Single-cell RNA-seq assisted synthesis of a Boolean network to model early hematopoiesis aging

Léonard Hérault[1,2], Mathilde Poplineau[2], Estelle Duprez[2*#] and Élisabeth Remy[1*#]

1. Aix Marseille Université, CNRS, Centrale Marseille, I2M, Marseille, France
2. Epigenetic Factors in Normal and Malignant Hematopoiesis Team, Aix Marseille Université, CNRS, INSERM, Institut Paoli-Calmettes, CRCM, Marseille, France

* These authors contributed equally: Elisabeth Remy, Estelle Duprez
#Corresponding authors.

Abstract

We previously analyzed 15 000 transcriptomes of mouse hematopoietic stem and progenitor cells (HSPCs) from young and aged mice and characterized the early differentiation of the hematopoietic stem cells (HSCs) according to age, thanks to cell clustering and pseudotime analysis [1]. In this study, we propose an original strategy to build a Boolean gene network explaining HSC priming and homeostasis based on our previous single cell data analysis and the actual knowledge of these biological processes **(graphical abstract)**.

We first made an exhaustive analysis of the transcriptional network on selected HSPC states in the differentiation trajectory of HSCs by identifying regulons, modules formed by a transcription factor (TFs) and its targets, from the scRNA-seq data., From this global view of transcriptional regulation in early hematopoiesis, we chose to focus on 15 components, 13 selected TFs (Tal1, Fli1, Gata2, Gata1, Zfpm1, Egr1, Junb, Ikzf1, Myc, Cebpa, Bclaf1, Klf1, Spi1) and two complexes regulating the ability of HSC to cycle (CDK4/6 - Cyclines D and CIP/KIP). We then defined the relations in the differentiation dynamics we want to model ((non) reachability, attractors) between the HSPC states that are partial observations of

binarized activity levels of the 15 components. Besides, we defined an influence graph of possibly involved TF interactions in the dynamic using regulon analysis on our single cell data and interactions from the literature. Next, using Answer Set Programming (ASP) and considering these inputs, we obtained a Boolean model as a final solution of a Boolean satisfiability problem. Finally, we perturbed the model according to aging differences underlined from our regulon analysis. This led us to propose new regulatory mechanisms at the origin of the differentiation bias of aged HSCs, explaining the decrease in the transcriptional priming of HSCs toward all mature cell types except megakaryocytes.

Introduction

Hematopoiesis is the process of cell differentiation that allows the hematopoietic stem cell (HSC) to produce all types of mature blood cells, each with its own function. 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 to instruct the differentiation of the hematopoietic stem and progenitor cell (HSPC) compartment (reviewed in [2]). It is also well known that deregulations of transcriptional programs underlie the decline in HSC function during aging [3]. This leads to an alteration of the HSC pool phenotype, resulting in an increase in myeloid and megakaryocytic cells at the expense of lymphoid and erythroid ones in aged individuals. As a consequence, eldery people are subject to blood disorders such as anemia and acute myeloid leukemia [4]. Given the current aging population, deciphering the molecular mechanisms and specifically the gene regulatory network (GRN) underlying age-induced deregulation of HSCs is of great interest and is currently the subject of extensive research, particularly with respect to the early events of HSC commitment called early hematopoiesis. 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 [5,6]. Comparison of young and aged HSPCs using single-cell RNA-seq (scRNA-seq) analysis helped to accurately map lineage priming and cell cycle changes in aged mice [1,7,8], leading to the identification of groups of HSPCs that are distinct in their potential to maintain early hematopoiesis and whose proportions are altered during aging. 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 [9,10]. Some inference methods,

based on mutual information or regression trees, have been successfully used to analyze regulatory networks in the HSC microenvironment [11]. They have also permitted the identification of regulons, modules formed by a transcription factor and its targets, in the HSPC compartment during human ontogeny [12] or mouse aging [1]. However, these studies offer only a static view of the GRN governing the biological process, so the precise molecular mechanisms and interactions involved in cellular decisions such as differentiation to a particular lineage are not resolved. To address these issues, it is relevant to study the dynamics of networks using Boolean network (BN) modeling. BN modeling approach provides a good abstraction of the long-term behaviors of a biological system, although continuous changes in component activity and the 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 [13]. 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 [14–16].

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 when ordered along a pseudo-trajectory and after binarization of the component activities of interest (usually gene expression), can be interpreted as an observation of a trajectory generated by a 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 [16,17]. More recently a method, called Bonesis, has been developed to infer, from a gene regulatory network, a BN satisfying dynamic constraints between cell states [18]. 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 scRNA-seq analysis, we and others recently observed new priming events in the HSC pool [1,5] which are deregulated with aging of the BM. Here, we wanted to take advantage of our scRNA-seq data to construct a BN to understand the early priming of HSCs and to precisely characterize molecular mechanisms leading to HSC aging. We first defined key HSPC states of early hematopoiesis based on our previous scRNA-seq analysis of young and aged mouse HSPCs [1]. Next, we characterized these states by an exhaustive analysis of the transcriptional activity in each cell composing a given state using the single-cell regulatory network inference and clustering (SCENIC) method [19]. Then, based on these characterisations and taking into account the current knowledge on the regulation of HSC

fate, we adapted an existing GRN of early myelopoiesis [20]. Then, we used Bonesis to infer on the new GRN a BN of early hematopoiesis whose dynamic fits with our pseudo-trajectory of HSC priming. Finally, we performed and analyzed perturbations of this BN to propose key factors and mechanisms behind the HSC differentiation bias observed during aging. Altogether, our results provide a mathematical model of the early hematopoiesis process which allows to assess the HSPC changes under physiological aging.

Methods

scRNA-seq dataset

We used the scRNA-seq dataset presented in our previous study available in the Gene Expression Omnibus database under accession code GSE147729 [1]. This dataset is composed of two pools of young (2/3 months) mice HSPCs and two pools of aged (18 months) mice HSPCs. Our previous results (cell cycle phase assignment, cell clustering, pseudotime ordering) were considered in this study to define the HSPC states at the basis of our modeling [1] (**Supplementary Table 1**).

Regulon analysis

Identification of regulons with pySCENIC. We used the Single-Cell Regulatory Network Inference and Clustering (SCENIC) approach [19] to identify regulons, 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 [21] 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 [1]. In this study we used the whole 1721 TFs with an available motif in the motif database as input. 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_t is defined as follows:

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

where $IS(r_t,g)$ is the interaction score defined as the product of the number of SCENIC runs in which the interaction from r_t 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 aucell 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 [22], 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 same sign and above 0.001 in the two batches presenting a combined p value < 0.001 were kept (**supplementary Table 4)**

Regulon 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 [21]. The cluster_louvain function, from igraph R package [23] was used to find TF communities in the undirected transformation of this network with edges weighted by the NIS scores. The Cytoscape software [24] was used to visualize the results from graph clustering.

Cistrome database analysis

Available mouse TF ChiP-seq experiments annotated for bone marrow tissue in the Cistrome database were analyzed using Cistrome database workflow [25]. 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 [26]. 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 follow:

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

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 & 5**).

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 the edges representing activations and inhibitions. 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 systems (succession of consecutive configurations). Here, we used the Most Permissive (MP) semantics that is required to use the Bonesis tool). This recently proposed semantic takes into account additional transient states reflecting increasing (↗) or decreasing (↘) dynamical states. A component in (↗) or (↘) 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 [27].

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 respectively a loss of function or deletion (denoted KO, Knocked Out), by maintaining its variable at 1, respectively at 0.

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.

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:

If the node of the network represents a TF heading a regulon with more than 10 targets, we considered AUCell scores of all cells of the HSPC state and used a Kmeans clustering (with K=2) to decide whether the node was active (1) or inactive (0). Otherwise, because the AUCell scores are less reliable, we discretized the activity on 3 levels, active (1), inactive (0) or free/unknown (*) using Kmeans clustering (K=3) on averaged RNA levels per HSPC states.

For the cell cycle complexes (CDK4/6CycD and CIP/KIP complexes) we took the discretization of the RNA levels genes coding for their component and attributed them a value from -1 to 1 (-1 inactive, 0 free/unknown, 1 active). Then we considered the sum of gene values for each complex: above 1 active complex, below -1 inactive complex, between -1 and 1 free/unknown complex activity).

In order to be less constraining, we relaxed some constraints on the component activities, by replacing their discretized value 0 or 1 to a free (*) status.

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 the Boolean satisfiability problem and enumerates all the possible Boolean models that satisfy the constraints in the Most Permissive (MP) semantic [18].

We chose to limit the number of clauses per logical rule to 3. The solver Clingo [28] was used in the inference steps.

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 previously reported in [29]. For each of them in silico KO perturbation on each source node 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 led to the addition of 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 number and then a minimization in the other interaction numbers in the inferred models.

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 Most Permissive (MP) semantics with the mpbn python package [27].

Code availability

All R, python, and ASP codes used in this study are integrated in a global snakemake workflow available at: https://github.com/leonardHerault/scRNA_infer.git.

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 primed HSPC state and in non-primed clusters gathered, the enrichment of age was tested using a hypergeometric test (phyper R function).

Results

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

In order to define references for establishing dynamical constraints for the inference method of BN, we defined HSPC states, mixing three layers of information: cell cluster identity, pseudotime trajectory and cell cycle phases, retrieved from our previous scRNAseq analyses [1]. We chose to take into account these different parameters considering that each of them is importantly linked to HSPC functionality. The clusters provide a meaningful functional partition of the HSPCs, the shape of the trajectory reflects well the priming of HSPCs toward different lineages, while cell cycle status is important to distinguish the dividing HSPCs.

Hence, we defined nine states that were visualized on the pseudotime trajectory (**Figure 1A**). We considered two HSPC states at the beginning of the pseudotime trajectory (pseudotime <2); one was composed of non-cycling cells 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 ifnHSC state gathering cells of the ifn cluster (interferon response signature), a state gathering the 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 most of the short-term hematopoietic stem cells (STHSC) and spreading on the terminal branches of the trajectory. 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 (Herault et al, 2021): 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 nine defined states resume several initial (iHSC, srHSC), terminal (pEr, pMk, pNeuMast, pLymph) and branching (preDiff) points of the pseudo-trajectory and gather 40% of the cells of the scRNA-seq dataset **(Supplementary Table 1**). Thus, we provide a detailed 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 [21]. We characterized 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 [19] and performing a hierarchical clustering, we revealed a specific regulon activity profile for each of the HSPC states (**Figure 1B**). Regulon activities supported the TF 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 ifnHSC; 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 feature [1].

Next, to connect the regulons to each other, we built a transcriptional network whose nodes are TFs at the head of the 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) (**Figure 1C, Supplementary Table 2**). We further confirm these regulatory interactions by analyzing the presence of the TFs in the regulatory regions of their targets using ChiP-seq data from the Cistrome database (Liu et al., 2011), which covered 33 % of the TFs of the transcriptional network in cell populations close to ours. 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 3)**. We then performed a clustering analysis by weighting the network using a normalized interaction score (NIS) calculated from the SCENIC results and applying Louvain clustering. We underlined 10 regulon communities and three isolated regulons (Zbtb7b, Brf2, Sp4). By associating each TF in the network to the HSPC state that its regulon most characterizes, we observed that half of the communities regroups TFs whose activity characterizes the same HSPC state **(Figure 1C and 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 were markers of the qHSC state, whereas the C2 community contained mainly TFs leading regulon markers of ifnHSC 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 we defined and characterized with specific transcriptional activity. This also highlighted that TFs, regulators of these activities, interact with each other in a structured network that supports the differentiation pathway of HSCs: from the i-srHSC states to one of the four lineage primed HSC states, through the transient HSC states (ifn-qHSC,preDiff).

Inference of a gene 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 the use of Bonesis, a recently proposed approach for Boolean network inference [18], which needs the synthesis of an influence graph and 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 [20]. This model provided megakaryocyte and erythrocyte stable states and a granulo/monocyte branching state that corresponded quite well to our defined states. Yet, we had to adapt it to encompass the priming of lymphoid HSCs and cell cycle regulation that are likely to be involved in early HSC commitment and aging [1,7]. We extracted a subgraph from the Krumsiek model 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 Zfpm1, the cofactor of Gata1, which was expressed in pEr and pMk HSPC states (**Supplementary Figure 1**). To represent the lymphoid priming, we added Ikzf1 whose regulon marked pLymph state according to our analysis (**Supplementary Table 2)**, in agreement with prior knowledge of early lymphoid specification in HSPC [30].

We also added two components that regulate the ability of HSC to cycle: the CDK4/6-Cyclines D (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) marking the quiescence of the HSCs [31]. To connect CIP/KIP complex to the network we added Junb and Egr1, two factors involved in HSC quiescence [32,33], which we identified in our regulon analysis as markers of qHSC state and activators of CIP/KIP genes (**Figure 1B; Supplementary Tables 2 & 3**). Finally, to connect CDK4/6CycD to the network, we added Myc and Bclaf1, two factors involved in HSC cell cycle [34,35]. Both were active regulons in the preDiff state whereas only Bclaf1 regulon was active in srHSC state with CDK4/6CycD complex genes in its targets (**Figure 1B; Supplementary Tables 2 & 3**). To connect all

these nodes, we considered interactions from Krumsiek's model that we complemented with interactions we identified with SCENIC analysis and in the literature.

Finally, we obtained an influence graph with 15 components and 60 interactions (**Figure 2Ai**), more than 75% of which were confirmed by at least two of the following information sources; SCENIC, literature or Cistrome (**Supplementary Figure 2A; Supplementary Table 5A, B & C**).

For the discretization of the data, we associated to each HSPC state a meta-configuration, *i.e.,* a vector representing the discretized activity level of each of the 15 components (see Methods). 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 CIPKIP 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 [1]. Thus, we let Gata2 free in pNeuMast HSCP states. Finally, we also let free Egr1 in srHSC in agreement with a previous study suggesting its role in HSC maintenance in the niche [32].

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 and we 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 (self-renewal) or qHSC (quiescence) state as suggested by the literature [36,37]. 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 fixpoints pNeuMast, pER, and pMk were accessible. We allowed a cell from iHSC to directly reach the pLymph fixpoint based on the shape of the pseudo-trajectory and the high hscScore of the pLymph cells [1]. All these constraints are resumed in the HSC differentiation journey **(Figure 2Aiii)**.

With this first inference, Bonesis provided a set of more than 100,000 solutions. To reduce the set of solutions, we therefore developed a strategy to refine the solution search and obtain a final solution (**Figure 2B**). We considered possible mutant behaviors of some of our solutions that match previously described biological phenotypes of the mutants reported in

the literature (see methods and **Supplementary Table 6** for the references): the Ikzf1 KO conducting to an absence of pLymph fixpoint , the Spi1 KO to absences of both pNeuMast and pLymph fixpoints, the Klf1 KO to an absence of pEr fixpoints and the Junb KO to an apparition of an additional proliferative (active CDK4/6CycD complex) pNeuMast fixpoint. We also considered KI perturbations on nodes Egr1 and Junb, as these components were previously found upregulated in HSC upon aging [7]. 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 [1]. These 6 altered behaviors are resumed in **Supplementary Figure 3** and were added to the Bonesis constraint set (**Figure 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 (**Figure 2Bii)**, of which more than 80% were supported by at least two sources of information among SCENIC, Cistrome and literature (**Supplementary Figure 2B)**. 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 [38]. 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 (**Figure 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 Zfpm1 are present in both clauses. Finally, we obtained the final Boolean network presented in **(Figure 2C & D)**.

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

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

Simulations of the model were done within the MP semantic, they displayed 5 fixed points (whose complete descriptions are given **Figure 3A**). We observed that node Tal1 was active

in the fixed point pEr (while its value was left free in the fixed point constraint), as it is in the erythroid fixed point in the Krumsiek model [20]. According to our requests, all fixed points were reachable from iHSC, regardless of the initial value of the Zfpm1 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 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 [39].

We conducted a fine analysis of the transient dynamics to highlight events that are responsible for some salient dynamical properties along the trajectory. We observed that Gata2 is 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, which is marked by the activity of Ikzf1 whose only regulator is the activator Gata2 (**Figure 3B**). We highlighted a transient pME configuration, described in **Figure 3A,** that can reach pEr and pMk configurations but not pNeuMast configuration thanks to a fine analysis of paths in the trajectory space. We observed that the pME configuration is reachable from the preDiff state, when Junb is activated and Spi1 is inactivated (**Figure 3B**).

Interestingly, the choice between pMk and pEr fixed points relied on the Fli1-Gata1-Klf1 module (**Figure 3C**), and on a transient state of Fli1 catched thanks to the MP semantics. Indeed, starting from the branching point pME in which the three components of the module are absent, Fli1 activity can increase (thanks to the presence of Junb) allowing *Gata1* to be also activated and lead to the stable configuration pMk. Moreover, as long as Fli1 has not reached its activity level allowing it to inhibit *Klf1*, activation of *Klf1* by Gata1 can occur and lead to pEr configuration (**Figure 3B**). Note that we are 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* functions as a switch to maintain the differentiation between pMk and pEr (**Figure 3C**).

Furthermore, our model presents a proliferation configuration, in which CDK4/6CycD and Myc are active and CIP/KIP inactive. This configuration is accessible from the iHSC state,

and all fixed points can be reached from it. This is in agreement with our previous observation of an increase of HSC proliferation during the priming toward the different lineages [1].

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 state depends on the fine tuning of Fli1.

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 animals at the expense of the priming toward pLymph, pEr and pNeuMast [1]. To decipher molecular mechanisms and factors at the origin of this alteration, we simulated perturbations in the inferred Boolean network according to alterations observed in the transcriptome of aged HSPCs.

Alterations of regulon activity due 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, srHSC states with similar amounts of decrease and increase in activity (**Supplementary Figure 4**), and very few in pNeuMast and pEr. Almost all activity alterations of regulons were found in more than one state. 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 Figure 5 & 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 Table 5B**). More precisely we found Junb, Egr1 and Fli1 (resp. Spi1) activities significantly increased (resp. decreased) in more than a half of the 8 HSPC states considered for the model inference (**Figure 4A**).

To identify possible altered TF regulations, we compared for each regulation the normalized interaction score (NIS) of young and aged cells analyzed separately using SCENIC workflow

and computed a score difference between young and aged conditions (**Supplementary Table 4 and Supplementary Figure 6**). 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 Figure 7**). 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 an increase of activation by Spi1 (NIS increase by 0.2) upon aging (**Figure 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 (**Figure 4C**). The simulations of these 5 perturbations led to the loss of reachability of the fixed points pLymph and pNeuMast from the i-sr-qHSC configurations, only the simulation of the 3 KIs led to the loss of reachability of pEr, which makes pMK the unique reachable fixed point (**Table 1**). This was expected for Egr1, Junb KIs and Spi1 KO as we have imposed these behaviors in the dynamical constraints regarding previous experimental studies (**Supplementary Table 6**). Note that, still for these two perturbations, the fixed point pMk is quiescent (CIP/KIP active). These results agree with our single cell analysis at the population scale, 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 (**Figure 4Di**).

We also observed that preDiff was no longer accessible from i-sr-qHSC configurations with Junb KI disruption 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 (**Figure 4D**) and are 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 **Figure 1A**).

Although the topology of the gene regulatory network consists in a unique strongly connected component and 3 output nodes, we distinguished two functional modules, one regulating the cell cycle complexes (CDK4/6CycD and CIP/KIP) constituted of nodes Bclaf1, Myc, Junb, and a second one constituted of seven TFs governing HSC fate, with nodes Cebpa, Fli1 and Egr1 making the connections between them (**Figure 4C**). The model perturbations affected these modules, in particular the functional circuits that control them (Egr1-Junb, Fli1-Gata1, and the self-regulation on Spi1). The perturbation *Egr1* KI implies

an activation of *Junb* which in turn activates *Fli1*. Thus, the cross-inhibitory circuit (Fli1-Klf1) that ensures the differentiation pMk/pEr is affected in each of these perturbations with a loss of pEr. Besides, the KO of *Cebpa* activation by Gata2 prevents Spi1 from becoming active from any of the i-sr-qHSC configurations. Thus, we highlighted the major roles of *Egr1* overexpression and loss of *Cebpa* activation by Gata2 in early hematopoiesis aging. On another note, regarding the global TF network from scenic (**supplementary Table 3**) *Junb* and *Egr1* upregulation with aging could be mediated by Klf factors such as Klf2-4-6 also upregulated with aging (**supplementary Table 4**).

Therefore, our model perturbation analysis of aging of early hematopoiesis highlights *Egr1* and *Junb* upregulation and loss of *Cebpa* activation by Gata2 alterations as two major molecular mechanisms that lead to HSC aging resulting in the decrease in all lineage priming except the megakaryocyte one.

Discussion

The mechanisms governing the balance between self-renewal and differentiation of HSCs are the guarantee of functional hematopoiesis. Several BNs have been proposed to understand the key regulatory elements of HSC commitment to lymphoid or myeloid progenitors [14–16]. However, none of them fully addressed the early priming of the HSCs, which has been recently emphasized by the development of scRNAseq analyses [1,5]. Previous approaches to inferring BN from scRNA-seq data have been successfully applied to hematopoiesis [16,17]. However, results from these works, relying solely on the data, may be biased by the imprecision of the pseudotime values and the discretization of the data they require. Here we developed a new modeling strategy using scRNAseq to highlight key regulatory points of early HSC differentiation that are dysregulated in aged HSCs. The originality of our modeling approach lies in taking full advantage of the scRNA-seq data by incorporating the knowledge available in the rich literature and databases to compensate for the "grey areas" of scRNA-seq [40]. The synthesis of our BN was assisted by constant cross-referencing of data from the literature and our scRNA-seq-deduced observations. On one hand, we inferred new transcriptional regulations between the components that we selected using regression trees (added to the literature regulations). On the other hand, we defined dynamic constraints ((non)accessibility, stable states) between the meta-configurations, defined from our observations of key process states. We then implemented a BN inference strategy using Bonesis to select an early hematopoiesis BN as consistent as possible with

prior knowledge of the behaviors and regulations of experimentally characterized mutants**.** Studying the dynamics of our BN solution in MP semantics allowed us to trace a succession of events leading to HSC priming to the different lineages. We showed that from the iHSC, activation of Ikzf1 by Gata2 stabilizes early lymphoid priming of HSC. From a committed state with Spi1 and Myc active and Gata2 inactive HSC can prime for the neutrophil/mastocytic lineage or activation of Fli1 regulates the choice of priming towards the erythroid or megakaryocytic lineage. Our model also shows that the pre-Diff committed state is not the unique path for HSC priming to differentiation pathways. An alternative trajectory is possible, for example, with early activation of Gata1 by Gata2 leading directly to the primed megakaryocyte or erythrocyte state. Thus, our model is in agreement with previous lineage tracing studies highlighting the coexistence of multiple hematopoietic hierarchies [41,42]. However, we suggest that under normal conditions, HSC initiates mainly following the trajectory observed in our scRNA-seq data on which the dynamic constraints satisfied by our model have been defined.

Our model correctly reproduces behaviors of mutants observed in vivo/in vitro for a majority of the TFs in the network. This is not the case for the Myc KO for which we observed no difference in the accessibility of the primed states in silico, whereas an increase in HSCs and a decrease in its differentiation have been reported experimentally with this mutation [35]. According to this study, this is due to intercellular interactions not taken into account in our model. We also did not observe any changes in the dynamics for the Egr1 KO mutant although, similarly, a previous study showed a decrease in HSC priming along with an increase in HSC [32]. These observations could correspond to a transient accumulation before a delayed priming that our modeling formalism cannot capture. In the condition perturbed by aging our study suggests an alteration of the accessibility of the primed states. Our model perturbations based on alterations in TF activity (over-activation of Egr1 and Junb) and regulation (loss of Cebpa activation by Gata2) observed with aging in our scRNA-seq data, as a whole result in the loss of priming to all lineages except the megakaryocytic lineage. These observations are in agreement with our scRNA-seq data, which show a decrease in the proportion of aged cells in all primed state clusters except for the cell cluster primed for megakaryopoiesis. Our model is thus able to describe in part the mechanisms of immunosenescence occurring during early hematopoiesis with a loss of lymphoid potential and a myeloid bias [43].

Our results highlight the role of overactivated Egr1 and Junb factors in quiescent myeloid-biased HSCs that accumulate with aging [7] and have previously been identified as factors in HSC quiescence [32,33]. Our model shows that these alterations impact a positive circuit between Egr1 and Junb that may be necessary for multiple priming to the different fixed points in the model. 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-beta signaling in other biological contexts [44–46]. We have seen that cells in the qHSC state have a strong TGF-beta signature to which a myeloid bias is added with aging (over-activation of Cebpe-b in particular). Moreover, it is known that megakaryocytes can promote HSC quiescence by producing TGF-beta [47,48]. Our results thus allow us to propose a self-activating loop of HSC aging with differentiation towards megakaryocytes that would promote a myeloid-biased quiescent state of HSC from which a single priming towards the megakaryocytic lineage would occasionally be possible. In agreement with this hypothesis, a direct differentiation of HSC into megakaryocytes was shown in a lineage tracing study [5]. This trajectory would thus be the one best preserved by the aged quiescent myeloid biased HSC. In parallel, the loss of Cebpa activation by Gata2 also participates in the loss of lymphoid and neutrophil/mastocyte priming. This alteration could have an epigenetic origin given the strong impact of ageing on histone marks regulating chromatin opening to the regulatory elements of HSC differentiation genes [49]. It could also originate from the emergence with age of a clone with a mutation at a Cebpa regulatory element recognised by Gata2.

To be noted that the selected solution and our simulations of aging perturbations are valid in MP semantics which allows many more transitions than the asynchronous and generalized semantics classically used. A comparison of the asynchronous and MP trajectories of our model would be interesting in order to identify the set of multivalued refinements necessary to reproduce in asynchronous semantics the MP trajectories. Here, we have done this work to explain the choice of differentiation towards the fixed points pEr and pMk from the preDiff state which is based on the existence of two thresholds of influence of Fli1 on these targets Klf1 and Gata1. Finally, one can notice that if our model describes the differentiation of a HSC, it could be used in a future work to model a population of HSCs using stochastic simulations. Such method could allow for an even finer description of the biological observations by proposing probability of stable states as output [50], which would correspond to the proportions of cells found in the primed clusters in the scRNA-seq data.

Figure & Table legends

Graphical abstract: From single cell-RNA seq data and current knowledge in early hematopoiesis (literature and biological database investigation), 3 inputs are obtained to define the network synthesis as a Boolean Satisfiability Problem depending on observations of states in the differentiation process: **1** influence graph of the possible component interactions, **2** discretized component activity levels in the considered states (blue: 0, inactive, white: *, unknown/free, red: 1, active). **3** dynamic relations ((non) reachability, attractors) between the considered states. Then, these inputs were encoded as constraints in Answer Set Programing (ASP) thanks to the Bonesis tool. After the solving, a final solution of a Boolean model of early hematopoiesis is obtained. This model is altered according to the characteristics of aging observed in our scRNA-seq data, in order to identify the main molecular actors and mechanisms of aging.

Figure 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 data [1]. On the right, the cells are ordered on the trajectory and are coloured according to their pseudotime value. The 5 branches of the trajectory are circled. Lower panel: the defined HSPC states are highlighted (labels) in the trajectory and cells coloured accordingly. **B** Heatmap of the average AUCell scores of the activity of regulons in each HSPC state. The scores were standardized and used to cluster regulons hierarchically. **C** Transcriptional regulation network of the regulon markers of the defined HSPC states. Regulons were clustered in 10 communities (from C1 to C10 plus 3 isolated nodes) with Louvain graph clustering. Node label color highlights the states where the regulon is the most active (same color code as in B). Red (grey) edges indicate that the transcriptional regulation is (is not) supported by peak analysis in Cistrome database with BETA tool. Edge thickness represents the normalized interaction score (NIS) obtained from SCENIC.

Figure 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** Discretization of component activities in the configurations. The activities of TFs at the head of a regulon with more than 10 targets are discretized with a Kmeans clustering of two on all cell regulon activity scores (the value of the cluster with the most cells is retained). Blue:

inactive; red: active. The activities of the other TFs (Tal1, Ikzf1, Zfpm1, and gene belonging to the two complexes CDK4/6CycD and CIPKIP) are discretized with a Kmeans clustering of three on averaged RNA levels in the corresponding states. Blue: inactive; white: unknown/free; red: active. For the CDK4/6CycD and CIPKIP complexes the final discretization is deduced from the sum of the discretization of each of its genes (see **supplementary Figure 2**). Red (resp. blue) hatched cases mark node activities freed from 1 (resp 0) to * in the final configuration settings compare to the discretized data in order to get some solutions. **iii** Graph representation of the dynamical constraints (edges) defined between the configurations (nodes). Arrows (crossed arrows) indicate reachability (resp. unreachability) between source and target configurations. Framed configurations are constrained as fixpoints. **B** Scheme of the Strategy to refine the solution search and obtain a final solution. **i** Obtention of a new influence graph after the addition of constraints resulting from mutant behaviours. 36 interactions are kept. For the updated constraints see **supplementary Figure 3A. ii** Pruning of the influence graph through maximization of high-confident interactions and minimization of others. **iii** A last inference step is applied to result in 616 possible solutions. **iv** A manual curation is necessary to obtain a final model. **C** Logical rules of the Boolean model of HSC priming. **D** Gene regulatory network of the inferred Boolean model. Rectangular nodes are cell cycle complexes and ellipse node TFs. Nodes are colored according to the HSPC states in which they are highly active according to our regulon analysis: Grey for qHSC, yellow for pL, orange for pNeuMast , blue for pMk and pEr, white for the nodes highly active in several HSPC states.

Figure 3: Analyses of the Boolean model evidence a sequence of transcriptional events to prime HSCs. A Table relating the HSPC and pME configurations identified by the model analysis (column: HSPC states, lines: components of the model). Colors represent the activation levels of the nodes (blue: inactive; red : active, white: free). **B** Graph representation of the dynamics between the configurations (nodes) from iHSC toward the different fixed points (framed nodes). Arrows (resp. crossed arrows) indicate reachability (resp. unreachability) between source and target configurations. Black edges are constrained dynamic properties whereas the red ones and pME configuration result from the dynamic study of the model. Zfpm1: * highlights the two possible values of this node in iHSC. Spi1: 1, Gata2: 0 indicate the irreversible inactivation of Gata2 by Spi1 observed in the preDiff non-return configuration. Junb: 1, Spi1: 0 indicate the necessary change in preDiff to reach the branching configuration pME. In MP semantics, from pME 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** Module of Gata1, Fli1 and Klf1 of the BN with cross inhibitory circuits between Klf1 and Fli1 maintaining HSC priming to pMk or pEr.

Figure 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), only young cells (orange) and only aged cells (purple). **C** Aging perturbations of the Boolean gene network. Rectangular nodes are cell cycle complexes and ellipse nodes TFs. Nodes are coloured according to the HSPC states in which they are highly active according to our single cell analysis: Grey 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 KO/KI perturbations. Crossed activation of *Cebpa* by Spi1 illustrates its edgetic mutation. **D** Reachability of HSCP states from any configuration in i-sr-qHSCs for WT and altered dynamic of the model. **(i)** WT case. Young (orange) and aged (purple) cell proportion is given below each HSPC state node. A star highlights a significant shift from expected proportion (hypergeometric test p value < 0.05). **(ii)** *Egr1* KI perturbation. **(iii)** *Junb* KI perturbation. **(iv)** *Cebpa* edgetic mutation. Colored arrows represent the remaining reachabilities.

Table 1: Aging perturbations of the early hematopoiesis Boolean model. This table summarizes the reachabilities of 5 HSPC-states (4 fixed points pLymph, pNeuMast, pER, pMk, and preDiff) from states i-sr-qHSC, in different simulations (a cross mean that the HSPC-state is reachable). The last column reports the observations in our data: Up (resp.down) arrows indicate an increase (resp a decrease) upon aging in component activity or interaction score revealed by our scRNA-seq analysis.

Supplementary figure legends

Supplementary Figure 1: Discretization of gene expressions for genes and regulons with less than 10 targets. Results of k-means clustering on averaged RNA levels of the selected HSPC states. Blue: inactivated; grey: unknown/free; red: activated.

Supplementary Figure 2: Venn diagrams of influence graph interaction sources. A Initial influence graph interactions retrieved from SCENIC results and/or literature investigation and supported or not by the Cistrome database analysis (see Supplementary Table 3). **B** Same diagram for interactions remaining after the influence graph pruning.

Supplementary Figure 3: Constraints and discretization used for the influence graph pruning and the final rule inference. A Dynamical constraint used for the update of the influence graph and the final rule inference. Black Arrows (resp. crossed arrows) indicate reachability (resp. unreachability) between source and target configuration. Framed configurations are constrained as fixpoints. Dashed line highlights the allowed reachability of a fixpoint with all node activities at 0 from iHSC**.** Red (crossed) arrow highlights the additional (non) reachable constraints of mutant behaviors. G0pMk is the only reachable fixpoint from iHSC in pEr/pMk KI (large blue arrow). 1: loss of pLymph reachability with *Ikzf1*/*Spi1* KO, 2: loss of pNeuMast reachability with *Spi1* KO, 3 loss of pEr reachability with *Klf1* KO, 4: additional pNeuMast cycling fixpoint with *Junb* KO, 5: a unique pMk quiescent fixpoint with *Junb*/*Egr1* KI **B** Discretization of component activities in the configurations used for the pruning of the influence graph and the final rule inference. Blue: 0, inactivated; white: *, unknown/free; red: 1, activated. G0pMk and G2MpNeuMast configuration were defined according to the first solution space exploration.

Supplementary Figure 4: HSPC state repartition of the regulon activity alterations upon aging. Up (Down) marks significant increases (decreases) in activity upon aging (average differences > 0.001, p-value < 10^{-3}). An alteration of a regulon activity can be recovered in several HSPC states.

Supplementary Figure 5: Heatmap of AUCell scores of regulons activity averaged by group of cells from the HSPC states in young and aged cells. Scores are standardized on aged (A) and young (Y) cells of the different states. Rows are ordered as in Figure 1.

Supplementary Figure 6: Retrieved scenic interaction from all cell analysis in aged or young only cell analysis. We did not consider interactions retrieved only in young or only in aged cells.

Supplementary Figure 7: Histogram of normalized interaction score differences with aging. Interactions non retrieved in SCENIC analysis of aged or young cells have a null difference. A difference of one points out that a regulation was retrieved only in young or only in aged cells analysis

Supplementary table legends

Supplementary Table 1: Definition of the HSCP states. Nine HPSC states were defined according to the results of cell clustering, cell cycle phase assignment and pseudo-trajetory analysis [1]. Cell number and cell proportion given the entire scRNA-seq dataset are given for each state.

Supplementary Table 2: Transcriptional network inferred with SCENIC. The table gives all the transcriptional interactions recovered in at least 80% (40) runs of SCENIC on all cell dataset from a TF head of a regulon toward a target gene with a mor (mode of regulation) of 1 for activation and -1 for inhibition. The recoveredTimes columns give the number of SCENIC runs in which the regulation is recovered. NIS: Normalized Interaction Score computed from importance score of SCENIC. NIS_diff: Normalized Interaction Score differences between NIS obtained from aged cell analysis versus NIS obtained from young cell analysis. NIS difference is 1 when the interaction is recovered in young (resp. aged) cell analysis and not in aged (resp. young) cell analysis. NIS is 0 when the interaction is recovered neither in young or aged cell analysis and only in all dataset analysis. Cistrome_BM column indicates if some ChIP-seq experiments in the Cistrome database for the TF head of regulon were available and analyzed (TRUE) or not (FALSE) enabling the computation of the Cistrome Regulatory Score (CRS) for the interaction.

Supplementary Table 3: Regulon activity markers of HSPC states. For each HSPC state, list of the regulons with an average AUCell score difference (avg_diff one state vs all others) > 0.001, a p value (p_val Wilcoxon rank sum test) and a p-adjusted value (p_val_adj Bonferonni correction) < 0.05. regulons were assigned to a community (C1 to C10) of the regulon network.

Supplementary Table 4: Regulon markers of aging in the different HSPC states.
In each HSPC state Wilcoxon Rank sum tests were performed on the AUCell activity scores between young versus aged cells in batch A and B separately. Only regulons with an activity in at least 10% of either young or aged cells of the state in both batches were tested. The

two p-values for each regulon were combined using the Tipett's method (minimum_p_pval column). In each cluster only regulon differences presenting the same variation and with an average score difference > 0.001 in the two batches were kept.

Supplementary Table 5: Influence graph interactions. A Interactions between the 15 components considered in the influence graph. List of the transcription factors (TFs) with their mode of regulation (mor: 1 for activation, -1 for inhibition) of a target. When available, references characterizing experimentally the interaction are given. In that case the interaction proof level can be a transcriptional regulation, a physical protein-protein interaction (physical interaction), or a functional interaction: Knock Down (KD), KO (Know Out), retrieved in the specified cell line (cell_line) and or cell type/tissue (cell_type_tissue). The interactions were reliably identified by SCENIC (present in more than 90% of the runs) analysis or not and when it was possible a Cistrome Regulatory Score (CRS) was computed. For cell cycle complexes (CIP/KIP, CDK4/6-CycD) the CRS is the sum of the CRS of each regulation of a considered TF toward one of the genes of the complex. A confidence level of A (high) or B (low) was given depending of references information and CRS and NIS (see **B**) score. After the pruning of low confidence level interactions 36 interactions remained in the solution (solution = TRUE). **B** SCENIC interactions considered for the influence graph. The table gives all the transcriptional interaction recovered in at least 90% (45) runs of SCENIC on all cell dataset from a TF head of a regulon toward a target gene with a mor (mode of regulation) of 1 for activation and -1 for inhibition. The recoveredTimes(_young/_aged) columns give the number of SCENIC runs on all dataset (young dataset/aged dataset) in which the regulation is recovered. NIS(_young/_aged): Normalized Interaction Score computed from importance score of SCENIC on all dataset (young dataset/aged dataset). NIS_diff = NIS_aged – NIS_young. NIS difference is set to 1 when the interaction is recovered in young (aged) cell analysis and not in aged (young) cell analysis. NIS is set to 0 when the interaction is recovered neither in young or aged cell analysis and only in all dataset analysis. Cistrome_BM column indicates if some ChIP-seq experiments in the Cistrome database for the TF head of regulon were available and analyzed (TRUE) or not (FALSE) enabling the computation of the CRS: Cistrome Regulatory Score for the interaction. After the pruning of low confidence level interactions 36 interactions remained in the solution (solution = TRUE). **C** List of all the references used to support some interactions.

Supplementary Table 6: Comparison of *in silico* KO perturbations in the final BN selected with previous in vivo/in vitro mutant studies.

For each altered gene (KO perturbation), the reachability of fixed points from iHSC in the perturbed BN was assessed. References presenting in vivo/in vitro related experiments are provided. Some additional fixed points compared to wild type conditions were found (column Remarks). In the last column precises if the perturbed behavior is constrained in the inference process (only the mutant behaviors observed in the 1000 selected BN solutions are constrained for the inference of the final solution).

Supplementary Table 7: Possible rules for CDK46CycD, Fli1, Gata1 and Gata2 after the final inference. The rules manually selected to select a final solution are in red. Of the 77 possible rules for Gata2 only the 7 ones with two clauses are presented.

REFERENCES

1. Hérault, L. *et al.* Single-cell RNA-seq reveals a concomitant delay in differentiation and cell cycle of aged hematopoietic stem cells. *BMC Biol* **19**, (2021).
2. Laiosa, C. V., Stadtfeld, M. & Graf, T. Determinants of lymphoid-myeloid lineage diversification. *Annu Rev Immunol* **24**, 705–738 (2006).
3. Svendsen, A. F. *et al.* A comprehensive transcriptome signature of murine hematopoietic stem cell aging. *Blood* (2021) doi:10.1182/blood.2020009729.
4. Rossi, D. J., Jamieson, C. H. M. & Weissman, I. L. Stems cells and the pathways to aging and cancer. *Cell* **132**, 681–696 (2008).
5. Rodriguez-Fraticelli, A. E. *et al.* Clonal analysis of lineage fate in native haematopoiesis. *Nature* **553**, 212–216 (2018).
6. Velten, L. *et al.* Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol* **19**, 271–281 (2017).
7. Kirschner, K. *et al.* Proliferation Drives Aging-Related Functional Decline in a Subpopulation of the Hematopoietic Stem Cell Compartment. *Cell Rep* **19**, 1503–1511 (2017).

8. Kowalczyk, M. S. *et al.* Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.* **25**, 1860–1872 (2015).
9. Hu, X., Hu, Y., Wu, F., Leung, R. W. T. & Qin, J. Integration of single-cell multi-omics for gene regulatory network inference. *Comput Struct Biotechnol J* **18**, 1925–1938 (2020).
10. Pratapa, A., Jalihal, A. P., Law, J. N., Bharadwaj, A. & Murali, T. M. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat Methods* **17**, 147–154 (2020).
11. Desterke, C. *et al.* Inferring Gene Networks in Bone Marrow Hematopoietic Stem Cell-Supporting Stromal Niche Populations. *iScience* **23**, (2020).
12. Roy, A. *et al.* Transitions in lineage specification and gene regulatory networks in hematopoietic stem/progenitor cells over human development. *Cell Reports* **36**, 109698 (2021).
13. Kauffman, S. A. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology* **22**, 437–467 (1969).
14. Bonzanni, N. *et al.* Hard-wired heterogeneity in blood stem cells revealed using a dynamic regulatory network model. *Bioinformatics* **29**, i80–i88 (2013).
15. Collombet, S. *et al.* Logical modeling of lymphoid and myeloid cell specification and transdifferentiation. *PNAS* **114**, 5792–5799 (2017).
16. Hamey, F. K. *et al.* Reconstructing blood stem cell regulatory network models from single-cell molecular profiles. *PNAS* **114**, 5822–5829 (2017).
17. Moignard, V. *et al.* Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat Biotechnol* **33**, 269–276 (2015).
18. Chevalier, S., Froidevaux, C., Paulevé, L. & Zinovyev, A. Synthesis of Boolean Networks from Biological Dynamical Constraints using Answer-Set Programming. in *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)* 34–41 (2019). doi:10.1109/ICTAI.2019.00014.

19. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083–1086 (2017).
20. Krumsiek, J., Marr, C., Schroeder, T. & Theis, F. J. Hierarchical Differentiation of Myeloid Progenitors Is Encoded in the Transcription Factor Network. *PLoS One* **6**, (2011).
21. Van de Sande, B. *et al.* A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nature Protocols* **15**, 2247–2276 (2020).
22. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
23. Csardi, G., Nepusz, T., & others. The igraph software package for complex network research. *InterJournal, complex systems* **1695**, 1–9 (2006).
24. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498–2504 (2003).
25. Liu, T. *et al.* Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol* **12**, R83 (2011).
26. Wang, S. *et al.* Target analysis by integration of transcriptome and ChIP-seq data with BETA. *Nat Protoc* **8**, 2502–2515 (2013).
27. Paulevé, L., Kolčák, J., Chatain, T. & Haar, S. Reconciling qualitative, abstract, and scalable modeling of biological networks. *Nature Communications* **11**, 4256 (2020).
28. Gebser, M., Kaminski, R., Kaufmann, B. & Schaub, T. Multi-shot ASP solving with clingo. *CoRR* **abs/1705.09811**, (2017).
29. Chevalier, S., Noël, V., Calzone, L., Zinovyev, A. & Paulevé, L. Synthesis and Simulation of Ensembles of Boolean Networks for Cell Fate Decision. in *18th International Conference on Computational Methods in Systems Biology (CMSB)* vol. 12314 193–209 (Springer, 2020).
30. Ng, S. Y.-M., Yoshida, T., Zhang, J. & Georgopoulos, K. Genome-wide lineage-specific transcriptional networks underscore Ikaros-dependent lymphoid priming in hematopoietic stem cells. *Immunity* **30**, 493–507 (2009).

31. Pietras, E. M., Warr, M. R. & Passegué, E. Cell cycle regulation in hematopoietic stem cells. *J Cell Biol* **195**, 709–720 (2011).

32. Min, I. M. *et al.* The transcription factor EGR1 controls both the proliferation and localization of hematopoietic stem cells. *Cell Stem Cell* **2**, 380–391 (2008).

33. Santaguida, M. *et al.* JunB protects against myeloid malignancies by limiting hematopoietic stem cell proliferation and differentiation without affecting self-renewal. *Cancer Cell* **15**, 341–352 (2009).

34. Dell'Aversana, C. *et al.* miR-194-5p/BCLAF1 deregulation in AML tumorigenesis. *Leukemia* **31**, 2315–2325 (2017).

35. Wilson, A. *et al.* c-Myc controls the balance between hematopoietic stem cell self-renewal and differentiation. *Genes Dev* **18**, 2747–2763 (2004).

36. Bernitz, J. M., Kim, H. S., MacArthur, B., Sieburg, H. & Moore, K. Hematopoietic Stem Cells Count and Remember Self-Renewal Divisions. *Cell* **167**, 1296-1309.e10 (2016).

37. Wilson, A. *et al.* Hematopoietic Stem Cells Reversibly Switch from Dormancy to Self-Renewal during Homeostasis and Repair. *Cell* **135**, 1118–1129 (2008).

38. Wiedemann, D. A computation of the eighth Dedekind number. *Order* **8**, 5–6 (1991).

39. Moussa, O. *et al.* Thrombocytopenia in mice lacking the carboxy-terminal regulatory domain of the Ets transcription factor Fli1. *Mol Cell Biol* **30**, 5194–5206 (2010).

40. Watcham, S., Kucinski, I. & Gottgens, B. New insights into hematopoietic differentiation landscapes from single-cell RNA sequencing. *Blood* **133**, 1415–1426 (2019).

41. Rodriguez-Fraticelli, A. E. *et al.* Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis. *Nature* **583**, 585–589 (2020).

42. Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D. & Klein, A. M. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* **367**, (2020).

43. 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* **208**, 2691–2703 (2011).

44. Botella, L. M. *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* **419**, 485–495 (2009).

45. He, M. *et al.* KLF4 mediates the link between TGF-β1-induced gene transcription and H3 acetylation in vascular smooth muscle cells. *The FASEB Journal* **29**, 4059–4070 (2015).

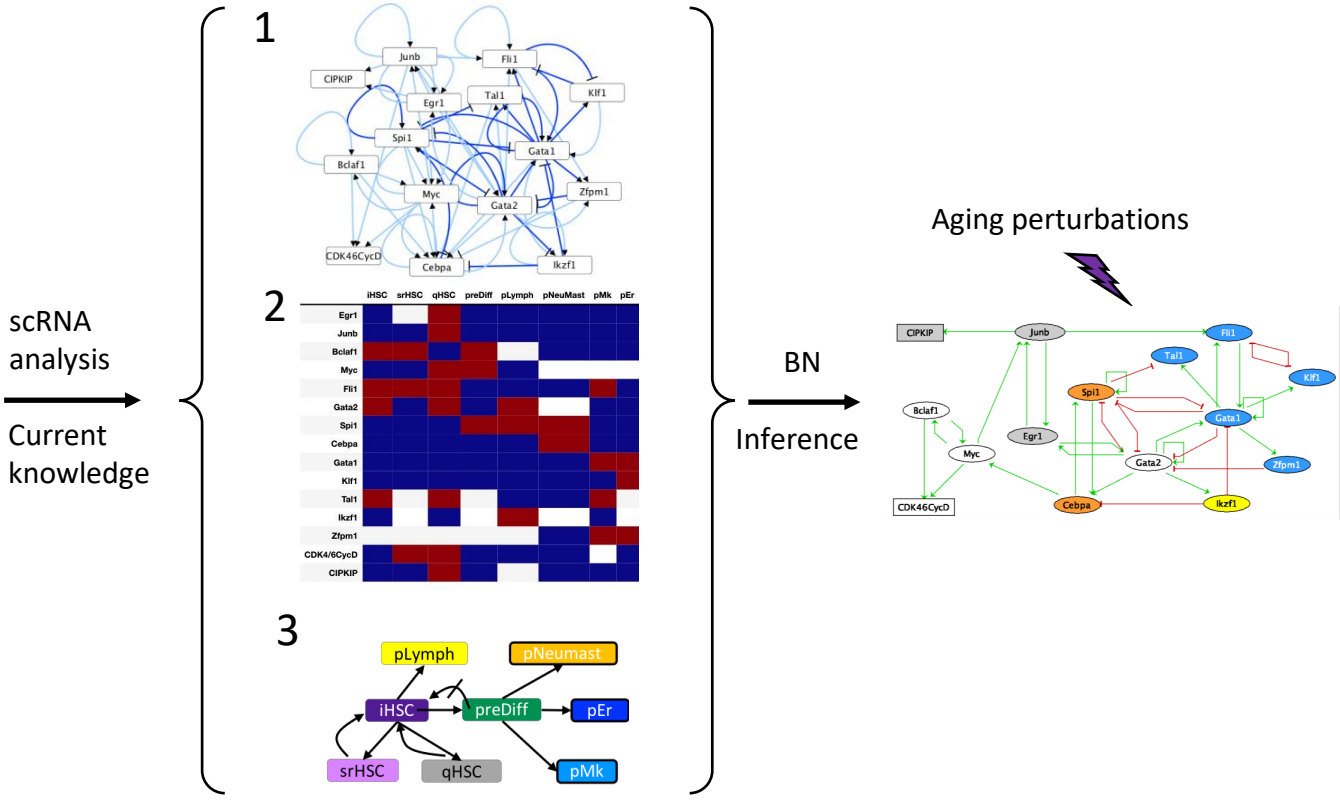
46. Yan, X., Xiong, X. & Chen, Y.-G. Feedback regulation of TGF-β signaling. *Acta Biochimica et Biophysica Sinica* **50**, 37–50 (2018).

47. Gong, Y. *et al.* Megakaryocyte-derived excessive transforming growth factor β1 inhibits proliferation of normal hematopoietic stem cells in acute myeloid leukemia. *Exp Hematol* **60**, 40-46.e2 (2018).

48. Zhao, M. *et al.* Megakaryocytes maintain homeostatic quiescence and promote post-injury regeneration of hematopoietic stem cells. *Nat Med* **20**, 1321–1326 (2014).

49. Sun, D. *et al.* Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal. *Cell Stem Cell* **14**, 673–688 (2014).

50. Stoll, G. *et al.* MaBoSS 2.0: an environment for stochastic Boolean modeling. *Bioinformatics* **33**, 2226–2228 (2017).



Graphical abstract: Overview of the scRNA-seq assisted gene Boolean network synthesis strategy.

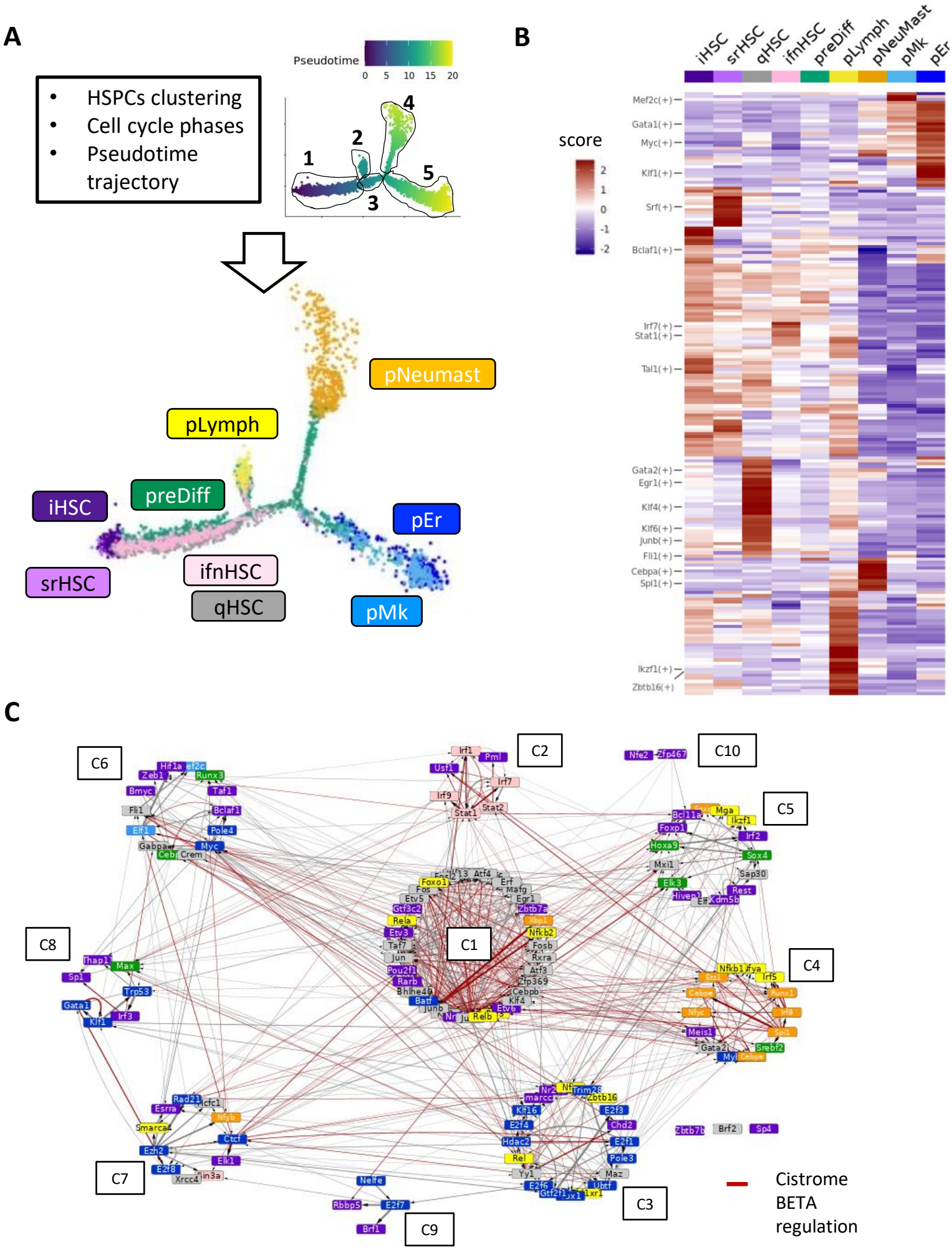
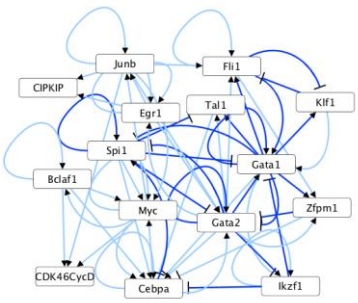


Figure 1: Regulon analysis identified distinct HSPC states with specific transcription factor activities and interactions.

Figure 2

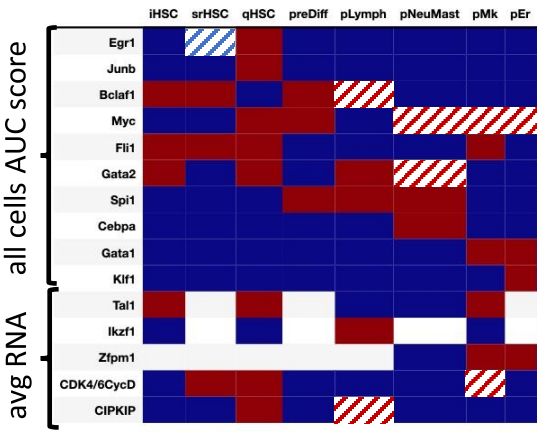
A

i

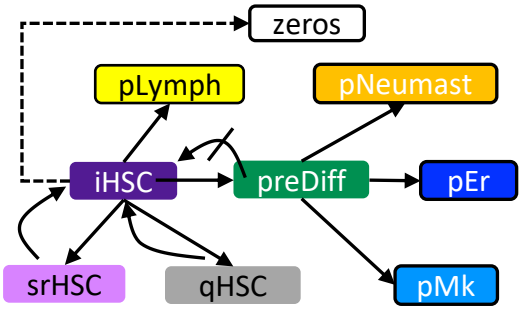


- 15 nodes
- 60 interactions
- >10^{22} possible models

ii



iii

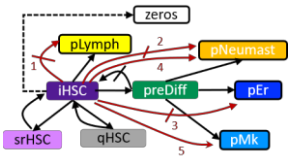


> 100 000 BN solutions

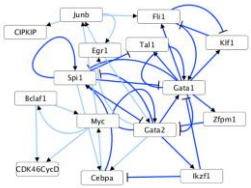
B

Perturbations (KO/KI) in 1000 solutions

i Adding constrained mutant behaviors



ii Influence graph pruning



iii Final inference

616 solution

iv Manual curation

Final solution

C

Component	Logical rules
$Egr1$	$Gata2 \wedge \text{Junb}$
$Junb$	$Egr1 \vee \text{Myc}$
$Bclaf1$	Myc
Myc	$Cebpa \wedge \text{Bclaf1}$
$Fli1$	$Junb \vee (\text{Gata1} \wedge \overline{Klf1})$
$Gata2$	$(Gata2 \wedge \overline{Gata1} \wedge \overline{Zfpm1}) \vee (\text{Egr1} \wedge \overline{Gata1} \wedge \overline{Zfpm1} \wedge \overline{Spi1})$
$Spi1$	$(Spi1 \wedge \overline{Gata1}) \vee (\text{Cebpa} \wedge \overline{Gata1} \wedge \overline{Gata2})$
$Cebpa$	$(Gata2 \wedge \overline{Ikzf1}) \vee (Spi1 \wedge \overline{Ikzf1})$
$Gata1$	$Fli1 \vee (Gata2 \wedge \overline{Spi1}) \vee (Gata1 \wedge \overline{Ikzf1} \wedge \overline{Spi1})$
$Klf1$	$Gata1 \wedge \overline{Fli1}$
$Tal1$	$Gata1 \wedge \overline{Spi1}$
$Ikzf1$	$Gata2$
$Zfpm1$	$Gata1$
$CDK46CycD$	$Bclaf1 \vee \text{Myc}$
$CIPKIP$	$Junb$

D

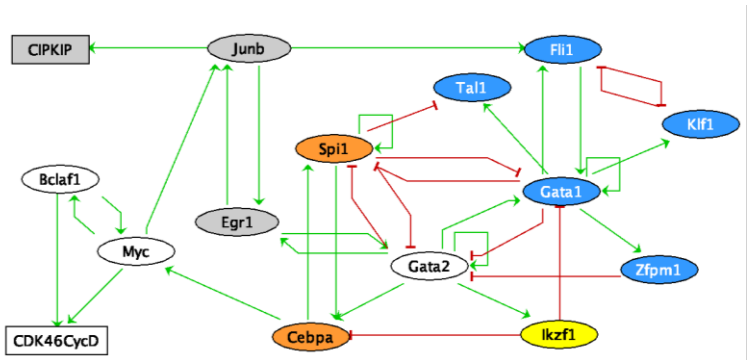


Figure 2: Inference of a gene Boolean network to model HSC priming.

Figure 3

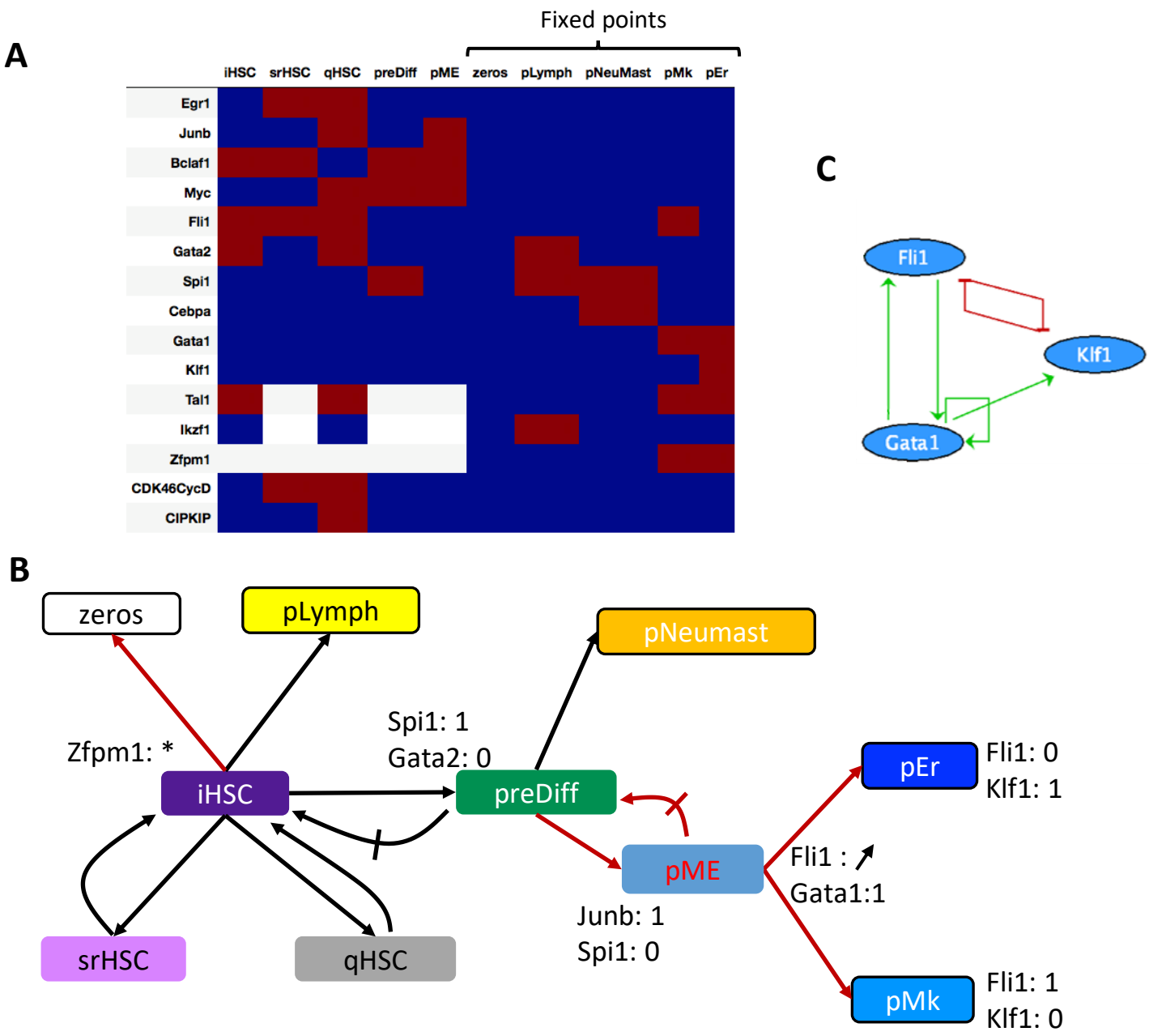


Figure 3: Analyses of the Boolean model evidence a sequence of transcriptional events to prime HSCs.

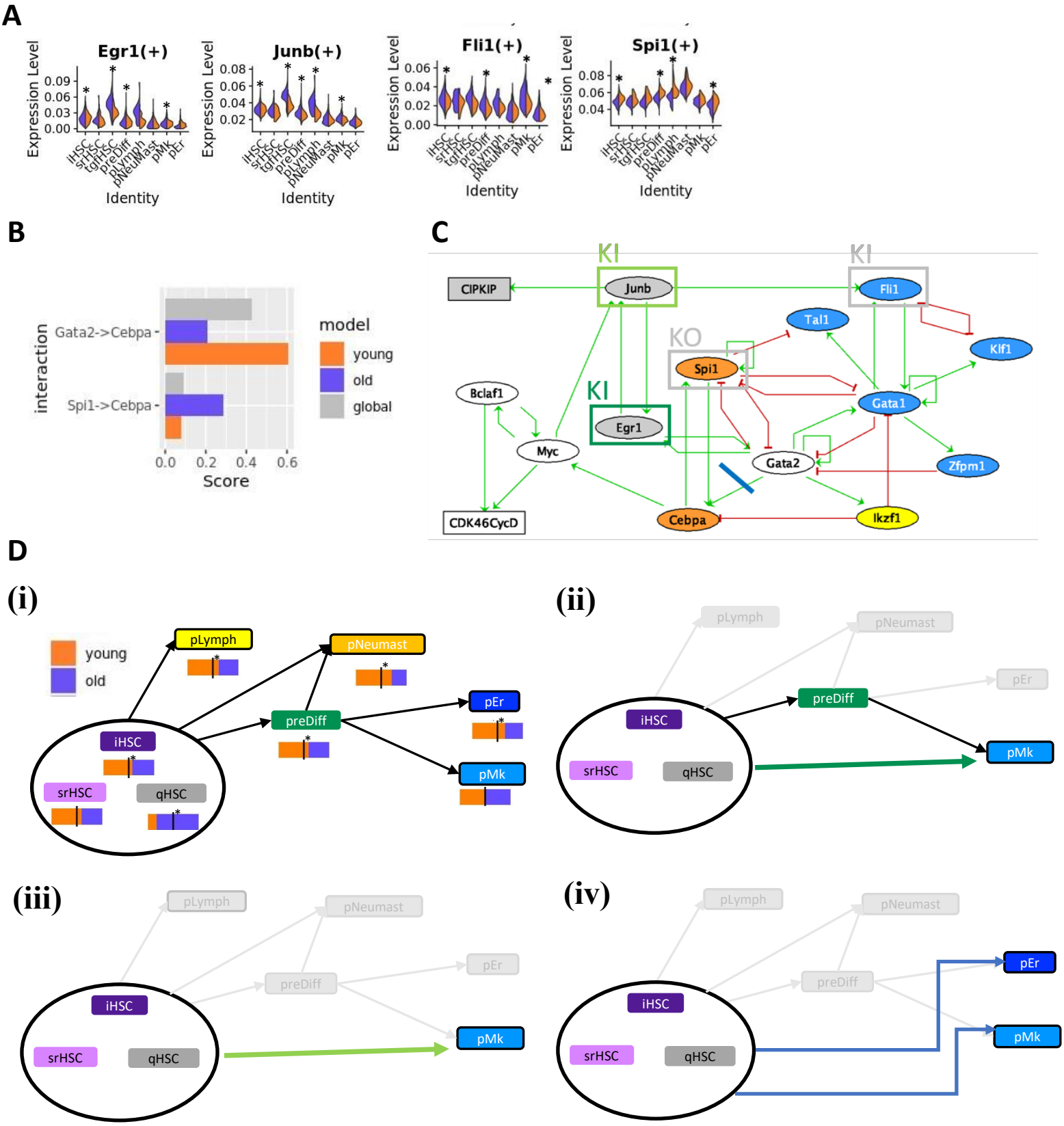
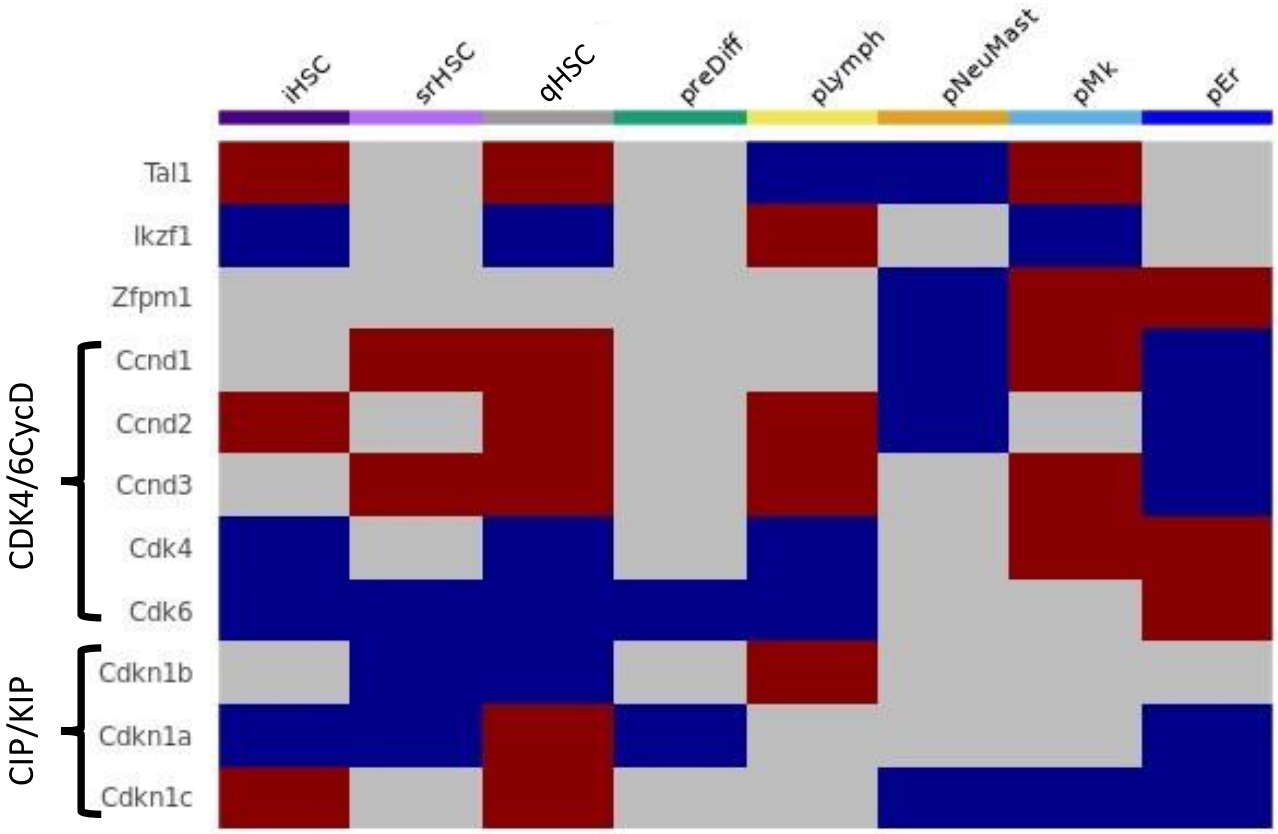


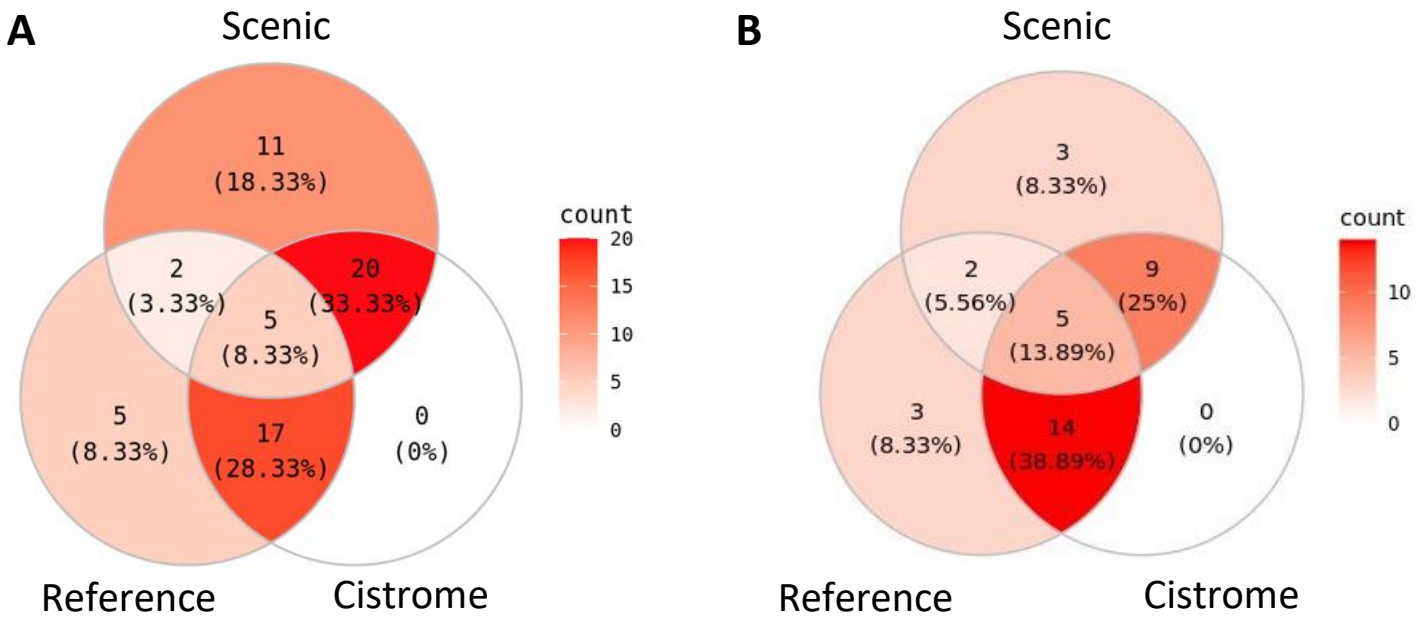
Figure 4: Perturbations of early hematopoiesis model explains some HSC aging features.

	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, pEr, pMk
Spi1 KO			X	X	X	Spi1 ↓ in iHSC, preDiff, pLymph, pEr
Edgetic Gata2-Cebpa			X	X		Gata2 -> Cebpa ↓ Spi1 -> Cebpa ↑

Table 1: Aging perturbations of the early hematopoiesis Boolean model.

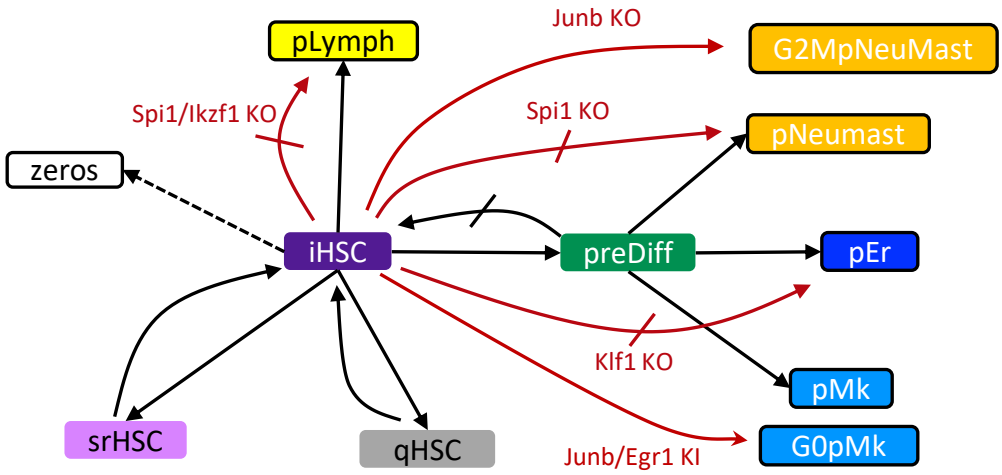


Supplementary Figure 1: Discretization of gene expressions for genes and regulons with less than 10 targets.

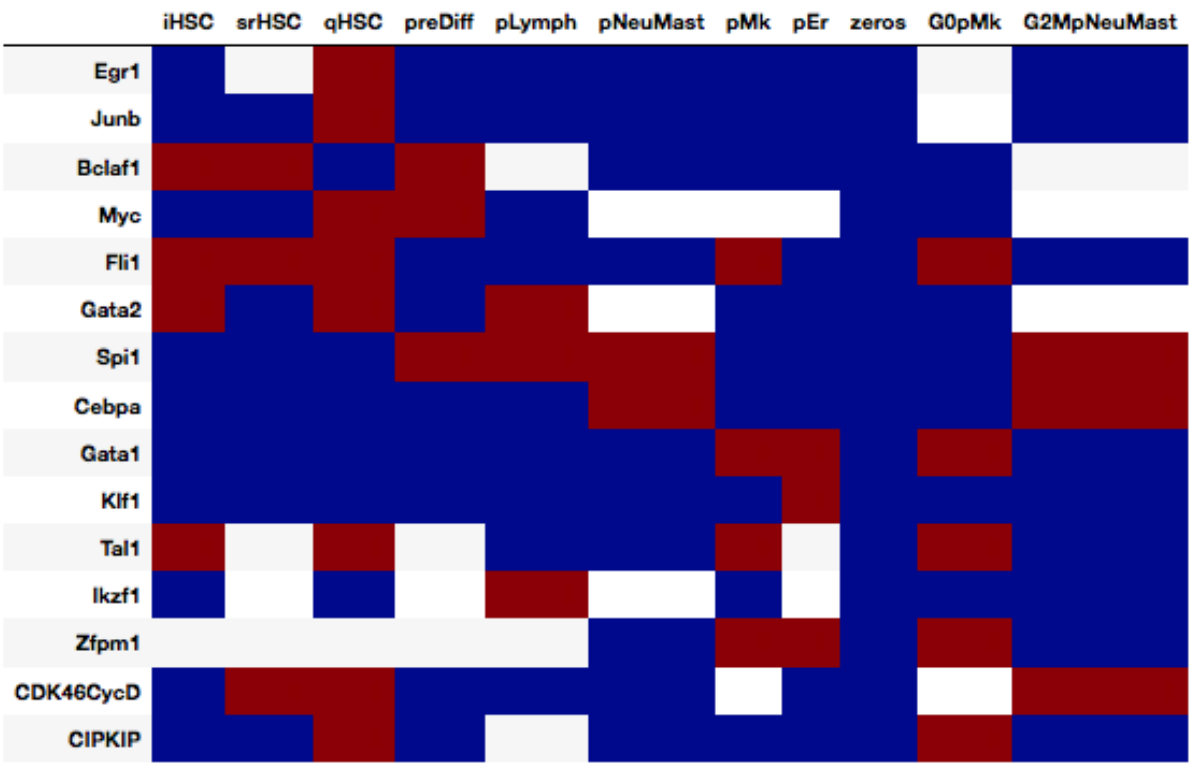


Supplementary Figure 2: Venn diagrams of influence graph interactions sources.

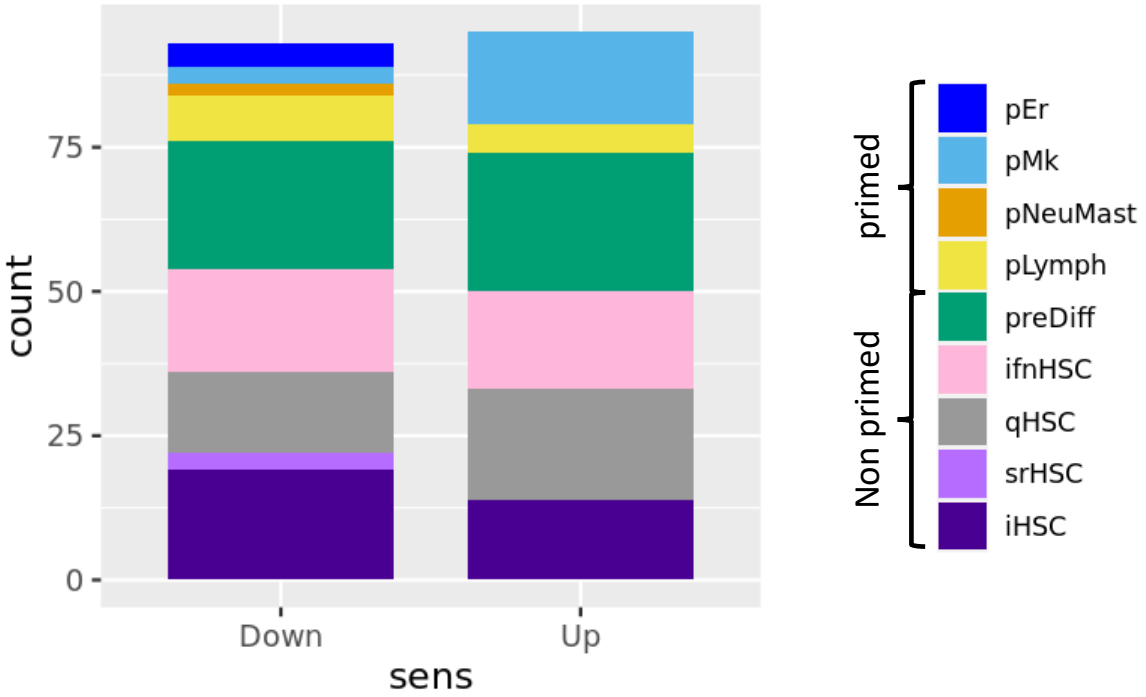
A



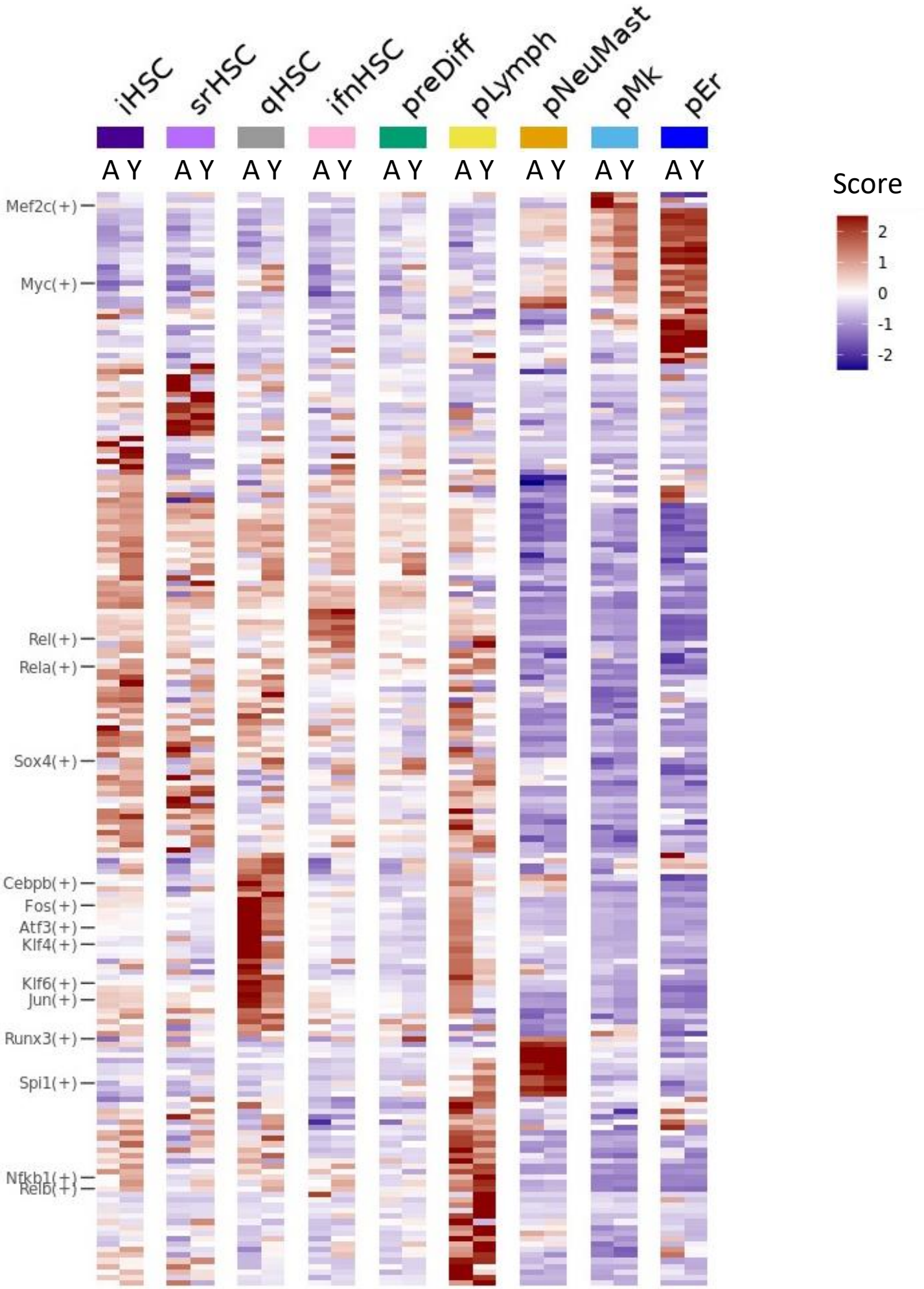
B



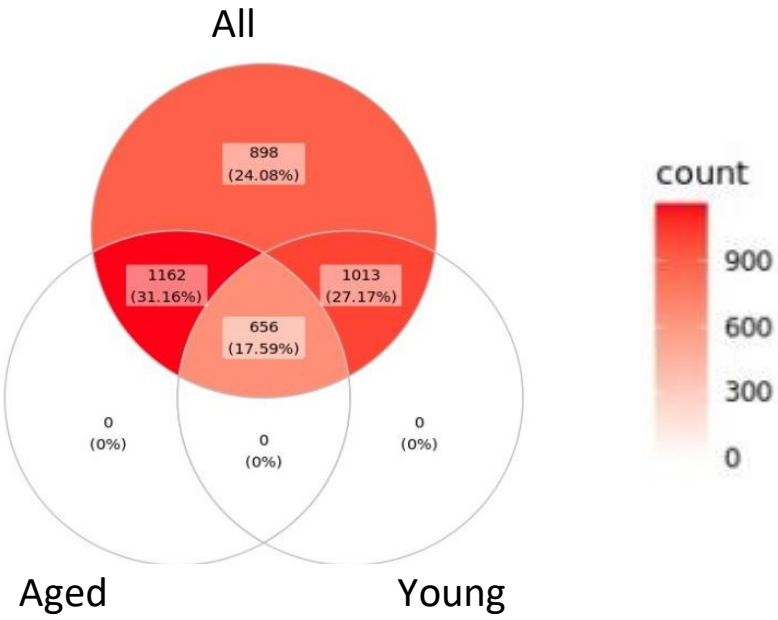
Supplementary Figure 3: Constraints and discretization used for the influence graph pruning and the final rule inference.



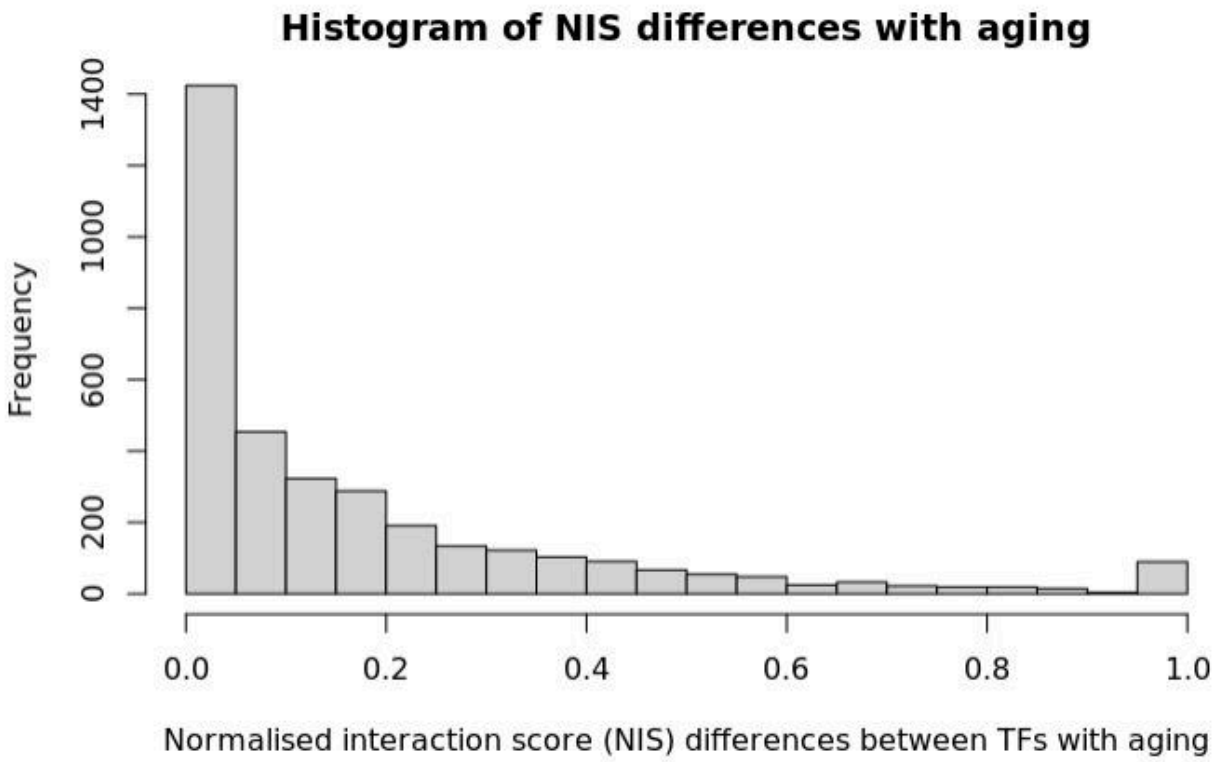
Supplementary Figure 4: HSPC state repartition of the regulon activity alterations upon aging.



Supplementary Figure 5: Heatmap of AUCell scores of regulon activity averaged by group of cells from the HSPC states in young and aged cells.



Supplementary Figure 6: Retrieved scenic interaction from all cell analysis in aged or young only cell analysis.



Supplementary Figure 7: Histogram of normalized interaction score differences with aging.