



HAL
open science

A novel Boolean network inference strategy to model early hematopoiesis aging

Léonard Hérault, Mathilde Poplineau, Estelle Duprez, Élisabeth Remy

► **To cite this version:**

Léonard Hérault, Mathilde Poplineau, Estelle Duprez, Élisabeth Remy. A novel Boolean network inference strategy to model early hematopoiesis aging. Computational and Structural Biotechnology Journal, 2022, 21, pp.21-33. 10.1016/j.csbj.2022.10.040 . hal-04277075

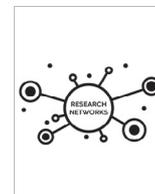
HAL Id: hal-04277075

<https://hal.science/hal-04277075>

Submitted on 9 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



A novel Boolean network inference strategy to model early hematopoiesis aging

Léonard Héroult ^{a,b}, Mathilde Poplineau ^b, Estelle Duprez ^{b,1}, Élisabeth Remy ^{a,*,1}

^a Aix Marseille Université, CNRS, Marseille I2M, France

^b Epigenetic Factors in Normal and Malignant Hematopoiesis Team, Aix Marseille Université, CNRS, INSERM, Institut Paoli-Calmettes, CRCM, Marseille, France



ARTICLE INFO

Article history:

Received 12 July 2022

Received in revised form 24 October 2022

Accepted 24 October 2022

Available online 02 November 2022

Keywords:

Aging

Hematopoietic stem cells

Single-cell RNA seq

Gene regulatory network Inference

Boolean modelling

ABSTRACT

Hematopoietic stem cell (HSC) aging is a multifactorial event leading to changes in HSC properties and functions, which are intrinsically coordinated and affect the early hematopoiesis. To better understand the mechanisms and factors controlling these changes, we developed an original strategy to construct a Boolean model of HSC differentiation. Based on our previous scRNA-seq data, we exhaustively characterized active transcription modules or regulons along the differentiation trajectory and constructed an influence graph between 15 selected components involved in the dynamics of the process. Then we defined dynamical constraints between observed cellular states along the trajectory and using answer set programming with in silico perturbation analysis, we obtained a Boolean model explaining the early priming of HSCs. Finally, perturbations of the model based on age-related changes revealed important deregulations, such as the overactivation of *Egr1* and *Junb* or the loss of *Cebpa* activation by *Gata2*. These new regulatory mechanisms were found to be relevant for the myeloid bias of aged HSC and explain the decreased transcriptional priming of HSCs to all mature cell types except megakaryocytes.

© 2022 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Hematopoiesis is the process of cellular differentiation that allows the hematopoietic stem cell (HSC) to produce all types of mature and functional blood cells. A critical balance between HSC self-renewal and differentiation into different hematopoietic lineages must be maintained throughout an individual's life in order to maintain an effective immune system and normal oxygen transport. As with many biological systems, transcriptional regulations orchestrated by transcription factors (TFs) and their networking are key mechanisms for instructing differentiation occurring in the hematopoietic stem and progenitor cell (HSPC) compartment (reviewed in [1,2]). The characterization of their activities has led to refine our view of the multiple branch points for early hematopoiesis [3].

It is also well known that deregulations of transcriptional and epigenetic programs underlie the decline in HSC function during aging [4,5]. This leads to an alteration of the HSPC pool phenotype, resulting in an increase in myeloid and megakaryocytic cells at the expense of lymphoid and erythroid ones in aged individuals [6,7].

As a consequence, elderly people are subject to various blood disorders such as anemia and acute myeloid leukemia. Given the current aging of the population, deciphering the molecular mechanisms and more particularly the gene regulatory networks (GRNs) underlying age-induced deregulation of HSPCs is of great interest and is the subject of extensive research.

With recent technology developments allowing single-cell resolution transcriptome analysis and lineage tracing, hematopoiesis is now considered as a continuous process with a very early and gradual priming of the HSPC compartment into different lineages [8,9]. Comparative transcriptome studies from young and aged HSPCs at the single-cell level (scRNA-seq) have accurately mapped lineage priming and cell cycle changes in aged mice, allowing the identification of subgroups of HSPCs distinct in their ability to maintain early hematopoiesis and whose proportions are altered during aging [10–12]. With thousands of gene expressions measured in thousands of cells, scRNA-seq also provided the amount of data needed to significantly improve GRN inference methods [13,14]. Some inference methods, based on mutual information or regression trees, have been successfully used to analyze regulatory networks in the HSC microenvironment [15]. They have also permitted the identification of regulons, modules formed by a TF and its target genes, in the HSPC compartment during human ontogeny [16] or mouse aging [10]. However, these studies only pro-

* Corresponding author.

E-mail address: elisabeth.remy@univ-amu.fr (É. Remy).

¹ These authors contributed equally: Elisabeth Remy, Estelle Duprez.

vide a static view of the GRN governing HSC fate. It would be relevant to have a dynamic view of the molecular mechanisms and interactions involved in cellular decisions such as commitment to a particular lineage. To address these issues, one possibility is to study the dynamics of networks using Boolean network (BN) modeling [17]. BN modelling approach provides a good abstraction of the long-term behaviors of a biological system, although continuous changes in component activities and timing of regulations are not captured. It also provides mechanistic explanations on the functioning of regulatory processes without the need for kinetic parameters and is therefore particularly suitable for the analysis of large biological networks [18]. In the context of hematopoiesis, several logical models of HSC differentiation have been proposed, which helped us to understand the connection between the major TFs specifying hematopoietic lineage differentiation [19–21].

The recent development of scRNA-seq technology has opened up new possibilities and challenges in the field of BN modeling. Indeed, the scRNA-seq data represent observations of a large number of cell states that can be ordered along a pseudo-trajectory reflecting the temporal organization of the cells according to variations in their gene expression from an initial state to one or more terminal or differentiated states. Binarization of the component activities of interest (usually gene expressions) provide discrete observations of the model whose connections are interpreted as possible trajectories generated by the BN. It is then possible from the transitions between states observed in the data to find logical functions for each component by a reverse engineering approach, in view of the observed dynamics [21,22]. More recently a method, called BoNesis, has been developed to infer, from a gene regulatory network, a BN satisfying dynamic constraints between cell states [23] and it is most likely that scRNA-seq data through the analysis of the pseudo-trajectory are well suited to extract the dynamical constraints used as input for BoNesis.

Based on single-cell RNA-seq (scRNA-seq) analysis, we and others recently observed new priming events in the HSC pool [8,10] which are deregulated with aging. Here, we wanted to take advantage of our scRNA-seq data of young and aged mouse HSPCs [10] to construct a BN to understand the dynamic of early priming of HSCs and to precisely characterize transcriptional determinants leading to HSC dysfunctions and aging. We first defined the key states of early hematopoiesis by selecting and grouping HSPCs according to their transcriptome in coherence with their TF marker activity. Next, combining our analyses with the current literature on HSC biology, we adapted an existing GRN of early myelopoiesis [24], by inferring with BoNesis a BN whose dynamics corresponds to our pseudo trajectory of HSC priming. Finally, we performed and analyzed perturbations of this BN to propose key factors and mechanisms supporting the differentiation bias observed in aged HSCs. Overall, our results provide a mathematical model of early hematopoiesis that allows us to assess the regulation of age-related physiological changes in HSCs.

Thus, our novel strategy judiciously combines recent inference tools and prior knowledge to recapitulate the known biology of early hematopoiesis while providing new insights into its transcriptional regulation. We applied it here in the context of hematopoietic aging, but it could be adapted to other biological processes modeled by Boolean networks.

2. Materials and methods

2.1. scRNA-seq dataset

We used the scRNA-seq dataset presented in our previous study available in the Gene Expression Omnibus database under acces-

sion code GSE147729 [10]. This dataset is composed of two pools of young (2/3 months) mouse HSPCs and two pools of aged (18 months) mouse HSPCs. Our previous results (cell-cycle phase assignment, cell clustering, pseudotime ordering and hscScore [25]) were considered in this study to define the HSPC states at the basis of our modeling [10] (Supplementary Table 1).

2.1.1. Regulon analysis

Identification of regulons with pySCENIC. We used the Single-Cell Regulatory Network Inference and Clustering (SCENIC) approach to identify regulons [26], which are modules of one TF and its potential targets, and their activities. We ran SCENIC workflow using pySCENIC v1.10.0 with its command line implementation [27], as in our previous study regarding gene filtering, TF motifs (motifs-v9-nr.mgi-m0.001-o0.0), cis-target (+/- 10 kb from TSS mm9-tss-centered-10 kb-7species.mc9nr) databases and command line options [10]. In this work, we used as input all 1721 TFs with a motif available in the motif database. We processed with SCENIC workflow all cells together as well as only young or only aged cells. For each cell set, regression per target step with grnboost2 followed by cis-target motif discovery and target pruning were run 50 times using a different seed for the pySCENIC grn command. The regulons and their targets recovered in at least 80 % of the runs were kept.

For a gene g with n regulators (r_1, \dots, r_n) , the normalized interaction score (NIS) of the transcriptional regulation of g by r is defined as follows:

$$NIS(r_i, g) = \frac{IS(r_i, g)}{\sum_{j=1}^n IS(r_j, g)}$$

where $IS(r_i, g)$ is the interaction score defined as the product of the number of SCENIC runs in which the interaction from r_i to g was found, by the average importance score given by grnboost2 for the interaction across these scenic runs. The results for the interactions found in pySCENIC analysis of all cells are available in Supplementary Table 2.

Regulon markers of HSPC states. We scored the activating regulons (i.e., regulons with a positive correlation between the TF and its targets) with AUCell (pySCENIC auctell command, default option with a fixed seed). Averaged AUCell scores by HSPC states were computed. These scores were standardized in order to hierarchically cluster the regulons using ward.D2 method of the R function hclust with Euclidean distance. The DoHeatmap function from the Seurat v3 package [28], was used to display the results. Averaged AUCell enrichment scores for young and aged cells by HSPC states were also computed in the same way.

Activating regulon markers of HSPC states were identified based on their AUCell scores using FindAllMarkers Seurat function (min.pct = 0.1, logfc.threshold = 0) with Wilcoxon rank sum tests. Only regulons with an average AUCell score difference above 0.001 between one state versus all the others were kept. A p-adjusted value (Bonferroni correction) threshold of 0.001 was applied to filter out non-significant markers (Supplementary Table 3).

Activating regulon activity differences with aging in each state were identified using the FindConservedMarkers Seurat function (sequencing platform as grouping variable, min.pct = 0.1 and logfc.threshold = 0) with Wilcoxon rank sum tests. For each HSPC state, only average AUCell score differences of the same sign and above 0.001 in the two batches presenting a combined p value < 0.001 were kept (Supplementary Table 4).

TF network. A network based on interactions between TFs found in at least 90 % of SCENIC runs on all cells was built (discarding self-inhibitions because of their uncertainty [26]). The cluster_louvain function, from igraph R package [29] was used to find TF communities in the undirected transformation of this net-

work with edges weighted by the NIS scores. The Cytoscape software was used to visualize the results from graph clustering [30].

2.2. Cistrome database analysis

Available mouse TF ChIP-seq experiments annotated for bone marrow tissue were analyzed using Cistrome database workflow [31]. More specifically, for each BED file of the selected experiments in the databases, the top 10,000 peaks with more than 5-fold signal to background ratio were conserved for downstream analysis. Then, target transcripts were identified with BETA in each TF experiment [32]. We considered all TF peaks in an experiment j inside a +/- 10 kb window from a Transcriptional Start Site (TSS).

BETA gave us a regulatory score s_j for each TSS_i^g of potential target genes g of a TF t . Then we defined a global cistrome regulatory score (CRS) for a TF t on a potential target gene g as follows:

$$CRS(t, g) = \frac{m}{N} \sum_{j=1}^m \sum_{i=1}^{n_m} s_j(t, TSS_i^g)$$

where, the TSS_i^g are the n_m TSSs of g for which a regulatory score s_j by t is obtained in experiment j among the m experiments where the regulation is found. This score is weighted by the m, N ratio where N is the number of experiments for the given TF available in the considered cistrome datasets. Only CRS for Scenic interactions or referenced regulations were retained (Supplementary Tables 2 and 5).

2.3. Boolean modeling

A **Boolean Network** (BN) is an influence graph parameterized with logical functions. An influence graph is a directed signed graph that is an abstraction of regulatory and molecular interactions (in binary relations). Nodes stand for biological components (here TFs and cell-cycle protein complexes) that are connected through edges representing activations and inhibitions. In this study, we constructed an influence graph of 15 nodes and their interactions involved in early hematopoiesis (see results). From this influence graph we define a discrete dynamical model using logical formalism. Each node of the influence graph is associated with a Boolean variable representing its level that can be 0 (component inactive) or 1 (active), this level reflecting its ability to regulate its targets. The effect of regulators on the level of the target node is expressed through **logical functions** (using connectors & for AND, | for OR and ! for NOT). Given a **configuration of the network**, i.e., a vector containing the level of all the components, several components may be called to update their level by the logical functions. The choice of the updating policy defines the trajectories of the system (succession of consecutive configurations). Here, we used the Most Permissive (MP) semantics (that is required to use the BoNesis inference tool). This recently proposed semantic considers additional transient states reflecting increasing (\uparrow) or decreasing (\downarrow) dynamical states. A component in (\uparrow) or (\downarrow) state can be read non-deterministically as either 0 or 1 to take into account the uncertainties of its actual influence thresholds on its different targets. This generates a non-deterministic dynamic with a large set of trajectories, which considerably reduces the complexity of the exploratory analysis of the dynamics [33].

The **attractors** of the model capture the asymptotic behaviors of the system. They are a set of configurations from which it is not possible to escape, and can be fixed points, i.e., a configuration whose all components are stable, or cyclical attractors, containing more than two configurations among which the system oscillates. Finally, a biological interpretation of these attractors, based on the level of some nodes or read-outs of the model, allows them to be associated with biological phenotypes.

BN offer the possibility to easily simulate **perturbations of gene activity**, such as a gain-of-function or overexpression denoted KI (Knocked In), or a loss-of-function or deletion denoted KO (Knocked Out), by maintaining its variable at 1 or 0 respectively.

We can also simulate an edgetic mutation by perturbing not a node but an edge of a network. For that, we removed the edge of the network and updated the logical rules of the target nodes.

2.4. Data discretization

We associated each HSPC state with a vector, called **meta-configuration**, representing the discretized activity level of each of the 15 components of the model. The discretization method depends slightly on the nature of the components:

- For TFs with more than 10 targets we applied a Kmeans clustering with $K = 2$ on all cell regulon activity scores. For each HSPC state the value of the cluster (active or inactive) with the most cells was retained.
- For TFs with <10 targets (Tal1, Ikzf1 and Zfp1), because the AUCell scores were less reliable, we applied a Kmeans clustering with $K = 3$, active (1), inactive (0) or free/unknown (*) on averaged RNA levels per HSPC states. This was the case for Tal1, Ikzf1 and Zfp1.
- For the two cell-cycle complexes (CDK4/6CycD and CIP/KIP), we chose to discretize the RNA levels of their genes as for the TFs with few targets on three levels: active (1), inactive (-1) or free/unknown (0). Then, if the sum of gene values for each complex was above 1, we considered the complex as active, below 1 as inactive.

In order to be less constraining, we relaxed some component activities by replacing their discretized value 0 or 1 to a free (*) status (according to data and biological assumptions, see results).

2.5. Boolean network inference with BoNesis

The influence graph, meta-configurations and dynamical constraints (expressed for instance in terms of stability or reachability of meta-configurations of the network) were encoded in Answer Set Programming (ASP) language with the BoNesis tool which solves our Boolean satisfiability problem and enumerates all the possible Boolean models that satisfy the constraints in the MP semantic [23].

To reduce the number of solutions we chose to limit the number of clauses per logical rule to 3, as it is the case for most of the logical rules of gene regulatory BN. The solver Clingo was used in the inference steps [34].

A subset of 1000 BNs representing a variety of possible behaviors was selected during the generation of the solution space by the Clingo solver, as reported in [35]. For each of them, in silico KO perturbations were performed one by one in the aim of recovering some mutant phenotypes previously described experimentally. In the same way, in silico KI perturbations for nodes with a TF activity upregulated upon aging were conducted. This provides new mutant constraints matching literature evidence for the following next inference steps (Supplementary Table 6).

The influence graph was then pruned by adding two optimizations to reduce the number of possible solutions: in priority a maximization of the confident interaction (see Results & Supplementary Table 5) number and then a minimization in the other interaction numbers in the inferred models.

2.6. Dynamical analysis of Boolean networks

Dynamical analysis (e.g. attractors reachability from iHSC state, (un)reachabilities between states) of the inferred Boolean models was done in the MP semantics with the `mpbn` python package [33].

2.7. Statistics

Statistics were computed with R software v4.0.2. The statistical tests for regulon activity scores were performed with Seurat and are detailed above. In each HSPC state corresponding to the considered configurations of the model, the enrichment for young or aged cells was tested using a hypergeometric test (`phyper` R function).

3. Results

3.1. Regulon analysis identified distinct HSPC states with specific transcription factor activities and their interactions

In order to get references for establishing dynamical constraints for the inference method of BN, we defined HSPC states. We chose to take into account three layers of information, extracted from our previous scRNA-seq analyses [10], considering that each of them is importantly linked to HSPC functionality: cell cluster identity (a meaningful functional partition of the HSPCs), cell-cycle phases (distinguishing the dividing HSPCs) and pseudotime trajectory (a good representation of HSPC priming toward different lineages) with its pseudotime, in agreement with the HSCscore [25], measuring the stemness of hematopoietic cells profiled with scRNA-seq (Supplementary Fig. 1). Hence, we defined nine states that were visualized on the pseudotime trajectory (Fig. 1A). We considered two HSPC states at the beginning of the pseudotime trajectory (pseudotime <2); one was composed of non-cycling cells (without a *tgf* signature of quiescence, see below) and was called the initiating HSC state (iHSC) and the other one was composed of cells in the G2/M phase and was then considered as the self-renewing HSC state (srHSC). We considered three states based on their cluster identity; the *ifn*HSC state gathering cells of the *ifn* cluster (interferon response signature), a state gathering all cells of the *tgf* cluster that we named the quiescent HSPC state (qHSC state) as all of the cells (except one) were in G1/G0 phase and the *preDiff* state gathering the cells of the *diff* cluster representing and spreading on the core of the trajectory [10]. Finally, we defined four lineage-primed HSPC states based on their position at the terminal branches of the trajectory and their belonging to the lineage-primed clusters: pLymph (primed lymphoid clusters pL1 on branch 2), pNeuMast (primed neutrophils and primed mastocytes clusters gathered together, on branch 4), pEr (primed erythrocytes, on branch 5) and pMk (primed megakaryocytes, on branch 5). The selection of these 9 states, excluding cells from our original data set (Supplementary Table 1), resumed the initial (iHSC, srHSC), transient (qHSC, *ifn*HSC), terminal (pEr, pMk, pNeuMast, pLymph) and branching (*preDiff*) states of early hematopoiesis. Thus, we provide an accurate view of the key states that an HSC can reach during early hematopoiesis.

To functionally characterize these HSPC states, we studied their regulons using the SCENIC workflow [27]. We identified 197 activating and 132 inhibiting regulators (Supplementary Table 2). Among them, 140 were regulon markers of at least one of the 9 HSPC states (see **Methods regulon marker analysis**; Supplementary Table 3). Next, by quantifying the regulon activities with the AUCell enrichment score [26], and performing a hierarchical clustering, we revealed a specific regulon activity profile for each of the HSPC states (Fig. 1B). Regulon activities supported the HSPC state identity. *Klf1* was active in pEr, *Gata1* in pEr and pMk, *Spi1*

and *Cebpa* in pNeuMast and *Ikzf1* and *Zbtb16* in pLymph. We observed *Stat* and *Irf* regulon activity in *ifn*HSC; *Gata2*, *Junb*, *Egr1* and *Klf1* in qHSC and *Bclaf1* and *Srf* in respectively iHSC and srHSC states. *Fli1* regulon was active in both iHSC and pMk. The *preDiff* state was marked with *Spi1* and *Myc* regulons, two factors involved in HSPC commitment. Except for the srHSC state, uniquely marked by *Zbtb7a*, probably due to its low cell number, each state was characterized by a combination of regulons, consistent with its transcriptional and lineage feature [10].

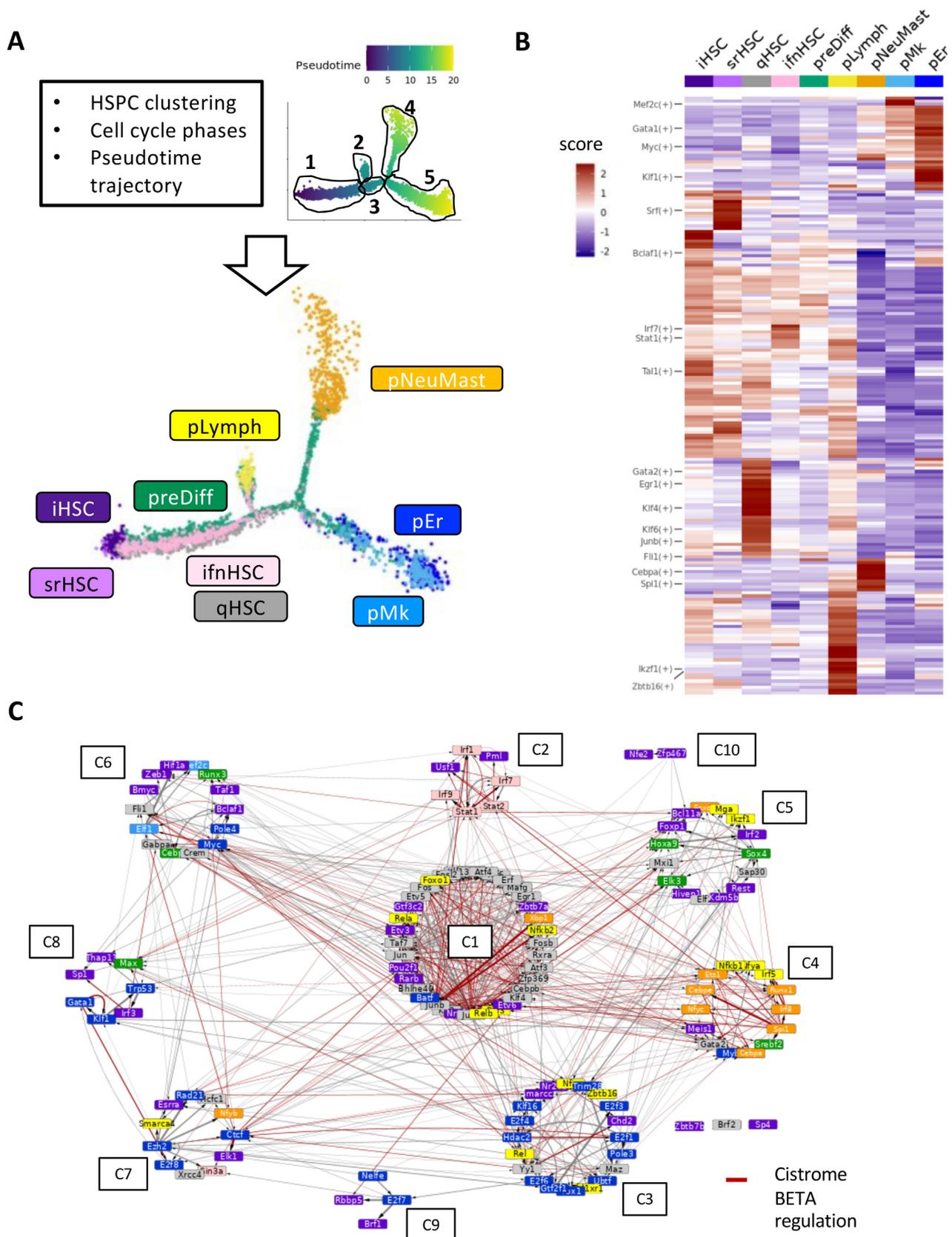
Next, to connect the regulons to each other, we built a transcriptional network whose nodes are TFs at the head of regulons significantly marking at least one of the HSPC states, and directed edges represent the transcriptional regulations between them. When considering only the reliable transcriptional regulations (found in 90 % of the SCENIC runs) and after removing auto-regulations, we obtained a directed graph of 133 nodes (TFs) and 670 edges (regulations) (Fig. 1C, Supplementary Table 2). We further confirmed these regulatory interactions by analyzing the presence of TF peaks in the regulatory regions of their target genes using ChIP-seq data from the Cistrome database [31]. Approximately 60 % (302) of the network interactions with an available TF node in the Cistrome database were confirmed by the presence of a peak in the regulatory regions of its targets (Supplementary Table 2). We then performed a clustering analysis by weighting the network using a normalized interaction score (NIS, cf Materials and Methods) calculated from the SCENIC results and by applying Louvain clustering. We underlined 10 regulon communities and three isolated regulons (*Zbtb7b*, *Brf2*, *Sp4*). By associating each TF in the network to the most relevant HSPC state, we observed that half of the communities regroups TFs whose activity characterizes the same HSPC state (Fig. 1C, Supplementary Table 3). Indeed, most of TFs from the C1 community (*Klf* factors, *Jun* and *Fos* AP-1 factors, *Egr1*) are known to be related to quiescence and their regulons mark the qHSC state, whereas the C2 community contained mainly TFs marking *ifn*HSC state (eg *Irf1-7-9*, *Stat1-2*). In the same way, C3 community was associated with pEr state, C4 community with pNeuMast and C5 with iHSC (*Kdm5b*, *Foxp1*) and *preDiff* (*Sox4*, *Hoxa9*) states. It was more difficult to define the smaller communities (C6 to C10) as they presented a more heterogeneous composition of TFs.

Altogether, our analysis revealed a functional relevance of the 9 HSPC states harboring a specific transcriptional activity. We also revealed a well-structured interconnection of TFs that supports the HSC differentiation journey starting with the i-srHSC states, continuing through the transient HSC states (*ifn*-qHSC, *preDiff*) and ending with one of the four lineage-primed HSC states.

3.2. Inference of a Boolean network to model HSC priming

To decipher the key molecular mechanisms governing HSC fate, we constructed a Boolean gene network. We developed a strategy based on BoNesis, a recently developed approach for Boolean network inference [23], which relies on two steps: the synthesis of an influence graph and the definition of dynamical constraints.

For the influence graph synthesis, we built a gene network based on a previous published Boolean model of early myeloid differentiation [24]. This model provided megakaryocyte and erythrocyte stable states and a granulo/monocyte branching state that fits well with our defined states [24]. We extracted from it a subgraph of 9 relevant TFs and their mutual interactions. Eight of them were regulon markers of the HSPC states in our analysis: *Gata1* a marker of pEr and pMk; *Fli1* of pMk; *Klf1* of pEr; *Spi1* and *Cebpa* of pNeuMast; *Tal1*, *Fli1* and *Gata2* of qHSC (Supplementary Table 3). We also selected *Zfp1*, the cofactor of *Gata1*, which was expressed in pEr and pMk HSPC states (Supplementary Fig. 2). To adapt the graph to early HSC commitment and aging, we added *Ikzf1*, a TF



involved in early lymphoid specification in HSPC [36] and whose regulon marked pLymph state according to our analysis (Supplementary Table 3). We also added two components that allow the HSC to regulate the cell cycle: the CDK4/6-CyclinD (CDK4/6CycD) complex (*Ccnd1-3* and *CDK4/6* genes) required for the HSC quiescence exit, and its inhibitory complex CIP/KIP (*Cdkn1a-b-c* genes) driving the quiescence of the HSCs [37]. To connect CIP/KIP complex to the transcriptional network, we added Junb and Egr1, two TF involved in HSC quiescence [38–40], and identified in our regulon analysis as markers of qHSC state and activators of CIP/KIP genes (Fig. 1B; Supplementary Tables 2 and 3). Finally, to connect CDK4/6CycD to the network, we added Myc and Bclaf1 known to be involved in HSC cell cycle [41,42]. Both were active regulons in the preDiff state whereas only Bclaf1 regulon was active in srHSC state and had CDK4/6CycD complex genes as targets (Fig. 1B; Supplementary Tables 2 and 3). To connect all these nodes, we considered interactions from the original model [24] that we complemented with interactions identified by SCENIC and in the literature.

Finally, we obtained an influence graph with 15 components and 60 interactions (Fig. 2Ai), more than 75 % of which were confirmed by at least two of the following information sources; SCENIC, literature or Cistrome (Supplementary Fig. 3A; Supplementary Table 5A–C). The complexity of a modeling study increases with the number of nodes, so we added the minimum number of components that allow us to capture the HSPC states identified in the data. Thus, we did not retain many TFs known to play a role in hematopoiesis, such as Runx1, Erg, Hhex, Smad6, Eto2 or Ets1. The results of SCENIC showed that the TFs retained in our model are not regulated them (Supplementary Table 2) suggesting that they probably have little role in the early priming of HSCs that we study in this work.

For the discretization of the data, we have assigned to each HSPC state a *meta*-configuration, i.e., a vector representing the discretized activity level of each of the 15 components (see Methods and Fig. 2Aii). This discretization turned out to be too strict regarding the first set of constraints. Thus, we decided to release empirically some constraints on *meta*-configurations by attributing a free (*) state to some nodes. First, we allowed the cycling configurations CDK4/6CycD in pMk and CIP/KIP in pLymph to be free since they were linked to a HSPC state composed of cells in different cell-cycle phases. Following the same idea, we let free the CDK4/6CycD activators, Bclaf1 in pLymph and Myc in pNeuMast, pER and pMk. We also found that the pNeuMast states presented a bimodal activity for Gata2, with this gene marking pMast and not pNeu cells (see Supplementary Table 1 in [10]). Thus, we let Gata2 free in pNeuMast HSPC states. Finally, we also let free Egr1 in srHSC in agreement with a previous study suggesting its role in HSC maintenance in the hematopoietic niche [38].

In order to infer with BoNesis a BN whose dynamics fits with our pseudo-trajectory of HSC priming, we enunciated dynamical constraints between the HSPC states. We required that the model presented at least four fixed points, one for each of the 4 primed *meta*-configurations pLymph, pNeuMast, pER and pMk, reachable

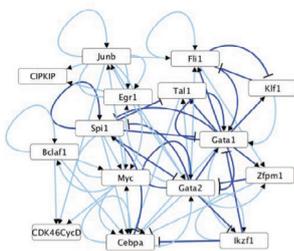
from iHSC. We also added a zeros configuration in which all the components are inactive, reachable only from iHSC directly. We allowed a cell to go back and forth from the iHSC state to the srHSC or qHSC state as suggested by the literature [43,44]. Based on the trajectory and the differentiation committed state of preDiff, we considered this state as a “no return state” and blocked its return to the iHSC state. From this state, any of the 3 primed fixed points pNeuMast, pER, and pMk were accessible. We allowed a cell from iHSC to directly reach the pLymph fixed point based on the shape of the pseudo-trajectory and the high hscScore of the pLymph cells [10]. All these constraints are resumed in the HSC differentiation journey (Fig. 2Aiii).

With this first inference, BoNesis provided a set of over 100,000 solutions. We therefore developed a strategy to refine the solution search and get a Boolean network solution (Fig. 2B). We added constraints by considering mutant behaviors: we retained solutions whose mutant simulations agree with the biological phenotypes of these same mutants previously described in the literature (see Materials and Methods, and Supplementary Table 6 for the references): the *Ikzf1* KO conducting to an absence of pLymph fixed point, the *Spi1* KO to absences of both pNeuMast and pLymph fixed points, the *Klf1* KO to an absence of pER fixed points and the *Junb* KO to an apparition of an additional proliferative (active CDK4/6CycD complex) pNeuMast fixed point. We also considered KI perturbations on *Egr1* and *Junb* nodes, as these components were previously found upregulated in HSC upon aging [11]. For these KIs, we observed a loss of reachability of all fixed points except a quiescent (CIP/KIP active) pMk one consistent with the HSC priming bias we previously described [10]. These 6 altered behaviors are resumed in Supplementary Fig. 4 and were added to the BoNesis constraint set (Fig. 2Bi). Next, we performed a graph pruning that consists in reducing the number of edges in the influence graph by favoring the most confident ones, which were chosen based on strong literature supports (Supplementary Table 5A). There remained 36 interactions (Fig. 2Bii), of which more than 80 % were supported by at least two sources of information among SCENIC, Cistrome and literature (Supplementary Fig. 3B). We required solutions containing all these 36 interactions and another run of BoNesis provided 616 solutions (to compare with the 10^{22} possible solutions with the initial influence graph according to the Dedekind number [45]) (Fig. 2Biii). These solutions differed on the logical rules of 4 nodes (Supplementary Table 7): CDK4/6CycD, *Fli1*, *Gata1* (2 inferred rules for each) and *Gata2* (77 inferred rules) and needed a manual curation (Fig. 2Biv). For the CDK4/6CycD, we chose the rule making the activation possible through Myc in the preDiff state or through Bclaf1 in the srHSC state. For *Fli1*, we chose the logical rule containing the least number of clauses (2). For *Gata1*, we chose the rule for which the auto activation is possible only when the two repressors *Ikzf1* and *Spi1* are inactive. Finally, for *Gata2* among the 77 possibilities, 7 contains only two clauses and we chose the one in which the inhibition by *Gata1* and its co-factor *Zfp1* are present in both clauses. Thus, we obtained the final Boolean network presented in (Fig. 2C and D). Regarding the computational cost of the infer-

Fig. 1. Regulon analysis identified distinct HSPC states with specific transcription factor activities and interactions. **A** Upper panel: HSPC states are defined according to results of cell clustering, cell cycle phase assignment and pseudotime trajectory analysis of scRNA-seq data¹⁰. On the right, cells are ordered on the pseudotime trajectory and are coloured according to their pseudotime value. The 5 branches of the trajectory are circled. Lower panel: pseudotime trajectory where cells are colored according to their HSPC state: initial HSCs (iHSC, dark violet); self-renewal (schHSC, violet); quiescent (qHSC, gray); interferon (ifnHSC, pink); differentiation (preDiff, green). And the primed states: lymphoid (pLymph, yellow); neutrophils and mastocytes (pNeuMast, orange); erythrocytes (pER, dark blue) and megakaryocytes (pMk, blue). **B** Heatmap of the average AUCell scores of the regulon activity in each HSPC state. The scores were standardized and used to cluster regulons hierarchically. **C** Transcriptional regulation network of the regulon markers of the HSPC states. Regulons were clustered in 10 communities (from C1 to C10) plus 3 isolated nodes with Louvain graph clustering. Node color highlights the states where the regulon is the most active (same color code as in Fig. 1A). Red (resp. grey) edges indicate transcriptional regulations that are (resp. are not) supported by peak analysis in the Cistrome database. Edge thickness represents the normalized interaction score (NIS) obtained from SCENIC. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

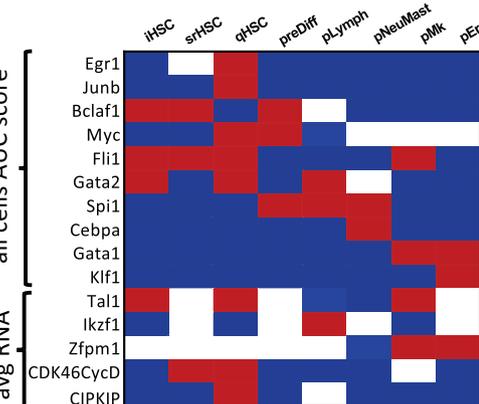
A

i Influence graph

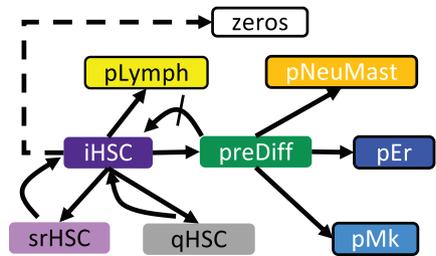


- 15 nodes
- 60 interactions
- >10²² possible models

ii Data discretization



iii Dynamical constraints

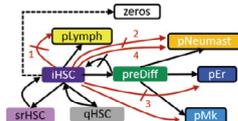


> 100 000 BN solutions

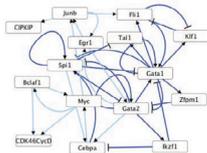
B

Perturbations (KO/KI) in 1000 solutions

i Adding constrained mutant behaviors



ii Influence graph pruning



iii Bonesis inference

616 solutions

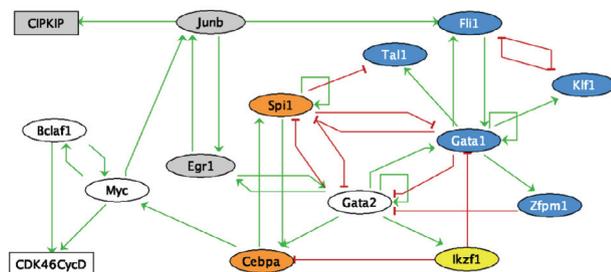
iv Manual curation

Final solution

C

Component	Logical rules
<i>Egr1</i>	$Gata2 \wedge Junb$
<i>Junb</i>	$Egr1 \vee Myc$
<i>Bclaf1</i>	<i>Myc</i>
<i>Myc</i>	$Cebpa \wedge Bclaf1$
<i>Fli1</i>	$Junb \vee (Gata1 \wedge Klf1)$
<i>Gata2</i>	$(Gata2 \wedge \overline{Gata1} \wedge \overline{Zfp1}) \vee (Egr1 \wedge \overline{Gata1} \wedge \overline{Zfp1} \wedge \overline{Spi1})$
<i>Spi1</i>	$(Spi1 \wedge \overline{Gata1}) \vee (Cebpa \wedge \overline{Gata1} \wedge \overline{Gata2})$
<i>Cebpa</i>	$(Gata2 \wedge \overline{Ikzf1}) \vee (Spi1 \wedge \overline{Ikzf1})$
<i>Gata1</i>	$Fli1 \vee (Gata2 \wedge \overline{Spi1}) \vee (Gata1 \wedge \overline{Ikzf1} \wedge \overline{Spi1})$
<i>Klf1</i>	$Gata1 \wedge \overline{Fli1}$
<i>Tal1</i>	$Gata1 \wedge \overline{Spi1}$
<i>Ikzf1</i>	<i>Gata2</i>
<i>Zfp1</i>	<i>Gata1</i>
<i>CDK46CycD</i>	$Bclaf1 \vee Myc$
<i>CIPKIP</i>	<i>Junb</i>

D



ence process, once the initial influence graph built from SCENIC results and prior knowledge, the graph pruning optimization followed by the final solution inference with all remaining edges could be performed in <5 min on a desktop computer with 32 GB of RAM and an i7 core, demonstrating the reasonable computational resources cost of our inference strategy and its applicability to other BN construction from analyzed scRNA-seq data combined with prior knowledge.

Finally, coupling a customized implementation of BoNesis with multiple sources of biological knowledge allowed to reduce the large number of possible solutions. We successfully synthesized a Boolean network of early hematopoiesis based on a regulatory network consisting of 15 components and 36 interactions.

3.3. Analyses of the Boolean model evidence a sequence of transcriptional events to prime HSCs

Simulations of the model were done within the MP semantic. The dynamics displayed 5 fixed points whose complete descriptions are given Fig. 3A. According to our requests, all fixed points were reachable from iHSC, regardless of the initial value of the Zfp-m1 component. We verified that published mutants related to the genes of the model could be recovered by the simulations. To do this, we simulated the corresponding KO perturbations in the model and compared the results of the simulations to the expected behaviors described in the relevant publications, particularly regarding the reachability of HSPC configurations from iHSC. The large majority of the *in-silico* KO simulations matched the corresponding *in-vivo/in-vitro* perturbations reported in the literature (Supplementary Table 6). For example, *in silico* *Fli1* KO (simulations) led to the loss of pMK fixed point from iHSC in agreement with the *Fli1* KO BM that harbors a megakaryopoiesis defect [46].

We conducted a fine analysis of the trajectory space to highlight events that are responsible for some salient dynamical properties along the trajectory. We observed that *Gata2* was active in the initial state iHSC (and also in qHSC and pLymph), inactive when the cell reaches preDiff and cannot be re-activated. This event may explain the early branching of the trajectory from the iHSC to the pLymph state (Fig. 3B), which was characterized by the activity of *Ikzf1* whose only regulator is the activator *Gata2* (Fig. 2D). We highlighted a novel transient state, named pME and described in Fig. 3A, that can reach pEr and pMk configurations but not pNeuMast configuration. We observed that the pME configuration was reachable from the preDiff configuration, when *Junb* was activated and *Spi1* inactivated (Fig. 3B).

Interestingly, the choice between pMk and pEr fixed points relied on the *Fli1*-*Gata1*-*Klf1* circuit (Fig. 3C), and on a transient state of *Fli1* caught thanks to the MP semantics. Indeed, starting from the branching point pME in which the three components of the module were absent, *Fli1* activity could increase (thanks to the presence of *Junb*) allowing *Gata1* to be also activated and led to the stable configuration pMk. Moreover, as long as *Fli1* had

not reached its activity level allowing it to inhibit *Klf1*, activation of *Klf1* by *Gata1* could occur and led to pEr configuration (Fig. 3B). It is important to note that we were able to capture this cascade of events thanks to the MP semantics that considers the intermediate states between 0 and 1. It means that a necessary condition to reach pEr from pME is that the inhibition threshold of *Klf1* is greater than the activation threshold of *Gata1*. Finally, the cross-inhibitory circuit involving *Fli1* and *Klf1* acted as a switch to maintain the differentiation between pMk and pEr (Fig. 3C).

Furthermore, our model presented a proliferation configuration, in which *CDK4/6CycD* and *Myc* were active and *CIP/KIP* inactive (Supplementary Fig. 5A). This configuration was accessible from the iHSC state, and all fixed points could be reached from this state (Supplementary Fig. 5B). This is in agreement with our previous results showing an increase in the proliferation of HSC during their priming toward different lineages [10].

To summarize, the dynamical analysis of our MPBN of early hematopoiesis gives new insights about the succession of early priming events in HSCs. It highlights a decisive role of *Gata2* inactivation to reach preDiff at the expense of pLymph from iHSC. The *Spi1* inactivation together with *Junb* activation are key events to reach from preDiff the pME branching point, whose commitment to the pMk or pEr states depends on the fine tuning of *Fli1*.

3.4. Perturbations of early hematopoiesis model explain some HSC aging features

Our previous single cell RNA-seq analysis revealed an alteration of HSC priming with an accumulation of quiescent HSCs in aged mice at the expense of pLymph, pEr and pNeuMast cells [10]. To decipher the molecular mechanisms and TFs responsible for this alteration, we simulated perturbations in the inferred Boolean network according to alterations observed in the transcriptome of aged HSPCs.

Alterations of regulon activity linked to aging were identified by comparing, for each HSPC state, the regulon activities between young and aged cells. Regulon transcriptional activity differences were found mainly (80 %) in the non-primed iHSC, ifnHSC, qHSC states with similar amounts of decrease and increase in activity (Supplementary Fig. 6), and very few in pNeuMast and pEr. Almost all activity alterations of regulons were found in more than one state (Supplementary Table 4). Aging features consisted mainly in a decrease of the activity in regulons related to HSC activation (*Runx3*, *Sox4*, *Myc* and *Spi1*) and NF-kappaB pathway (*Rel* and *Nfkb* factors), and an increase in regulons from the AP-1 complex (*Atf*, *Jun* and *Fos* factors) and involved in quiescence of HSCs (*Egr1*, *Klf* factors, *Gata2*, Supplementary Fig. 7 & Supplementary Table 4). To be noted that we observed a specific increase in *Cebpe-b* regulon activity in qHSC state marking the myeloid bias of these quiescent aged cells. Eight of the 13 TF components of our models were altered upon aging in their regulon activities (*Myc*, *Spi1*, *Junb*, *Egr1*, *Fli1*, *Klf1*, *Gata2* and *Gata1*, Supplementary

Fig. 2. Inference of a gene Boolean network to model HSC priming. **A** Inference steps performed with wild-type constraints. **(i)** A first influence graph is retained taking into account the possible interactions of the components deduced from the literature and the SCENIC results. Interactions with a high (low) confidence level are in dark (pale) blue. **(ii)** Table representing the discretisation of the 15 components in the 9 configurations. Blue indicated active, red inactive and white free state. Red (resp. blue) hatched cases mark node activities freed from 1 (resp 0) to * in the final configuration settings compared to the discretized data. **(iii)** Graph representation of the dynamical constraint imposed between the configurations (nodes). Arrows (resp. crossed out arrows) indicate reachability (resp. unreachability) between source and target configurations. Framed configurations are constrained as fixpoints. Dashed line highlights the reachability of the fixpoint with all node activities at 0 from iHSC. **B** Workflow of the strategy used to refine the search of solution and obtain a final solution. **(i)** Updating of the influence graph after the consideration of constraints coming from mutant behaviors. For the updated constraints see **supplementary Fig. 3A**. **(ii)** Pruning of the influence graph through maximization of high-confident interactions and minimization of others. **(iii)** As a last inference step, we forced the use of all remaining edges, this provided 616 possible solutions **(iv)**. A manual curation is necessary to obtain the final model. **C** Logical rules of the Boolean model. **D** Gene regulatory network of the Boolean model. Nodes, rectangular for cell cycle complexes and ellipse for TFs, are colored according to the HSPC states in which they are highly active according to our regulon analysis: gray for qHSC, yellow for pLymph, orange for pNeuMast, blue for pMk and pEr, white for the nodes highly active in several HSPC states. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

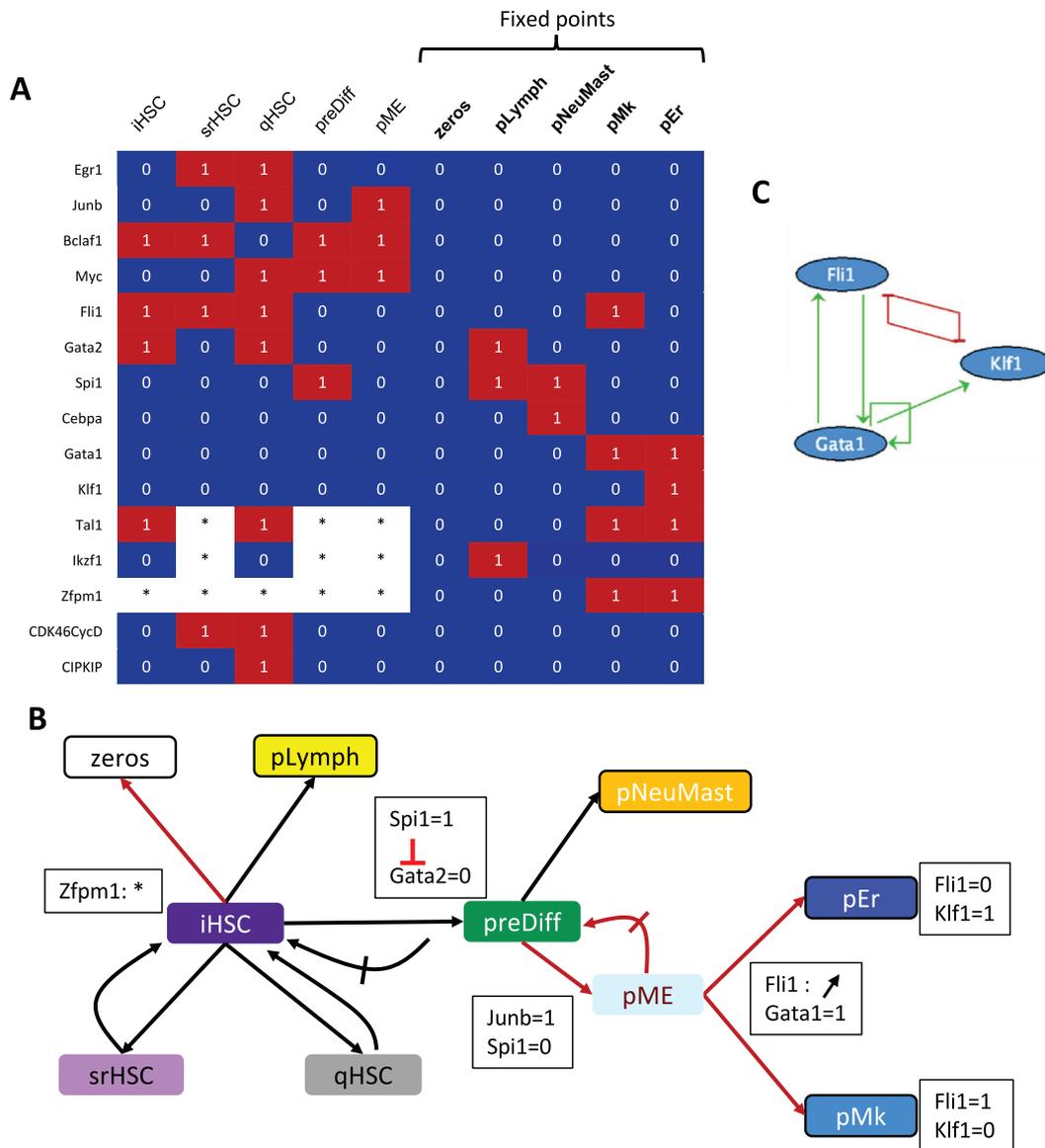


Fig. 3. Analyses of the Boolean model evidence a sequence of transcriptional events to prime HSCs. **A** Table describing the configurations of the model matching the HSPC states (column: HSPC states, lines: components of the model). Colors represent the activation levels of the nodes (blue: inactive; red: active, white: free). The five last columns are the fixed points of the model. pME configuration (5th column) results from the analysis of the model. **B** Graph representation of the (non)reachabilities between the configurations. Framed configurations represent fixed points, arrows (resp. crossed arrows) indicate reachability (resp. unreachability) from their source to their target configurations. Black arrows are constrained dynamic properties whereas the red ones result from the dynamic study of the model. The annotations in black boxes represent TF activities read in the dynamics, Zfpm1: * highlights the two possible values of this node in iHSC. Irreversible inactivation of Gata2 by Spi1 in the preDiff non-return configuration. necessary update of Junb (=1) and Spi1 (=0) to reach the configuration pME from preDiff. In MP semantics, from pME an increasing activity of Fli1(↑) can first activate Gata1 and then inhibit Klf1. Thus, depending on whether Gata1 activates Klf1 before it is inhibited by Fli1, pEr is reached rather than pMk. **C** Regulatory motif involving Gata1, Fli1 and Klf1 of the BN with a cross-inhibitory circuit between Klf1 and Fli1 maintaining HSC priming to pMk or pEr. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

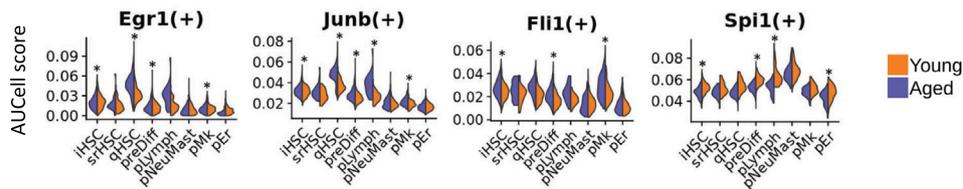
Table 5B). More precisely we found Junb, Egr1 and Fli1 (resp. Spi1) activities significantly increased (resp. decreased) in at least three of the 8 HSPC states considered for the model inference (Fig. 4A).

To identify possible altered TF regulations, we compared for each regulation the normalized interaction score (NIS, cf Materials and Methods) of young and aged cells analyzed separately using SCENIC workflow (Supplementary Fig. 8) and computed a score difference between young and aged conditions (Supplementary Table 4). The distribution of these score differences showed that most of the regulations were not strongly altered (14 % of the interactions have a score difference above 0.4; Supplementary Fig. 9). When focusing on the interactions of the model supported by SCENIC, we noticed an alteration of Cebpa activation by Gata2

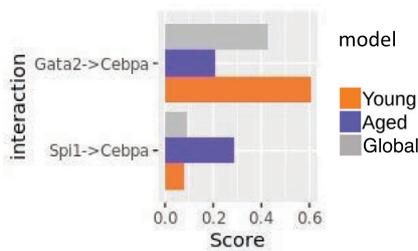
(decrease of the NIS of 0.4 upon aging), which was compensated by Spi1 activation of C/EBPalpha (NIS increase by 0.2) upon aging (Fig. 4B).

Thus, in order to simulate aging alteration, we performed KI perturbations on Junb, Egr1 and Fli1, KO perturbation on Spi1, and an edgetic mutation (loss of Cebpa activation by Gata2) on the network of early hematopoiesis (Fig. 4C). The simulations of each of these 5 perturbations led to the loss of reachability of the fixed points pLymph and pNeuMast from the i-sr-qHSC configurations. Additionally, the simulation of each of the 3 KIs led to the loss of reachability of pEr, which makes pMk the unique reachable fixed point (Table 1). Note that, still for these two perturbations, the constrained fixed point pMk is quiescent (CIP/KIP active). The

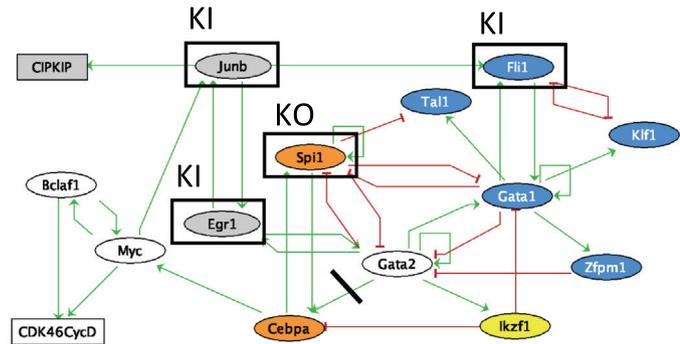
A



B



C



D

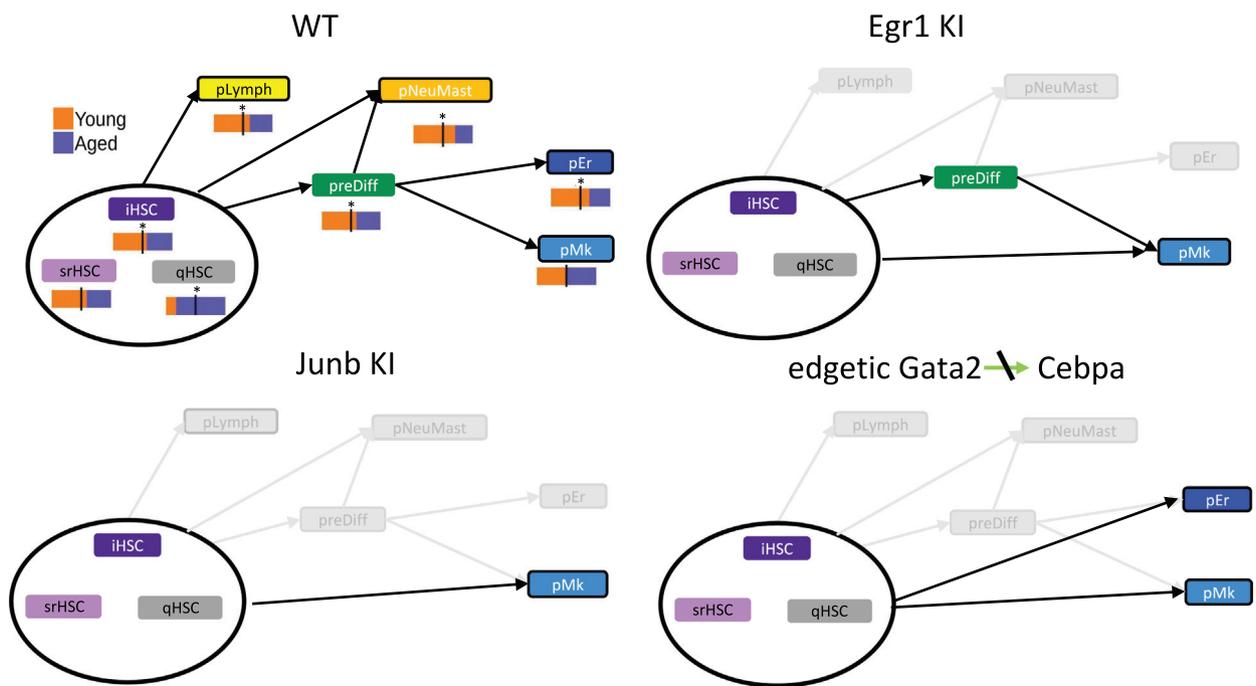


Fig. 4. Perturbations of the early hematopoiesis model explain some HSC aging features. **A** Combined violin plots of most altered TF (of the model) activities upon aging in young (orange) and aged (purple) cells from the different HSPC states. Stars show significant differences of activity score between young and aged cells (average difference > 0.001 and p value < 10⁻³). **B** Normalized interaction scores of *Cebpa* activation by *Spi1* and *Gata2* from SCENIC multiple runs on all cells (grey), young cells (orange) and aged cells (purple). **C** Aging perturbations of the Boolean gene network. Rectangular nodes are cell cycle complexes and ellipse nodes TFs. Nodes are colored according to the HSPC states in which they are highly active according to our single cell analysis: gray for qHSC, yellow for pL, orange for pNeuMast, blue for pMk and pEr, white for the nodes highly active in several HSPC states. Framed nodes highlight the 4 TF significantly altered with aging and crossed out activation of *Cebpa* by *Spi1* illustrates its edgetic mutation. **D** Reachability of HSPC states from any initial configuration from iHSC, srHSC or qHSC for WT and 3 altered dynamics of the model: **WT** case (top left) Young (orange) and aged (purple) cell proportion is given below each HSPC state node. A star highlights a significant shift from the global young/aged cell proportions in the single cell data (hypergeometric test p value < 0.05). **Egr1 KI** perturbation (top right); **Junb KI** perturbation (bottom left); **Cebpa edgetic** mutation (bottom right). In each graph, black arrows represent the reachabilities between configurations; pale gray represent the WT reachabilities lost with the mutation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

behaviors of *Egr1* KI, *Junb* KI and *Spi1* KO were expected as they were constrained for the inference of the model regarding previous experimental studies (see model inference strategy above). These results agree with our single cell analysis, as the 3 fixed points pLymph, pNeuMast, pEr correspond to the primed HSPC states

whose cell proportion significantly decreases with aging, while pMk remains reachable in any of our model perturbations and does not present any decrease in cell proportion with aging in the single cell data (Fig. 4D). We also observed that preDiff was no longer accessible from i-sr-qHSC configurations with *Junb* KI disruption

Table 1

Aging perturbations of the early hematopoiesis Boolean model. This table summarizes the reachabilities of 5 HSPC states (preDiff and 4 fixed points pLymph, pNeuMast, pEr, pMk) from states iHSC, srHSC or qHSC, in WT and 5 mutant simulations (a cross indicates that the HSPC-state is reachable). The last column reports the observations in our scRNA-seq data: Up (resp. down) arrows indicate an increase (resp. a decrease) upon aging in component activity or interaction score.

	pLymph	pNeuMast	pEr	pMk	preDiff	scRNAseq observations
WT	X	X	X	X	X	
Junb KI				X		Junb ↑ in iHSC, qHSC, preDiff, pLymph, pMk
Egr1 KI				X	X	Egr1 ↑ in iHSC, qHSC, preDiff, pMk
Fli1 KI				X	X	Fli1 ↑ in iHSC, preDiff, pMk
Spi1 KO			X	X	X	Spi1 ↓ in iHSC, preDiff, pLymph, pEr
Edgetic Gata2-Cebpa			X	X		Gata2 → Cebpa ↓ Spi1 → Cebpa ↑

and the edgetic Cebpa-Gata2 mutation, suggesting that HSC priming to pMk in aged mice follows an alternative differentiation pathway. This pathway could be directly derived from the state of qHSC cells, the proportion of which increases with aging (Fig. 4D) and were found at the end of the first branch of the pseudo-trajectory near the appearance of the first pMk (branch 3 and beginning of branch 5 of pseudotime trajectory Fig. 1A). The model emphasizes a clear dependence between the 5 perturbations related to aging (. Indeed, *Egr1* KI implies a definitive activation of *Junb* which in turn activates *Fli1*. Besides, the KO of *Cebpa* activation by *Gata2* prevents *Spi1* from becoming active from any of the i-sr-qHSC configurations. Thus, analysis of the model highlighted the major roles of *Egr1* overexpression and loss of *Cebpa* activation by *Gata2* as two major molecular mechanisms that led to HSC aging resulting in the decrease in all lineage priming except the megakaryocyte one.

4. Discussion

Aging alters the mechanisms governing the balance between self-renewal and differentiation of HSCs, which are the guarantee of functional hematopoiesis. Aging effects are difficult to define as they probably depend on multiple factors and regulations. To address HSC age-related alteration, we used Boolean networks, which provide modelling tools suitable for explanatory analysis of dynamical processes such as HSC differentiation process and its response to alterations in the interaction network. Several BNs have been proposed to understand the key regulatory elements of HSC differentiation to lymphoid or myeloid progenitors [19–21] but very few have addressed the impact of aging on HSC properties [47]. Moreover, none of them specifically addressed the early priming of the HSCs, which has been recently emphasized by the development of scRNA-seq analyses [8,10]. While models of BN are classically built from literature, more recent approaches to infer BN from scRNA-seq data have been applied to hematopoiesis [21,22] albeit presenting some bias due to the imprecision of the pseudotime values because of hidden variables (cell location, epigenetic modification, etc [48]) of scRNA-seq.

In this work, we provide a complete workflow for building a Boolean model from scRNA-seq data. While the methods proposed so far infer the influence graph and the logical rules together [21,22], we decoupled the problem into two parts, with first the construction of an influence graph, and then an inference of logical rules for each component.

GRN inference benchmark has shown that methods that require a quantitative ordering as input tend to perform worse than those that do not [14]. Decoupling the problem into two steps, we were able to construct a trustworthy influence graph by combining one of the best GRN inference methods SCENIC [26], and prior knowledge of our biological process. The use of BoNesis in the second step permitted us to avoid the need of a precise quantitative ordering of the cells that remain actually very challenging and open to

criticism as pseudotime ordering results can be highly variable depending on the tools and the parameters used [49]. Although the BoNesis approach clearly seems well suited for scRNA-seq analysis, the original publication [23] is mostly theoretical and we are the first to our knowledge to provide a complete workflow for using it from scRNA-seq data combined with prior knowledge. In particular, our entirely new graph pruning optimization process allowed us to handle the large number of possible solutions provided by this method.

We obtained a BN explaining the transcriptional mechanisms controlling the priming of HSC toward the different hematopoietic lineages and the impact of its alterations during aging. We conducted an original qualitative analysis of the model's trajectories which consists of tracing succession of events leading to the priming of HSCs in the different lineages. We could propose a main path starting from the initial state iHSC and passing through the intermediate state preDiff that was not captured before. From the iHSC, activation of *Ikzf1* by *Gata2* stabilizes early lymphoid priming of HSCs, whereas activation of *Spi1* and *Myc* together with an inactivation of *Gata2* leads to preDiff state. From preDiff, the positive circuit *Gata1/Fli1/Klf1* controls the priming to the neutrophil/mastocytic lineage or the erythroid or megakaryocytic lineages. Our model reproduces the main differentiation trajectory of HSCs under normal conditions, but also shows that alternative trajectories are possible, not passing through preDiff, but allowing direct priming of HSCs towards the different lineages. For example, with early activation of *Gata1* by *Gata2* the system bypasses preDiff and leads directly to the primed megakaryocyte or erythrocyte state, which is in agreement with previous lineage tracing studies highlighting the coexistence of multiple hematopoietic hierarchies [8,50]. It should be noted that all our analyses were conducted within the MP semantics. This recently defined semantics allows many more transitions between states than the classically used asynchronous or generalized semantics [33]. In particular, our analysis shows that the switch between pEr and pMk from the preDiff state depends on the existence of two thresholds of influence of *Fli1* on these targets *Klf1* and *Gata1*, and such a situation cannot be represented with the asynchronous Boolean dynamics. These situations provide perfect case studies to address the question of how the choice of semantics impacts the properties of the dynamics, and more specifically to identify the specific parts of the BN that would need to be refined in order for the MP trajectories to be reproducible in asynchronous semantic.

Our model correctly reproduces behaviors of mutants observed *in vivo/in vitro* for most of the TFs in the network, except for *Myc* and *Egr1* KO. Simulations of *Myc* KO showed no difference *in silico* in the accessibility of primed states. However, an increase in HSC self-renewal and a decrease in differentiation due to intercellular interactions not taken into account in our model have been reported experimentally with this mutation [41]. We also did not observe any changes in dynamics for the *Egr1* KO mutant although, again, a previous study showed a decrease in HSC priming along with an increase in HSC self-renewal [38]. These observations

could correspond to a transient accumulation of cells prior to delayed priming, not captured by the model.

The loss of lymphoid potential and the myeloid bias, mainly driven by a platelet bias [10,51] are the main features of aged HSCs [52–54]. Our model successfully reproduces this aging feature with new molecular mechanisms based on the overactivation of *Egr1* and *Junb*, or loss of *Cebpa* activation by *Gata2*. Our results highlight the overactivation of *Egr1* and *Junb* factors in quiescent myeloid HSCs that accumulate with aging [11], two factors that have been previously involved in HSC quiescence [38,39]. Our model shows that these alterations impact the positive circuit between *Egr1* and *Junb* required for the multiple HSC priming. Interestingly, the global transcriptional network inferred with SCENIC shows that these two factors are activated by *Klf2–4–6* factors known to be downstream of TGF- β signaling in other biological contexts [55–57]. We saw that aged cells forming the qHSC state have a strong TGF- β signature combined with a myeloid bias (overactivation of *Cebpe-b* in particular). Moreover, megakaryocytes are known to promote HSC quiescence by producing TGF- β [58,59]. Our results therefore sustain the hypothesis of a self-activating loop of HSC aging which would be triggered by TGF- β increase in aged HSC microenvironment favoring a myeloid-biased quiescent state from which a single priming towards the megakaryocytic lineage would occasionally be possible [60]. Our model proposed a path sustaining this single priming (without passing through preDiff), in agreement with a direct HSC differentiation into megakaryocytes that would therefore be the one that is preserved by aged quiescent HSCs [8]. Besides, our model shows that the loss of *Cebpa* activation by *Gata2* may drive the loss of lymphoid and neutrophil/mastocyte priming. Knowing the impact of aging on epigenetic [61], this edgetic alteration could have an epigenetic origin, due to changes either in histone marks on the regulatory elements of *Cebpa* [62], or in hypermethylation of the *Cebpa* promoter as found in leukemia cases [63].

Thus, we propose a novel model of the intracellular transcriptional network explaining the HSC early differentiation and its megakaryocytes bias related to aging. Analysis of this model could be quantitatively refined to reproduce the probabilities of the different HSC priming observed in the single cell data [64]. More globally, this model could be the basis of a modelling at the cell population level taking into account the HSC and its microenvironment, which is known to be important for HSC aging.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This work was supported by the Ligue Nationale contre le Cancer (E.D.), the Fondation A**MIDEX* (E.D. and E.R.), M.P. was supported by a Fondation de France and OPALE fellowships, L.H. was supported by a PhD grant from Aix Marseille University.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2022.10.040>.

References

- [1] Laiosa CV, Stadtfeld M, Graf T. Determinants of lymphoid-myeloid lineage diversification. *Annu Rev Immunol* 2006;24:705–38. <https://doi.org/10.1146/annurev.immunol.24.021605.090742>.
- [2] Mercer EM, Lin YC, Murre C. Factors and networks that underpin early hematopoiesis. *Semin Immunol* 2011;23:317–25. <https://doi.org/10.1016/j.smim.2011.08.004>.
- [3] Murre C. Developmental trajectories in early hematopoiesis. *Genes Dev* 2009;23:2366–70. <https://doi.org/10.1101/gad.1861709>.
- [4] Svendsen AF, Yang D, Kim KM, Lazare SS, Skinder N, Zwart E, et al. A comprehensive transcriptome signature of murine hematopoietic stem cell aging. *Blood* 2021. <https://doi.org/10.1182/blood.2020009729>.
- [5] Poplineau M, Verney J, Platel N, N'guyen L, Hérault L, Esposito M, et al. PLZF limits enhancer activity during hematopoietic progenitor aging. *Nucleic Acids Res* 2019;47:4509–20. <https://doi.org/10.1093/nar/gkz174>.
- [6] Geiger H, de Haan G, Florian MC. The ageing haematopoietic stem cell compartment. *Nat Rev Immunol* 2013;13:376–89. <https://doi.org/10.1038/nri3433>.
- [7] de Haan G, Lazare SS. Aging of hematopoietic stem cells. *Blood* 2018;131:479–87. <https://doi.org/10.1182/blood-2017-06-746412>.
- [8] Rodriguez-Fraticelli AE, Wolock SL, Weinreb CS, Panero R, Patel SH, Jankovic M, et al. Clonal analysis of lineage fate in native haematopoiesis. *Nature* 2018;553:212–6. <https://doi.org/10.1038/nature25168>.
- [9] Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol* 2017;19:271–81. <https://doi.org/10.1038/ncb3493>.
- [10] Hérault L, Poplineau M, Mazuel A, Platel N, Remy É, Duprez E. Single-cell RNA-seq reveals a concomitant delay in differentiation and cell cycle of aged hematopoietic stem cells. *BMC Biol* 2021;19:19. <https://doi.org/10.1186/s12915-021-00955-z>.
- [11] Kirschner K, Chandra T, Kiselev V, Flores-Santa Cruz D, Macaulay IC, Park HJ, et al. Proliferation Drives Aging-Related Functional Decline in a Subpopulation of the Hematopoietic Stem Cell Compartment. *Cell Rep* 2017;19:1503–11. <https://doi.org/10.1016/j.celrep.2017.04.074>.
- [12] Kowalczyk MS, Tirosh I, Heckl D, Rao TN, Dixit A, Haas BJ, et al. Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res* 2015;25:1860–72. <https://doi.org/10.1101/gr.192237.115>.
- [13] Hu X, Hu Y, Wu F, Leung RWT, Qin J. Integration of single-cell multi-omics for gene regulatory network inference. *Comput Struct Biotechnol J* 2020;18:1925–38. <https://doi.org/10.1016/j.csbj.2020.06.033>.
- [14] Pratapa A, Jalihal AP, Law JN, Bharadwaj A, Murali TM. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat Methods* 2020;17:147–54. <https://doi.org/10.1038/s41592-019-0690-6>.
- [15] Desterke C, Petit L, Sella N, Chevallier N, Cabeli V, Coquelin L, et al. Inferring Gene Networks in Bone Marrow Hematopoietic Stem Cell-Supporting Stromal Niche Populations. *IScience* 2020;23. <https://doi.org/10.1016/j.isci.2020.101222>.
- [16] Roy A, Wang G, Iskander D, O'Byrne S, Elliott N, O'Sullivan J, et al. Transitions in lineage specification and gene regulatory networks in hematopoietic stem/progenitor cells over human development. *Cell Reports* 2021;36:. <https://doi.org/10.1016/j.celrep.2021.109698>.
- [17] Thomas R. Boolean formalization of genetic control circuits. *J Theor Biol* 1973;42:563–85. [https://doi.org/10.1016/0022-5193\(73\)90247-6](https://doi.org/10.1016/0022-5193(73)90247-6).
- [18] Kauffman SA. Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol* 1969;22:437–67. [https://doi.org/10.1016/0022-5193\(69\)90015-0](https://doi.org/10.1016/0022-5193(69)90015-0).
- [19] Bonzanni N, Garg A, Feenstra KA, Schütte J, Kinston S, Miranda-Saavedra D, et al. Hard-wired heterogeneity in blood stem cells revealed using a dynamic regulatory network model i88. *Bioinformatics* 2013;29:i80. <https://doi.org/10.1093/bioinformatics/btt243>.
- [20] Collombet S, van Oevelen C, Ortega JLS, Abou-Jaoudé W, Stefano BD, Thomas-Chollier M, et al. Logical modeling of lymphoid and myeloid cell specification and transdifferentiation. *PNAS* 2017;114:5792–9. <https://doi.org/10.1073/pnas.1610622114>.
- [21] Hamey FK, Nestorowa S, Kinston SJ, Kent DG, Wilson NK, Göttgens B. Reconstructing blood stem cell regulatory network models from single-cell molecular profiles. *PNAS* 2017;114:5822–9.
- [22] Moignard V, Woodhouse S, Haghverdi L, Lilly AJ, Tanaka Y, Wilkinson AC, et al. Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat Biotechnol* 2015;33:269–76. <https://doi.org/10.1038/nbt.3154>.
- [23] Chevalier S, Froidevaux C, Paulevé L, Zinovyev A. Synthesis of Boolean Networks from Biological Dynamical Constraints using Answer-Set Programming. 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), 2019, p. 34–41. 10.1109/ICTAI.2019.00014.
- [24] Krumsiek J, Marr C, Schroeder T, Theis FJ. Hierarchical Differentiation of Myeloid Progenitors Is Encoded in the Transcription Factor Network. *PLoS ONE* 2011;6:e22649.
- [25] Hamey FK, Göttgens B. Machine learning predicts putative hematopoietic stem cells within large single-cell transcriptomics data sets. *Exp Hematol* 2019;78:11–20. <https://doi.org/10.1016/j.exphem.2019.08.009>.
- [26] Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* 2017;14:1083–6. <https://doi.org/10.1038/nmeth.4463>.
- [27] Van de Sande B, Flerin C, Davie K, De Waegeneer M, Hulselmans G, Aibar S, et al. A scalable SCENIC workflow for single-cell gene regulatory network

- analysis. *Nat Protoc* 2020;15:2247–76. <https://doi.org/10.1038/s41596-020-0336-2>.
- [28] Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive Integration of Single-Cell Data. *Cell* 2019;177:1888–1902.e21. <https://doi.org/10.1016/j.cell.2019.05.031>.
- [29] Csardi G, Nepusz T, et al. The igraph software package for complex network research. *InterJournal. Complex Systems* 2006, 1695:1–9.
- [30] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–504. <https://doi.org/10.1101/gr.1239303>.
- [31] Liu T, Ortiz JA, Taing L, Meyer CA, Lee B, Zhang Y, et al. Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol* 2011;12:R83. <https://doi.org/10.1186/gb-2011-12-r83>.
- [32] Wang S, Sun H, Ma J, Zang C, Wang C, Wang J, et al. Target analysis by integration of transcriptome and ChIP-seq data with BETA. *Nat Protoc* 2013;8:2502–15. <https://doi.org/10.1038/nprot.2013.150>.
- [33] Paulevė L, Kolčák J, Chatain T, Haar S. Reconciling qualitative, abstract, and scalable modeling of biological networks. *Nat Commun* 2020;11:4256. <https://doi.org/10.1038/s41467-020-18112-5>.
- [34] Gebser M, Kaminski R, Kaufmann B, Schaub T. Clingo = ASP + Control: Preliminary Report. *ArXiv:14053694 [Cs]* 2014.
- [35] Chevalier S, Noël V, Calzone L, Zinovyev A, Paulevė L. Synthesis and Simulation of Ensembles of Boolean Networks for Cell Fate Decision. 18th International Conference on Computational Methods in Systems Biology (CMSB), vol. 12314, Online, Germany: Springer; 2020, p. 193–209. 10.1007/978-3-030-60327-4_11.
- [36] Ng S-Y, Yoshida T, Zhang J, Georgopoulos K. Genome-wide lineage-specific transcriptional networks underscore Ikaros-dependent lymphoid priming in hematopoietic stem cells. *Immunity* 2009;30:493–507. <https://doi.org/10.1016/j.immuni.2009.01.014>.
- [37] Pietras EM, Warr MR, Passequé E. Cell cycle regulation in hematopoietic stem cells. *J Cell Biol* 2011;195:709–20. <https://doi.org/10.1083/jcb.201102131>.
- [38] Min IM, Pietramaggiore G, Kim FS, Passequé E, Stevenson KE, Wagers AJ. The transcription factor EGR1 controls both the proliferation and localization of hematopoietic stem cells. *Cell Stem Cell* 2008;2:380–91. <https://doi.org/10.1016/j.stem.2008.01.015>.
- [39] Santaguida M, Schepers K, King B, Sabnis AJ, Forsberg EC, Attema JL, et al. JunB protects against myeloid malignancies by limiting hematopoietic stem cell proliferation and differentiation without affecting self-renewal. *Cancer Cell* 2009;15:341–52. <https://doi.org/10.1016/j.ccr.2009.02.016>.
- [40] Passequé E, Wagner EF, Weissman IL. JunB deficiency leads to a myeloproliferative disorder arising from hematopoietic stem cells. *Cell* 2004;119:431–43. <https://doi.org/10.1016/j.cell.2004.10.010>.
- [41] Wilson A, Murphy MJ, Oskarsson T, Kaloulis K, Bettess MD, Oser GM, et al. c-Myc controls the balance between hematopoietic stem cell self-renewal and differentiation. *Genes Dev* 2004;18:2747–63. <https://doi.org/10.1101/gad.313104>.
- [42] Dell'Aversana C, Giorgio C, D'Amato L, Lania G, Matarese F, Saeed S, et al. miR-194-5p/BCLAF1 deregulation in AML tumorigenesis. *Leukemia* 2017;31:2315–25. <https://doi.org/10.1038/leu.2017.64>.
- [43] Wilson A, Laurenti E, Oser G, van der Wath RC, Bianco-Bose W, Jaworski M, et al. Hematopoietic Stem Cells Reversibly Switch from Dormancy to Self-Renewal during Homeostasis and Repair. *Cell* 2008;135:1118–29. <https://doi.org/10.1016/j.cell.2008.10.048>.
- [44] Bernitz JM, Kim HS, MacArthur B, Sieburg H, Moore K. Hematopoietic Stem Cells Count and Remember Self-Renewal Divisions. *Cell* 2016;167:1296–1309. e10. <https://doi.org/10.1016/j.cell.2016.10.022>.
- [45] Wiedemann D. A computation of the eighth Dedekind number. *Order* 1991;8:5–6. <https://doi.org/10.1007/BF00385808>.
- [46] Moussa O, LaRue AC, Abangan RS, Williams CR, Zhang XK, Masuya M, et al. Thrombocytopenia in mice lacking the carboxy-terminal regulatory domain of the Ets transcription factor Fli1. *Mol Cell Biol* 2010;30:5194–206. <https://doi.org/10.1128/MCB.01112-09>.
- [47] Schwab JD, Ikonomi N, Werle SD, Weidner FM, Geiger H, Kestler HA. Reconstructing Boolean network ensembles from single-cell data for unraveling dynamics in the aging of human hematopoietic stem cells. *Comput Struct Biotechnol J* 2021;19:5321–32. <https://doi.org/10.1016/j.csbi.2021.09.012>.
- [48] Watcham S, Kucinski I, Gottgens B. New insights into hematopoietic differentiation landscapes from single-cell RNA sequencing. *Blood* 2019;133:1415–26. <https://doi.org/10.1182/blood-2018-08-835355>.
- [49] Saelens W, Cannoodt R, Todorov H, Saeys Y. A comparison of single-cell trajectory inference methods. *Nat Biotechnol* 2019;37:547–54. <https://doi.org/10.1038/s41587-019-0071-9>.
- [50] Weinreb C, Rodriguez-Fraticelli A, Camargo FD, Klein AM. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* 2020;367. <https://doi.org/10.1126/science.aaw3381>.
- [51] Grover A, Sanjuan-Pla A, Thongjuea S, Carrelha J, Giustacchini A, Gambardella A, et al. Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells. *Nat Commun* 2016;7:11075. <https://doi.org/10.1038/ncomms11075>.
- [52] Dykstra B, Olthof S, Schreuder J, Ritsema M, de Haan G. Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells. *J Exp Med* 2011;208:2691–703. <https://doi.org/10.1084/jem.20111490>.
- [53] Yamamoto R, Wilkinson AC, Ooehara J, Lan X, Lai C-Y, Nakauchi Y, et al. Large-Scale Clonal Analysis Resolves Aging of the Mouse Hematopoietic Stem Cell Compartment. *Cell Stem Cell* 2018;22:600–607.e4. <https://doi.org/10.1016/j.stem.2018.03.013>.
- [54] Beerman I, Bhattacharya D, Zandi S, Sigvardsson M, Weissman IL, Bryder D, et al. Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. *Proc Natl Acad Sci* 2010;107:5465–70. <https://doi.org/10.1073/pnas.1000834107>.
- [55] Botella LM, Sanz-Rodríguez F, Komi Y, Fernandez-LA, Varela E, Garrido-Martin EM, et al. TGF-beta regulates the expression of transcription factor KLF6 and its splice variants and promotes co-operative transactivation of common target genes through a Smad3-Sp1-KLF6 interaction. *Biochem J* 2009;419:485–95. <https://doi.org/10.1042/BJ20081434>.
- [56] He M, Zheng B, Zhang Y, Zhang X-H, Wang C, Yang Z, et al. KLF4 mediates the link between TGF-β1-induced gene transcription and H3 acetylation in vascular smooth muscle cells. *FASEB J* 2015;29:4059–70. <https://doi.org/10.1096/fj.15-272658>.
- [57] Yan X, Xiong X, Chen Y-G. Feedback regulation of TGF-β signaling. *Acta Biochim Biophys Sin* 2018;50:37–50. <https://doi.org/10.1093/abbs/gmx129>.
- [58] Zhao M, Pery JM, Marshall H, Venkatraman A, Qian P, He XC, et al. Megakaryocytes maintain homeostatic quiescence and promote post-injury regeneration of hematopoietic stem cells. *Nat Med* 2014;20:1321–6. <https://doi.org/10.1038/nm.3706>.
- [59] Gong Y, Zhao M, Yang W, Gao A, Yin X, Hu L, et al. Megakaryocyte-derived excessive transforming growth factor β1 inhibits proliferation of normal hematopoietic stem cells in acute myeloid leukemia. *Exp Hematol* 2018;60:40–46.e2. <https://doi.org/10.1016/j.exphem.2017.12.010>.
- [60] SanMiguel JM, Young K, Trowbridge JJ. Hand in hand: intrinsic and extrinsic drivers of aging and clonal hematopoiesis. *Exp Hematol* 2020;91:1–9. <https://doi.org/10.1016/j.exphem.2020.09.197>.
- [61] Kramer A, Challen GA. The epigenetic basis of hematopoietic stem cell aging. *Semin Hematol* 2017;54:19–24. <https://doi.org/10.1053/j.seminhematol.2016.10.006>.
- [62] Sun D, Luo M, Jeong M, Rodriguez B, Xia Z, Hannah R, et al. Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal. *Cell Stem Cell* 2014;14:673–88. <https://doi.org/10.1016/j.stem.2014.03.002>.
- [63] Wouters BJ, Jordà MA, Keeshan K, Louwers I, Eerpelinc-Verschueren CAJ, Tielemans D, et al. Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced CEBPA and mutations in NOTCH1. *Blood* 2007;110:3706–14. <https://doi.org/10.1182/blood-2007-02-073486>.
- [64] Stoll G, Viara E, Barillot E, Calzone L. Continuous time boolean modeling for biological signaling: application of Gillespie algorithm. *BMC Syst Biol* 2012;6:116. <https://doi.org/10.1186/1752-0509-6-116>.