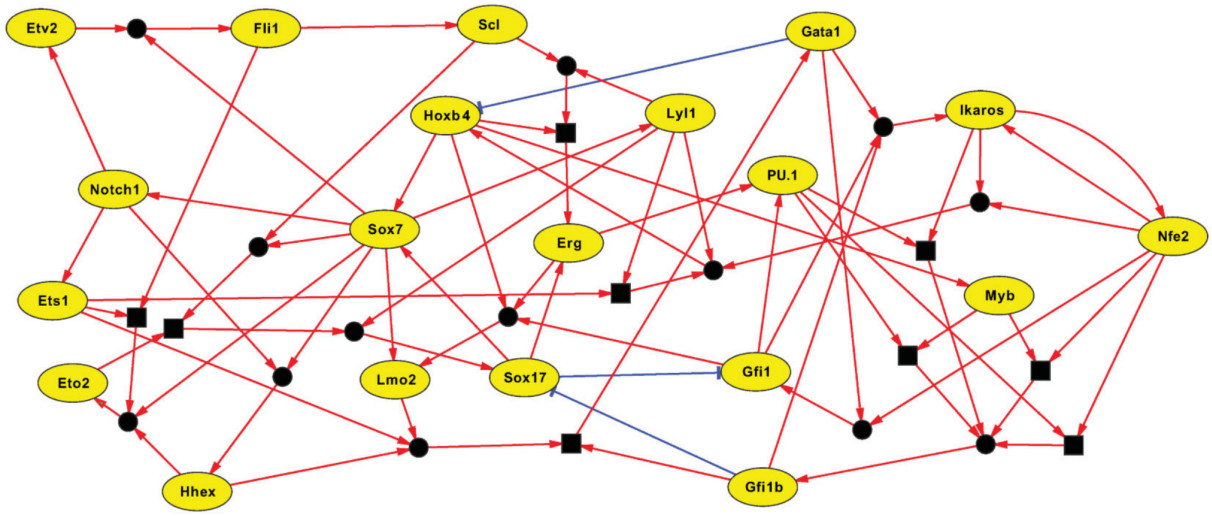

FIGURE 4 – Network constructed by the SCNS toolkit

Discussion One thing interesting about this type of boolean network is that you can find its stable state. A stable state will be assigning a value of expression (ie expressed or not) to each gene such that all boolean equalities are satisfied. We would expect the final cells (those that have already specified) to be stable state of the network. If the 4SG cells are indeed stable state (or some of them at least) then it does tell that the network is accurate. Also one such network would allow to simulate the evolution of a cell, it would be interesting to know how well this network fares in doing so.