bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Data-driven modeling predicts gene regulatory network dynamics during the differentiation of multipotential progenitors

Joanna Elzbieta Handzlik[1] and Manu[1,*,a]

[1]*Department of Biology, University of North Dakota, Grand Forks, 58202-9019 ND, USA*

[*]*Correspondence: manu.manu@und.edu*

[a]*Lead Contact*

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Summary

Cellular differentiation during hematopoiesis is guided by gene regulatory networks (GRNs) thought to be organized as a hierarchy of bistable switches, with antagonism between Gata1 and PU.1 driving red- and white-blood cell differentiation. We utilized high temporal-resolution gene-expression data from *in vitro* erythrocyte-neutrophil differentiation and a predictive data-driven dynamical modeling framework to learn the architecture and dynamics of gene regulation in a comprehensive twelve-gene GRN. The inferred genetic architecture is densely interconnected rather than hierarchical. The analysis of model dynamics revealed that neutrophil differentiation is driven by C/EBPα and Gfi1 rather than PU.1. This prediction is supported by the sequence of gene upregulation in an independent mouse bone marrow scRNA-Seq dataset. These results imply that neutrophil differentiation is not driven by the Gata1-PU.1 switch but by neutrophil-specific genes instead. More generally, this work shows that data-driven dynamical modeling can uncover the causality of developmental events that might otherwise be obscured.

Keywords

hematopoiesis, erythrocyte and neutrophil differentiation, gene regulatory networks, dynamical modeling, transcription factors, ordinary differential equations, bistable switch, cytokine receptors, positive feedback, gene expression time-series data

1 Introduction

Cell-fate decisions during hematopoiesis are thought to be made by transcriptional gene regulatory networks (GRNs) [1–3], which are comprised of genes that influence each others' expression through their products. The genetic architecture, by which we mean the regulators of genes, whether each regulator activates or represses, as well as the quantitative strength of regulation, of

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

hematopoietic GRNs is not fully understood. Hematopoietic cell-fate choice has often been interpreted in the context of a simple network motif, the bistable switch [3–5]. In the bistable switch model, two TFs repress each others' expression and cell-fate is chosen in a cell-autonomous manner when small stochastic fluctuations cause the system to shift to one of two steady states corresponding to the alternative cell fates. For example, the choice between the red- and white-blood cell fates is thought to be made by mutual repression between two transcription factors (TFs), Gata1 and PU.1 (encoded by *Spi1*) [4]. Similar bistable switches have been proposed for other binary cell-fate choices in hematopoiesis [2] and more generally in development [6].

A number of recent developments suggest that the bistable switch framework might be insufficient to explain cell-fate choice and that hematopoietic GRNs have a densely interconnected architecture. Network reconstructions based on genome-wide gene expression data reveal large modules of co-regulated genes [7] and genome-wide TF binding data show that most regulatory regions are co-bound by multiple TFs [8, 9]. A second issue is that the bistable-switch hypothesis is anchored in a developmental sequence of discrete binary cell-fate decisions with well-defined intermediate progenitors. Single-cell RNA sequencing data imply however that cellular states during hematopoiesis are situated along a continuum and may not involve binary decisions [10, 11]. Bistable switches, such as Gata1-PU.1, were inferred from genetic and biochemical analyses conducted at steady state, which lack information about the dynamics and causality of events. For instance, tracking the expression dynamics of fluorescently tagged Gata1 and PU.1 in live cells suggests that rather than initiating lineage choice, the divergent expression of the two proteins is itself a consequence of as-yet-unknown upstream regulatory events [12]. Finally, the cell-autonomous bistable-switch framework cannot integrate and account for the instructive influence of cytokines on hematopoietic differentiation [13, 14].

Here we take an alternative approach to inferring the genetic architecture and dynamics of the red- and white-blood cell-fate decision. Our approach utilizes a data-driven predictive modeling methodology called gene circuits [15, 16]. Gene circuits determine the time evolution of protein or

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

mRNA concentrations using coupled nonlinear ODEs in which synthesis is represented as a switch-like function of regulator concentrations. The values of the free parameters define the regulatory influences among the genes in the network. Gene circuits do not presuppose any particular scheme of regulatory interactions, but instead determine it by estimating the values of the parameters from quantitative data using global nonlinear optimization techniques [17–20]. Gene circuits infer not only the topology of the GRN but also the type, either activation or repression, and strength of interactions. Most importantly, the inference procedure yields ODE models that can be used to interrogate the dynamics and causality of regulatory events during differentiation as well as to simulate and predict the consequences of developmental perturbations [21–24].

We inferred the genetic architecture and gene regulation dynamics underlying red- and white-blood cell differentiation using gene circuit models comprising 12 genes. The gene circuits included receptors and effectors of cytokine siganling in addition to well-known lineage specifying TFs, such as Gata1 and PU.1, so that they could incorporate the potential influence of cytokines. The gene circuits were trained on publicly available high temporal resolution genome-wide gene expression data acquired during the differentiation of an inducible cell line, FDCP-mix [4, 25], into erythrocytes and neutrophils. Most of the inferred pair-wise regulatory interactions were consistent with available empirical evidence. The models also correctly predicted the effect of knock-out, knock-down, and overexpression of key TFs both qualitatively and quantitatively. The genetic architecture of the models, instead of being hierarchical, is densely interconnected and features extensive cross-repression between genes expressed in different lineages. Furthermore, analysis of the dynamics of gene regulation suggested that *Spi1* upregulation occurred in the latter half of neutrophil differentiation, which was driven instead by two other TFs expressed in the neutrophil lineage, C/EBPα and Gfi1. We tested this prediction of the model by inspecting the sequence of gene upregulation during neutrophil differentiation in a single-cell RNA-seq dataset [11] from mouse bone marrow. These data confirmed that *Cebpa* and *Gfi1* upregulation precede that of *Spi1* *in vivo*.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

2 Results

2.1 Data-driven modeling of gene expression dynamics during the differentiation of FDCP-mix cells

2.1.1 Specification of a gene circuit model for erythrocyte-neutrophil differentiation

We constructed a gene circuit model comprising 12 main lineage-specifying TFs and cytokine receptors implicated in erythrocyte-neutrophil differentiation. *Tal1* and *Gata2* are expressed in pluripotent stem cells and are necessary for the differentiation of multiple lineages including erythrocytes [26–31]. *Gata1*, it's partner *Zfpm1*, which encodes the Fog1 protein, and *Klf1* are necessary for erythroid and megakaryocytic differentiation [2, 32–37]. All white blood-cell lineages are absent in the bone marrow of $Spi1^{-/-}$ knockout mice [38], the products encoded by *Cebpa* and *Gfi1* specify the neutrophil cell fate [3, 39], and the TF encoded by *Stat3* acts downstream of GCSF signaling [40]. Previous work has suggested that the expression level of cytokine receptors can influence the activation of lineage specifying TFs [41]. We included three genes, *Epor*, *Csf3r*, and *Il3ra*, encoding the cytokine receptors Epor, GCSF-R, and the alpha subunit of the IL3 receptor respectively [42] in order to detect such potential regulatory mechanisms. Although all of these genes are well-known participants in erythrocyte-neutrophil differentiation, the precise genetic architecture of the network remains to be determined.

2.1.2 Time-series data for training the gene circuits

We trained the gene circuit on May *et al.*'s high temporal-resolution dataset [25] of genome-wide gene expression during erythrocyte-neutrophil differentiation. May *et al.* utilized FDCP-mix cells [34] which are maintained in a multipotent state in the presence of IL3 and can be induced to differentiate into erythrocytes or neutrophils by culturing in low IL3, Epo, and hemin or GCSF and SCF respectively. In the rest of the paper, we refer to the culture of FDCP-mix cells in low IL3,

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Epo, and hemin as erythrocyte conditions and culturing in GCSF and SCF as neutrophil conditions. The dataset comprises genome-wide gene expression measurements at 30 time points during the 7-day course of differentiation towards either cell fate, with sampling frequency reducing from once every two hours during the first day to once in three days during the last three days.

The trajectories of gene expression for the modeled genes (Fig. 1) exhibit rich temporal dynamics. Whereas the expression of some genes, such as *Klf1* and *Gfi1*, diverges between erythroid and neutrophil conditions during the first few hours of differentiation, the expression of genes such as *Il3ra*, *Gata2*, and *Spi1* does not diverge until 2-3 days in to the differentiation. Besides timing, the genes differ also in the magnitude of change during the course of differentiation. Although all the genes change expression significantly over 7 days, *Il3ra* is upregulated $\sim$2-fold in the neutrophil condition while *Csf3r* is upregulated $\sim$230-fold in the neutrophil condition. Furthermore, with the exception of *Gata2*, which is upregulated in both conditions, all genes demonstrate an "either-or" pattern of regulation, being upregulated in one condition, while being downregulated in the other (Fig. 1).

2.1.3 Training the gene circuits on time-series gene expression data

We trained the gene circuits on May et al.'s time series data using a global nonlinear optimization method called Parallel Lam Simulated Annealing [PLSA; 17, 22]. PLSA is a stochastic method and results in a distinct set of parameters each time a gene circuit is inferred from data. In order to ensure that our analysis was not influenced by any idiosyncrasy of a particular model, we inferred 100 independent gene circuit models, and chose 71 that met our goodness-of-fit of criteria (Section 4.2) for further analysis.

2.1.4 Simulation of the GRN during erythrocyte-neutrophil differentiation

The output of the 71 analyzed gene circuits agreed with data to within experimental error for all 12 genes and the vast majority of time points (Fig. 1). The sole exception was that the models

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

did not reproduce a spike in *Cebpa* expression occurring around the 70 hour time point, although it is unclear whether this spike is genuine or the result of experimental error. Models trained on randomly shuffled data fit the data poorly (Section 4.3), implying that the fits to the empirical data are statistically significant (Fig. S1). Consistent with the general agreement with the data, the models' outputs reproduce all the essential dynamical features of the data—the either-or differential expression, gene-specific timing of expression divergence, and gene-to-gene variation in the dynamic range of expression.

2.2 Gene circuits predict the consequences of genetic perturbations

Having obtained gene circuits that are able to quantitatively reproduce the observed time series data, we next tested whether the models could predict the outcomes of experimental treatments *de novo*, that is, without being trained on the data from the experiments. We simulated two kinds of experiments using the gene circuits. The first class are knockouts of *Gata1* and *Spi1*, experiments that were not carried out in FDCP-mix cells but in mice or other cell types. One should not expect the model to predict the outcomes of such knockout experiments at a quantitative level since the model was neither trained on the data from these cell types nor were all of it's state variables measured in the experiments. Therefore, we compare model predictions with the results of knockout experiments at a qualitative level. The second class of experiments involved the knockdown or overexpression of key gene products followed by genome-wide expression profiling conducted by May *et al.* in FDCP-mix cells [25]. Simulation of these perturbations may be compared to experiments at a quantitative level since they share the experimental system and all of the model's state variables were measured.

2.2.1 Simulation of *Spi1* and *Gata1* knockout

We simulated *Spi1* knockout by setting its initial expression and maximum synthesis rate to zero (Section 4.2). The consequences of this perturbation differed by condition (Fig. 2). In erythrocyte

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

conditions, although the change was more rapid in the mutant, the expression of all genes moved in the same direction and attained very similar values on day 7 as the wildtype. The model predicted therefore that erythrocyte differentiation is largely unperturbed in *Spi1* mutants, which matches experimental observations from *Spi1* knockout mice [38]. Gene expression temporal profiles differed markedly between mutant and wildtype in neutrophil conditions however, and changed very little from their initial values. A lack of change in gene expression implies that cells are arrested in a progenitor state in the *Spi1* mutant during neutrophil differentiation. This prediction is supported by the observations that *Spi1* knockout mice lack mature white-blood cells [38] and that their bone marrow contains IL3-dependent granulocyte-monocyte progenitors (GMPs) [43, 44], while disruption of *Spi1* in mouse granulocyte/monocyte-committed progenitors prevents their maturation but not proliferation [45].

The results of *Gata1* knockout (Fig. S2) were opposite to those of *Spi1* knockout. In neutrophil conditions, the expression of all genes changed in the same direction and reached the same endpoints as in wildtype, albeit more rapidly, implying that neutrophil differentiation is not affected by *Gata1* mutation. In erythrocyte conditions however, gene expression of all genes did not change much from initial conditions, implying an arrest in the progenitor state. These predictions match the empirical results that embryonic stem cells (ESCs) lacking *Gata1* undergo developmental arrest at the proerythroblast stage [46] and that *Gata1*-null ESCs cultured in the presence of Epo resemble proerythroblasts [47].

2.2.2 Simulation of knockdown and overexpression experiments in FDCP-mix cells

We simulated the knockdown of *Spi1* and *Gata2* in FDCP-mix cells and compared model output to the changes in gene expression observed in experiment [25]. Since the knockdown was performed in self-renewing IL3 conditions, we set the lineage condition parameter to zero (Section 4.2) and simulated knockdown by reducing the synthesis rate of either gene and computing the solution until equilibrium was achieved. Since the knockdown efficiency achieved in the experiment is

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

unknown, we set the synthesis rate to a value that results in a fold change in the expression of the targeted gene—*Spi1* or *Gata2*—that matches the empirically observed value. Therefore, we "fit" the knockdown model to the expression of the targeted gene to predict the changes in the expression of the remaining eleven genes. Finally, this analysis—and all subsequent analyses—were performed using one representative model (model #66) out of the 71 that matched the goodness-of-fit criteria (Section 4.2).

There is strong agreement between prediction and observation for *Spi1* knockdown (Fig. 3). Consistent with the well known regulatory role of PU.1, the model predicted the upregulation and downregulation of the erythrocyte and neutrophil lineage genes respectively, which matched the pattern of gene expression observed in the experiment. The only exception was *Il3ra*, which was predicted by the model to be slightly downregulated but in fact did not change in expression. In contrast to the results with *Spi1*, the model was unable to predict the consequences of *Gata2* knockdown (Fig. 3), suggesting that aspects of *Gata2*'s regulation were inferred poorly by model training. This is corroborated by the fact that many of the Gata2-related regulatory parameters were poorly constrained (Fig. 4).

The overexpression experiments were simulated differently than knockdown experiments since the induction of the ERT fusion proteins by OHT does not change their mRNA expression directly but changes their TF activity, instead. Since the genetic interconnectivity matrix elements parameterize the activity of the TFs in gene circuits, we simulated the induction of ERT protein activity by OHT by adding a bias term to the total regulatory input of each gene. The bias term of each gene is proportional to the interconnectivity element through which the gene is regulated by the overexpressed gene (Section 4.2). Similar to the knockdown experiments, the proportionality constant is unknown and was determined by fitting the overexpression model to the expression of one of the genes. Finally, we did not fit to the expression of the overexpressed gene since the observed mRNA includes an unknown contribution from the ERT fusion transgene.

The model was able to correctly predict the change in expression of all the genes except *Il3ra*

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

in the GATA1ERT experiment (Fig. 3). The quantitative agreement between model prediction and experiment was also good with the exception of *Epor*, for which a $\sim$1.5-fold upregulation was predicted while a $\sim$3-fold upregulation was observed. In the PU.1ERT experiment, the model predicted the change in expression of all genes except *Cebpa*, *Gfi1*, and *Gata2*. Whereas the model predicted an upregulation of these genes upon PU.1 overexpression, these genes were found to be downregulated in the actual experiment. Although the model did not match experiment for these genes, it is worth noting that the downregulation of *Gfi1* and *Cebpa* observed in experiment is inconsistent with the known role of PU.1 as an activator of these white-blood cell lineages genes [48–51] as well as their downregulation upon *Spi1* knockdown.

2.3 Erythrocyte-neutrophil GRN architecture is non-hierarchical and evolves in time

Having verified that the inferred models have predictive ability, we next determined the architecture of the GRN implied by the values of the genetic interconnectivity parameters, T_{ij}. T_{ij} determines how the product of gene j regulates gene i, where positive or negative values denote activation or repression respectively. The distributions of the majority of interconnectivity parameters across the ensemble of 71 analyzed models were well constrained and distinguishable as either activation or repression (Fig. 4 and Table S1). For example, the positive values of $T_{Gata1 \to Gata1}$ (Fig. 4A) and $T_{Spi1 \to Spi1}$ (Fig. 4B) in all but one model implies that both genes autoactivate while the negative values of $T_{Gata1 \to Spi1}$ (Fig. 4B) and $T_{Spi1 \to Gata1}$ (Fig. 4A) in all analyzed models implies that the two genes repress each other. We compared the inferred genetic interconnections to published empirical evidence (Fig. 4 and Table S2). The model inferred the correct role, activation or repression, for 58 of the 69 interconnections that we found empirical evidence for. The vast majority of the interconnections have not been previously examined and the model therefore implies novel inferences about the genetic architecture of the network.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The experimental evidence was inconclusive or conflicting in some instances (Table S2). Notably, the model inferred that Gfi1 activates *Spi1*, upregulated during FDCP-mix neutrophil differentiation, and represses *Gata1*, *Klf1*, and *Epor*, genes downregulated during FDCP-mix neutrophil differentiation (Fig. 1). These model inferences are supported by single-cell RT-qPCR data that show that *Gfi1* expression is positively correlated with *Spi1* expression in GMPs, LMPPs, and HSCs, while it is negatively correlated with *Gata1* expression in HSCs and GMPs [52]. Furthermore, Gfi1 is known to cooperate with C/EBPϵ to activate neutrophil genes [53, 54]. Contradicting the model's inferences and the above evidence, *Spi1* is upregulated in MPPs from *Gfi1*$^{-/-}$ mice [55, 56] while *Gata1*, *Klf1*, and *Epor* are downregulated in bone-marrow cells from *Gfi1*$^{-/-}$ mice [57]. The conflicting evidence and lack of agreement between the model and data may be a result of the pleiotropic roles that *Gfi1* plays in both HSC maintenance and neutrophil development [58]. As noted in the previous section, the regulatory parameters of *Gata2*, another gene acting pleiotropically in HSCs, the erythroid-megakaryocytic lineage, and the myeloid lineage [26, 59], were poorly or incorrectly inferred. These inconsistencies were, however, a small proportion of the total inferences and the overall good agreement between model inference and empirical evidence (Fig. 4 and Table S1) suggests a successful decoding of the genetic architecture.

The genetic architecture of the network, in fact, changes in time since the strength of the regulation of one gene by the products of another gene depends on the concentration of the latter, which evolves during the differentiation process. In order to gain insight into this "dynamical GRN", we computed the time-dependent regulatory contribution, given by the product of the genetic interconnectivity parameters by the concentrations of the cognate regulators ($T_{ij} \cdot x_j^l(t)$), for all pairs of regulators and targets in the model. The GRN may then be represented as a graph in which each gene is a node and the type—activation or repression—and time-dependent strength of regulation between each gene pair is an edge (Fig. 5).

The erythrocyte-neutrophil network inferred by the model is densely interconnected with genes

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

associated with the erythrocyte lineage repressing genes of the neutrophil lineage and vice versa. This conclusion is in agreement with other analyses based on genome-wide gene expression data [7] and contrasts the view that the genetic architecture consists of a hierarchy of bistable switches [2]. The time evolution of the network reveals two broad principles. First, there is a preponderance of repressive interactions at earlier time points during the differentiation suggesting that the cell-fate decision is dictated by loss of repression rather than a gain of activation. Conversely, activation between co-expressed genes gains prominence at later time points, suggesting that activation mainly reinforces the decision once it has been made.

2.4 Gene circuits predict that C/EBPα and Gfi1 drive neutrophil differentiation

How each gene in the network is regulated is, as discussed earlier, not static but changes as the concentrations of its regulators evolve in time during differentiation. We reasoned that the temporal dynamics of gene regulation could provide insight into the causality of the regulatory events underlying differentiation. The temporal dynamics of gene regulation can be analyzed by "looking under the hood" of the gene circuit model and decomposing the total regulatory input for each gene into contributions from individual regulators (Fig. 6; see Section 4.2 for details). In Figure 6, the total regulatory input (dotted black line) is plotted in time. A gene is at half its maximum activation when the total regulatory input is zero and thus the time at which this happens (black vertical lines) serves as a marker to order the sequence in which genes turn on or off as differentiation proceeds. The contributions of repressors and activators are shown as shaded sections above and below the total regulatory input respectively. The regulators accounting for the up- or down-regulation of a gene can be determined by noting their contribution to the change in the total regulatory input. For example, the bulk of the change in *Cebpa*'s regulatory input from the start of neutrophil differentiation to reaching half-max expression is the result of autoactivation (light blue) and activation by

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Gfi1 (dark blue; Fig. 6).

Several observations can be made regarding the temporal dynamics of gene regulation during erythrocyte-neutrophil differentiation (Fig. 6). All the genes are in a partially repressed state, since their total regulatory input is negative, in undifferentiated FDCP-mix cells. This is reminiscent of multilineage transcriptional priming [3, 60, 61]—the low-level expression of genes from multiple lineages in multipotential progenitors. What accounts for the repression varies by the target gene. Genes downregulated in neutrophil conditions, *Gata1*, *Zfpm1*, *Klf1*, *Tal1*, and *Epor*, are repressed by several genes of small effect. Genes downregulated in erythrocyte conditions however, *Spi1*, *Cebpa*, *Gfi1*, *Stat3*, *Gata2*, *Il3ra*, and *Csf3r*, are mainly repressed by a combination of *Zfpm1* and *Tal1*.

During erythrocyte differentiation, all the upregulated genes are activated more or less simultaneously since they reach half-max activation in a short $\sim$30 hour window (Fig. 6). Upregulation of the genes involves both the loss of repression as well as increased activation (Fig. 6). The three main activating influences are *Gata1*, *Klf1*, and *Epor*. The first two are well known activators of erythrocyte genes, while the activating influence of *Epor* implies that upregulation of the receptor's gene expression provides positive feedback, indirectly, to the TFs driving erythroid differentiation.

In contrast to erythrocyte differentiation, the sequence of activation of genes during neutrophil differentiation is spread out over $\sim$100 hours (Fig. 6). Surprisingly, *Spi1* is one of the last genes in the activation sequence, reaching half-max activation around day 5 of the differentiation process, while *Gata2* and *Cebpa* are the first ones to be activated. Unlike erythrocyte differentiation, during which three activators provided activation throughout the process, the genes accounting for activation change in time and with target gene.

While PU.1 and Csf3r provide activation during later stages of the differentiation, *Cebpa* and *Gfi1* together account for most of the early activation of the genes upregulated during neutrophil differentiation (Fig. 6). Although *Gfi1* expression is positively correlated with genes upregulated during neutrophil differentiation in FDCP-mix cells (Fig. 1) and with *Spi1* in GMPs [52],

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Gfi1 is known to function primarily as a repressor in MPPs and the lymphoid and myeloid lineages [53, 56, 62, 63]. The activation role inferred here for Gfi1 during neutrophil differentiation could result from indirect regulation of its targets. Another factor is the high level of similarity between the expression of *Cebpa* and *Gfi1* in the training data (Fig. 1) that renders the two factors interchangeable in the model. C/EBPα is known to directly activate itself, *Spi1*, *Csf3r*, and *Gfi1* during neutrophil differentiation [Table S1; 48–50, 64–68]. We conclude therefore that the activation of neutrophil targets by Gfi1 inferred by the model could, in fact, represent the activity of C/EBPα. Taken together this analysis implies that neutrophil differentiation in FDCP-mix cells is driven by C/EBPα and potentially Gfi1 acting indirectly [69], which activate *Spi1* at later time points. In other words, dynamical modeling suggests that *Cebpa* and *Gfi1* are upstream of *Spi1* in the genetic architecture of neutrophil differentiation.

2.5 *Cebpa* and *Gfi1* expression precedes *Spi1* upregulation in the neutrophil lineage in mouse bone-marrow hematopoietic progenitor cells

We sought confirmation of the sequence of gene activation implied by our model of FDCP-mix cell differentiation in an independent experimental system. We analyzed Tusi *et al.*'s single-cell RNA-seq (scRNA-Seq) data from Kit$^+$ mouse bone-marrow HPCs. Although scRNA-Seq data are a static snapshot of the progression of cell states during steady-state hematopoiesis, it is possible to infer the order of cell states under a few assumptions. Weinreb *et al.* developed Population Balance Analysis (PBA) [70], which computes the probability of transitions between the cell states—defined by genome-wide gene expression—observed in single-cell gene expression data and hence the probability that an intermediate cell state will evolve into some terminal cell fate (Fig. 7A). Cell states corresponding to multipotential progenitors—the origin of the differentiation process—and committed unilineage progenitors—the termini of the differentiation process—are identified by the expression of marker genes. PBA assumes that there are no oscillations in cellular state so

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

that the dynamics are governed by a potential function of cellular state and cells always move from higher to lower potential [70, Fig. 7C]. Under this assumption, it is possible to order the cells in developmental time by arranging them in order of decreasing potential [see 70, for details].

We profiled the expression of *Cebpa*, *Gfi1*, and *Spi1* in Tusi *et al.*'s dataset by identifying cells having a high probability of becoming neutrophils based on the fate probabilities assigned to them by PBA (Fig. 7A). The potential decreases with increasing neutrophil probability (Fig. 7B) and it is possible to visualize how gene expression changes with developmental age at a single-cell level (Fig. 7D) by following the direction of decreasing potential. Since single-cell read counts have considerable cell-to-cell variability, we also divided the potential into 11 bins containing an equal number of cells and averaged the expression over the cells in each bin (Fig. 7B). These data show that *Cebpa* and *Gfi1* expression precedes that of *Spi1* during differentiation toward the neutrophil fate. *Cebpa* is already at it's maximum level at the highest potential or earliest developmental stage. *Gfi1* rises rapidly at earlier stages and peaks at bin 7. *Spi1*, although expressed at lower levels at the earlier stages, changes relatively little until bin 6. *Spi1* is upregulated subsequently and reaches its maximum expression in bin 9 and maintains that level until the latest stage captured in this dataset. Interestingly, both *Cebpa* and *Gfi1* are downregulated to lower levels in the latest developmental stages. These inferred temporal patterns of gene expression during the granulocytic differentiation of bone-marrow HPCs are consistent with our model's predictions that *Cebpa* and *Gfi1* are upstream of *Spi1* in the genetic architecture of neutrophil development.

3 Discussion

Despite our knowledge of the main genes effecting hematopoietic cell-fate decisions, their genetic architecture as well as the causality of their regulation is not fully understood. Here we have taken the approach, complementary to empirical genetic analyses, of learning the genetic architecture by training gene circuit models on gene expression time-series data. We trained a comprehensive

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

model comprising 12 genes encoding TFs and cytokine signaling components on a high-temporal resolution dataset [25]. We showed that this model is biologically accurate by correctly predicting the consequences of genetic perturbations at a quantitative level. Similarly, we demonstrated through a detailed comparison with literature that the model correctly inferred the nature, activation or repression, of most known pairwise interactions. Our analysis implies that the genetic architecture of the erythrocyte-neutrophil decision is non-hierarchical and highly interconnected. There are extensive repressive interactions between genes from alternative lineages, while there is positive feedback from cytokine receptors. Furthermore, the gene circuit approach goes beyond static GRNs, and reveals their dynamics during the differentiation process. We found that repressive interactions dominate at the earliest stages of the cell-fate decision while activation gains importance only at later stages. Finally, we show through model analysis followed by validation in an independent scRNA-seq dataset [11] that *Spi1* is downstream of *Cebpa* and *Gfi1* in the chain of causation leading to the specification of neutrophils.

Hematopoietic cell-fate decisions have been modeled by two main approaches so far. In the first approach, the GRN is modeled using ODEs [3, 4, 71, 72] and the quantitative values of parameters are fixed by an exhaustive search of the parameter space to find regions that reproduce the qualitative behavior of the GRN. Such models have been mostly limited to 2–3 well-known "master" regulators, perhaps due to their relatively high computational expense. The second approach circumvents the high computational expense of ODEs by constructing logical or Boolean models that are more comprehensive and include 11–20 genes [73, 74]. The two approaches are similar in that the genetic architecture implemented by the models is based on prior empirical evidence.

In contrast to previous efforts, gene circuits do not rely on prior empirical data for defining the genetic architecture, but instead learn it from gene expression time-series data. Gene circuits therefore offer an independent means of decoding the genetic architecture to supplement, but also to potentially refine, what we know from purely empirical approaches. The utility of this independence is highlighted by our results about the placement of *Spi1* in the hierarchy and the highly

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

interconnected and non-hierarchical topology of the GRN. Our model is also the most comprehensive ODE model to date, simulating a GRN of a size previously only attempted with Boolean networks but without assuming that gene expression is restricted to a few discrete levels.

Analysis of gene regulation dynamics in the model followed by validation in an independent dataset [11] led us to the insight that *Cebpa* and *Gfi1* and not *Spi1* are the causal drivers of neutrophil differentiation. *Spi1* has been thought to reside at the top of the hierarchy [2–4, 6, 25, 26] of white-blood cell genes since *Spi1* knockout mice lack all white-blood cells [38]. Additionally, evidence that PU.1 inhibits Gata1 [75] and vice versa [76] led to a model in which *Gata1* and *Spi1* form a bistable switch that decides the fate, while all the other genes are downstream targets of Gata1 or PU.1 [6]. However, the Gata1-PU.1 bistable switch model has been questioned recently by experiments in which Gata1 and PU.1 expression was monitored in differentiating HSPCs [12]. These experiments failed to detect an intermediate stage where cells co-expressed low amounts of both Gata1 and PU.1, which is a necessary condition for the fate decision to be driven by the genes' mutual repression. Furthermore, in cells destined for a myeloid fate, PU.1 was expressed at a constant level before being upregulated during the later stages of commitment while Gata1 remained undetectable throughout. This observation suggested that some factor or factors other than Gata1, unknown heretofore, drive PU.1 upregulation during myeloid differentiation. We have thus arrived by independent means at conclusions similar to those of [12] and have further identified the upstream factors driving PU.1 upregulation during myeloid differentiation.

That *Spi1* is downstream of *Cebpa* and *Gfi1* could be reconciled with its mutant phenotype—the absence of all white blood cells in knockout mice [38]—if *Spi1* were necessary for the activation of all white-blood cell genes. *Spi1* would then be seen as a hub which integrates input from lineage-specifying genes such as *Cebpa* and *Gfi1* and coordinates the expression of all white blood-cell genes. Whereas *Spi1* was accorded both the role of an initiator and a hub in the bistable switch model [6], our results suggest that these two functions are separable and that *Spi1* only performs the latter function. Another question about the causality of neutrophil differentiation is

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

whether *Cebpa* and *Gfi1* are intermediaries or the ultimate cause of neutrophil differentiation. A potential ultimate cause is GCSF signaling because *Cebpa* is known to be upregulated by GCSF treatment [44, 49, 50, 77]. The C/EBPα protein is phosphorylated downstream of GCSF signaling [78] and autoactivates *Cebpa* transcription by binding to its promoter [66] and enhancers [48–50], providing a potential mechanism for GCSF-mediated *Cebpa* upregulation. GCSF signaling is certainly the ultimate cause of *in vitro* neutrophil differentiation of FDCP-mix cells since differentiation is induced by treating the cells with GCSF. Given that GCSF acts in an instructive manner to specify the granulocytic fate [14], we propose that this is the case *in vivo* as well.

Gene regulation during differentiation is dynamic; the contributions of the regulators modulating a gene's transcription and the overall balance of activation and repression change as the regulators' concentrations vary in time. Gene circuits, being dynamical models, allow us to determine how regulatory control varies in time both at the level of individual target genes (Fig. 6) and more broadly at the network level (Fig. 5). Our analysis indicates that, both at the individual and global levels, repression dominates over activation at earlier stages of eruthrocyte-neutrophil differentiation. As a result, all the genes in the network are partially repressed and expressed at low levels in progenitors. The data support this inference. Each gene in the network is upregulated by at least two-fold in one lineage or the other (Fig. 1), which implies that the expression level observed in the progenitors is significantly below that of an actively transcribed gene.

The predominance of repression in the earlier stages implies, in turn, that the divergence of gene expression during differentiation is driven by relief of repression rather than by activation. This is similar to the idea of lineage priming [3, 4, 60, 79–81] in the bistable switch model [3, 4], where genes from alternative lineages are expressed at low levels and repress each others' expression in progenitors. Our model differs from the bistable switch model in two ways. First, whereas cell fate is selected by the initial concentrations of the two genes in the bistable switch model, cytokines select the fate by exerting asymmetric effects on each gene in the gene circuits modeled here (Section 2). The second difference is that many more genes participate in cross-antagonism

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

than Gata1 and PU.1 as hypothesized in the bistable switch model.

The overall balance shifts in favor of activation at later stages of differentiation, leading to the establishment of positive feedback loops between genes co-expressed in the same lineage. Of note is the activation of lineage-specific TFs by cytokine receptors. In the model, *Csf3r*, which codes for GCSFR, provides substantial activation to most of the genes upregulated in the neutrophil condition, while *Epor* performs a similar function in the erythrocyte condition (Fig. 6). As discussed above, *Cebpa* is known to be downstream of GCSF signaling as are other myeloid TFs [40]. Similarly, EpoR phosphorylates and activates Gata1 through the PI3K-AKT pathway [82] and Epo signaling positively regulates several erythroid genes [83–86]. Cytokine receptor-mediated positive feedback has been shown to generate bistability in a model of Epo-dependent *Gata1* activation [41], resulting in greater sensitivity to Epo cytokine concentration. The positive feedback loops inferred in this bigger GRN might also result in bistability or multistability and sharp responses to cytokine concentration, a possibility that awaits confirmation through non-linear stability analysis [87].

Despite its general success in predicting the consequences of genetic perturbations, the model was unable to do so for *Gata2* knockdown (Fig. 3), highlighting a limitation of the gene circuit methodology. The model predicted nearly the exact opposite of the observed effects. The neutrophil lineage genes were predicted to be downregulated about two-fold, when in fact they were upregulated 1.2–4 fold, while erythrocyte lineage genes were predicted to be upregulated instead of being downregulated about two-fold (Fig. 3). These mispredictions may be traced to the fact that Gata2-related parameters were not inferred with much certainty during fitting. 4 of 12 of the interconnectivity parameters (T_{ij}) where Gata2 is the regulator and 4 of 12 of the interconnectivity parameters where *Gata2* is the target are indistinguishable from zero among the gene circuits that met goodness-of-fit criteria (Fig. 4 and Table S1). This implies that the goodness-of-fit was insensitive to the type, activation or repression, of those interconnections. The uncertainty about how Gata2 regulates its targets and how it is regulated itself likely arises from the fact that there

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

is almost no divergence in *Gata2* expression between the erythrocyte and neutrophil conditions (Fig. 1), with differences discernible only at one time point out of thirty. The lack of different patterns of expression in the two conditions means that the *Gata2* data do not bear sufficient information to constrain *Gata2*'s regulatory parameters. Similarly, some of the inferences, such as the activation of *Spi1* and repression of *Gata1*, *Epor*, and *Klf1* by Gfi1 (Fig. 4 and Table S2), that did not match empirical data probably resulted from a lack of training data from MPPs, monocytes, and lymphocytic progenitors, where Gfi1 exerts the experimentally observed effects [57, 58, 69]. This limitation of the gene circuit methodology—that the training dataset may not contain sufficient information to accurately infer certain regulatory parameters—may be overcome by experimental designs that either sample differentiation trajectories in a larger number of conditions and cell types or after genetic perturbations.

In gene circuits, the interconnection between a pair of genes can represent both direct and indirect regulation of one by the other. This design choice has both advantages and disadvantages. On the one hand, this flexibility leads to inferred GRNs that are not completely specified mechanistically. We could not hope to delineate GRNs with biochemical details relying exclusively on gene circuits. On the other hand, this very flexibility also makes predictive modeling of GRN dynamics feasible. Although biochemically detailed models of intracellular signaling [41] and gene regulation [49, 50] have been constructed for individual pathways and enhancers, it is currently not possible to model multiple signaling pathways or the gene regulation of multiple genes simultaneously. The challenges involved in constructing comprehensive but biochemically detailed models are many; the components are yet to be completely delineated, it is impractical to measure all the biochemical parameters, learning them from data leads to highly underdetermined problems, and the computational cost of such models would be prohibitive. Gene circuits, by coarse-graining much of the biochemical detail allow the construction of more complete models that are predictive in spite of a lack of biochemical detail.

Our results show that the temporal dynamics of gene expression bear information about the

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

genetic architecture underlying cell-fate choice. With a few exceptions such as the segmentation system of *Drosophila* [88], our current knowledge of the genetic architecture of most developmental systems is based on genetic analyses carried out at end points. Coupling gene circuits with high temporal resolution time series data is a viable complementary approach to decode the genetic architecture and reveal the causality of events during differentiation. One potential drawback of this approach is the cost of sequencing. However, the cost of sequencing is expected to decline exponentially over time [89] and is not likely to be a limitation in the future. Another concern is the high computational cost of fitting the gene circuits, which entails the use of parallel computers. This challenge was recently overcome by an algorithm called Fast Inference of Gene Regulation (FIGR) [16] that is much more computationally efficient and can infer models on a consumer-grade computer in a reasonable amount of time. We anticipate that with these improvements, it will be possible to collect time series datasets that span multiple hematopoietic lineages and genetic backgrounds and use the gene circuit approach to comprehensively decode the genetic architecture of hematopoietic cell-fate decisions.

Acknowledgments

We thank YL Loh for discussions and comments. This work was supported by the National Science Foundation [1615916 to M.].

Author contributions

J.E.H. and M. conceived and designed the study. J.E.H wrote software code and performed the analysis. M. supervised the analysis. J.E.H and M. wrote the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Declaration of interests

The authors declare no competing interests.

Figure Legends

Figure 1: Gene expression time series data vs. model output. Mean normalized gene expression measurements and model output for the 12 modeled genes are plotted as circles and lines respectively. Errors bars show standard deviation over 3 replicates. The output of the 71 models that met the goodness-of-fit criteria (Section 4.2) are shown simultaneously. Data and model output for FDCP-mix cells cultured in low IL3, Epo, and hemin, referred to as the erythrocyte condition hereafter, are shown in red. Data and model output for FDCP-mix cultured in GCSF and SCF, referred to as the neutrophil condition hereafter, are shown in blue.

Figure 2: Simulation of *Spi1* knockout. *Spi1* knockout was simulated in all 71 models that met the goodness-of-fit criteria. Their output is plotted as lines. The symbols and colors are the same as Figure 1.

Figure 3: Simulation of knockdown and overexpression of key transcription factors in FDCP-mix cells. The fold change in gene expression in simulations of *Spi1* and *Gata2* knockdown (top two panels) or PU.1 and Gata1 overexpression (bottom two panels) is plotted against the fold change observed in experiment. The dotted lines correspond to no change so that points in the green quadrants indicate qualitative agreement and points in the red quadrants indicate qualitative disagreement between prediction and observation. The green line represents perfect quantitative agreement between prediction and observation.

Figure 4: Inferred genetic architecture. The distribution of each genetic interconnectivity pa-

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

rameter (T_{ij}) over the ensemble of 71 models is shown as a box plot. The distribution of the each regulatory parameter representing the influence of cytokine conditions (b_i) is shown as a box plot ("Ext. Sig."). In the box plots, the box lines are the first quartile, median, and third quartile. The whiskers extend to the most extreme values lying within 1.5 times the interquartile range. Individual parameter values inferred by the models are shown as circles overlaid on the box plots. Each panel shows the regulation of a particular target. Positive and negative values of T_{ij} indicate activation and repression respectively. Positive values of b_i indicate activation by Epo and repression by GCSF while negative values indicate activation by GCSF and repression by Epo. Activation is inferred if the first quartile of the distribution is positive, while repression is inferred if the third quartile is negative. The type of regulation is considered to be poorly constrained when the interquartile range spans negative and positive values. Positive values of The parameters whose inferred sign agrees with prior empirical evidence (Table S2) are marked as dark green while those that are contradictory are marked as red. The parameters for which there is empirical evidence for an interaction but the type of interaction, activation or repression, is not known are marked as light green. The parameters for which we were unable to find experimental evidence, the experiments yielded negative results, or the sign was unconstrained are marked as brown.

Figure 5: The time evolution of the inferred GRN. The GRN is depicted as a graph at different time points during differentiation in both erythrocyte and neutrophil conditions. The contributions of each regulator to the regulation of its targets, given by the product of the pairwise genetic interconnectivity parameter and the regulator's concentration, are shown as edges from the regulator to the targets. Blue and red edges correspond to activation and repression respectively, while the opacity of the lines indicates the strength of regulation. The maximum opacities of activation and repression have been normalized to 1 separately.

Figure 6: The dynamics of gene regulation during differentiation. The total regulatory input

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

(u) is plotted as the dotted black line. The colored layers show the regulatory contribution of individual regulators. See Section 4.2 for the definitions of total regulatory input and regulatory contributions. The contributions of repressors and activators are shown above and below the dotted line respectively. The vertical dashed line in the center corresponds to uninduced FDCP-mix cells at the start of differentiation. Regulatory contributions during erythrocyte and neutrophil differentiation are shown to the right and left of the dashed line respectively. The vertical black line marks the time when the total regulatory input crosses zero so that synthesis occurs at half its maximum rate (Section 4.2).

Figure 7: The expression of *Cebpa*, *Gfi*, and *Spi1* in individual hematopoietic progenitor cells from murine bone marrow. Panels **A**, **C**, and **D** are SPRING plots [90] of Tusi *et al.*'s scRNA-Seq dataset [11] of mouse bone-marrow derived Kit^+ progenitors. Each point corresponds to an individual cell and cells are arranged as a k-nearest-neighbor (knn) graph according their pairwise distances in gene expression space [90]. **A**. The probability of a cell to adopt the neutrophil fate, as computed by the PBA algorithm, is shown as a color map if the probability is greater than 0.5. **B**. The potential landscape of the cells fated to be neutrophils is shown as a color map and orders the cells according to their maturity or developmental age. **C**. The expression of *Cebpa*, *Gfi1*, and *Spi1* is shown as a color map. Cells with no expression are colored black. **D**. The mean expression of *Cebpa*, *Gfi1*, and *Spi1* in cells binned according to their potential. Each bin contains 129 cells. The expression of each gene has been normalized to its maximum expression over the bins. The error bars show standard error.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

4 STAR Methods

4.1 RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Manu (manu.manu@und.edu).

Materials Availability

This study did not generate new unique reagents.

Data and code availability

- The gene circuit code and scripts used in this study are available at `https://github.com/mlekkha/EryNeu`.
- Parameter values of the generated gene circuits are included in the TableS1.

4.2 METHOD DETAILS

4.2.1 Gene Circuit Model of Erythroid-Neutrophil differentiation

A gene circuit [22] computes the time evolution of mRNA concentrations of a network of interacting genes by solving the coupled ordinary differential equations (ODEs)

$$\frac{dx_i^l}{dt} = R_i S\left(\sum_{j=1}^{N} T_{ij} x_j^l + b_i c^l + h_i\right) - \lambda_i x_i^l, \tag{1}$$

where $x_i^l(t)$ is the concentration of the mRNA of gene i at time t in lineage (or condition) l, and N is the total number of genes in the model. The synthesis rate depends on the concentrations of a gene's

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

regulators through the sigmoidal regulation-expression function $S(u) = \frac{1}{2}\left(u/\sqrt{(u^2+1)}+1\right)$. $S(u)$ determines the fraction of the maximum synthesis rate R_i attained by the gene given the total regulatory input $u = \sum_{j=1}^{N} T_{ij}x_j^l + b_i c^l + h_i$. The first term of u, $\sum_{j=1}^{N} T_{ij}x_j^l$, represents the regulation of gene i by the other genes in the network. Positive and negative values of T_{ij} signify activation and repression of gene i by gene j respectively. The regulation of gene i by factors specific to the condition l that have not been explicitly represented in the model is represented by the second term of u, $b_i c^l$, where c^l is -1, 0, or 1 for neutrophil, progenitor, and erythroid conditions respectively. The threshold h_i determines the basal synthesis rate and λ_i is the degradation rate of mRNA for gene i.

The initial conditions were given by the mRNA concentrations in progenitor cells. Equations 1 were solved numerically using the Bulirsch-Stöer adaptive step-size solver to an accuracy of 10^{-3} as described previously [23].

4.2.2 Training Data

The gene circuit was trained on May *et al.*'s genome-wide gene expression time-series dataset [GEO GSE49991; 25] acquired during the differentiation of FDCP-mix cells into erythrocytes or neutrophils. See [25] for the details of data processing and cross-sample normalization. The expression level of each gene was further normalized against it's maximum expression in either condition for model training and visualization.

4.2.3 Optimization by Parallel Lam Simulated Annealing (PLSA)

The parameters of Equation 1 were inferred by minimizing the cost function

$$E = \sum_{i,m,l} \left(x_i^l(t_m) - \hat{x}_i^l(t_m)\right)^2 + \text{Penalty}, \tag{2}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

where $x_i^l(t_m)$ and $\hat{x}_i^l(t_m)$ are model output and data respectively for gene i in lineage/condition l at time t_m. The penalty is a weighted regularization term that limits the search space or magnitude of the regulatory parameters T_{ij}, b_i, and h_i. The penalty is given by

$$\text{Penalty} = \begin{cases} \exp(\Pi) - \exp(1), & \text{if } \Pi > 1 \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

where

$$\Pi = \sum_i \Lambda_i \left(\sum_j (T_{ij}\hat{x}_j^{\max})^2 + (b_i^{\max})^2 + h_i^2 \right).$$

$\hat{x}_j^{\max}$ is the maximum expression of gene j observed in the dataset and $b_i^{\max}$ is the maximum value of b_i^l over all conditions l. Λ_i controls the magnitude of the regulatory parameters of gene i. Λ_i was set to 0.1 for all genes except *Csf3r*, for which Λ_i was set to 0.01. This allowed *Csf3r*'s regulatory parameters to have larger values, which was necessary for the model to be able to recapitulate the large dynamic range of *Csf3r* expression data (Fig. 1).

The cost function (Eq. 2) was minimized using parallel Lam simulated annealing (PLSA)—simulated annealing with the Lam cooling schedule [91]—running in parallel [17] as described previously [22]. PLSA was carried out on 10 CPUs (Intel Xeon E5-2643 v3 cores) in parallel.

4.2.4 Selection of gene circuits for analysis

Since PLSA is a stochastic method [17], each optimization attempt results in different values of inferred parameters and hence in a distinct gene circuit model. In order to evaluate their reproducibility, we repeated the optimization to obtain 100 different gene circuits. The root mean square (RMS) score,

$$\text{RMS} = \sqrt{\frac{E}{N_d}}, \tag{4}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

where N_d is the total number of data points, was used to measure the goodness-of-fit of each gene circuit model. We chose 71 gene circuits having RMS scores lower than 0.06, corresponding to an average error of 6% in expression levels. Models with higher RMS scores showed qualitative defects in their expression patterns compared to data.

4.2.5 Simulation of perturbation experiments

Gata1 and *Spi1* knockout was simulated by setting their initial concentrations and mRNA synthesis rates R_i to zero.

To simulate the knockdown and overexpression experiments carried out by [25] in FDCP-mix cells, we chose one representative model from the 71 that had met the goodness-of-fit criteria. For each model, we determined the number of regulatory parameters (T_{ij}) that had the same sign as the majority of the models. Of the 7 models having the largest number of regulatory parameters aligning with the consensus, one model, model #66, was chosen at random for perturbation simulations.

The knockdown *Spi1* or *Gata2* in FDCP-mix cells was simulated by decreasing the maximum synthesis rate of the gene, R_{Spi1} or R_{Gata2}, respectively. Since the efficiency of the knockdown achieved in the specific experiments was unknown, we chose the value of R_{Spi1} or R_{Gata2} so that the simulated expression of *Spi1* or *Gata2* matched the empirical values respectively. The simulations therefore could be said to predict the expression of only 11 of the 12 genes.

In the PU.1ERT and GATA1ERT experiments, 4-hydroxy-tamoxifen (OHT) treatment did not directly modulate the amount of *Spi1* or *Gata1* mRNA but instead increased the activity of the constitutively expressed PU.1ERT and GATA1ERT fusion proteins. We simulated the increase in the activity of PU.1 or Gata1 by introducing a constant bias term B_i in the total regulatory input u of each gene i,

$$u = \sum_{j=1}^{N} T_{ij} x_j^l + b_i c^l + h_i + B_i.$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The bias term is proportional to the genetic interconnectivity parameter corresponding to the regulation of each gene by PU.1 or Gata1 so that $B_i = T_{i\leftarrow Spi1} \cdot \beta_{Spi1}$ or $B_i = T_{i\leftarrow Gata1} \cdot \beta_{Gata1}$ respectively. The proportionality constants β_{Spi1} and β_{Gata1} represent the additional amount of active PU.1 and Gata1 induced by OHT respectively. Similar to the knockdown experiments, the efficiency of activation achieved in the overexpression experiments was unknown and we chose the values of the proportionality constants so that simulated expression of 1 of 12 genes matched the observed expression. We did not however fit to the observed expression of the overexpressed gene since it stems from a mixture of mRNAs transcribed from the endogenous locus and the constitutively expressed ERT fusion gene. Instead we chose the values of the proportionality constants so the simulations matched the observed expression of *Gata1* in the PU.1ERT and *Spi1* in the GATA1ERT experiments respectively.

The simulations were carried out with $c^l = 0$ to simulate the progenitor condition since the experiments had been conducted in undifferentiated FDCP-mix cells. The simulations were compared to experimental data at equilibrium. The GRN was simulated for 1000 hours to allow the solution to reach equilibrium. The ratio of each gene's expression in the perturbed condition to its expression in the unperturbed condition was computed to determine the fold change predicted by the simulation. This was compared to the empirical fold change, computed as the ratio of gene expression in treated cells to gene expression in control cells [25].

4.2.6 The analysis of gene regulation dynamics

The contribution of individual regulators to the activation or repression of a target was determined by decomposing the total regulatory input $u = \sum_{j=1}^{N} T_{ij}x_j^l + b_i c^l + h_i$ into its individual terms. The contribution of regulator j to the regulation of gene i was determined by computing $T_{ij}x_j^l(t)$, where T_{ij} is the genetic interconnectivity of the two genes and $x_j^l(t)$ is the model solution for the mRNA concentration of gene j at time t and condition l. Since the mRNA concentrations vary in time, the relative contributions of the regulators to the activation or repression of any target also

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

vary in time. When the total regulatory input crosses 0, that is $u = 0$, the regulation-expression $S(u) = \frac{1}{2}\left(u/\sqrt{(u^2+1)}+1\right) = \frac{1}{2}$ and the mRNA is synthesized at half the maximum rate (Eq. 1). The time at which different genes achieve half-maximum expression was used to order their activation in time.

4.2.7 Visualization of Tusi *et al.*'s scRNA-Seq data

The expression of *Cebpa*, *Gfi1*, and *Spi1* in individual Kit$^+$ hematopoietic progenitors cells from mouse bone marrow [GEO GSE49991; 11] was visualized as follows. The cells were arranged in 2D space as a k-nearest-neighbor (knn) graph according to their pairwise distances in gene expression space [SPRING algorithm; 90]. The potential landscape and the probability of each cell to adopt a given fate were given by Population Balance Analysis [PBA; see 70, for details]. Genome-wide normalized gene expression counts, the PBA potential, the PBA lineage probability, and the 2D SPRING coordinates of each cell were obtained from `https://kleintools.hms.harvard.edu/paper\_websites/tusi\_et\_al/`.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

4.3 QUANTIFICATION AND STATISTICAL ANALYSIS

4.3.1 Significance of fits

The optimization problem for the 12-gene circuit is overdetermined, having 720 data points and 192 free parameters, and the risk of overfitting is minimal. Nevertheless, we checked whether the model fits captured temporal patterns inherent in the data or whether the degrees of freedom were so numerous that the model could fit randomized non-biological data equally well. We randomized the data in a manner that preserved the dynamic range of the real data while creating non-biological temporal expression patterns and tested the ability of gene circuits to fit the latter compared to the former. For each gene, we created chimerical temporal expression patterns by combining erythrocyte training data up to the 96 hour time point with neutrophil data at later time points and vice versa. In each synthetic dataset, 10 of 12 genes were given chimerical expression patterns while the other two retained the original training data. 66 such synthetic datasets were generated for each combination of 10 genes (Algorithm 1). 10 gene circuits were trained per dataset resulting in a total of 660 gene circuits. The RMS scores of the resultant gene circuits were compared to the 100 gene circuits trained on the real data. The statistical significance of the differences between the RMS scores of gene circuits trained on random and real data was determined using the Wilcoxon rank sum test with continuity

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Algorithm 1 Generation of synthetic chimerical datasets

C_i: combination i for 10 out of 12 genes
c_i: 2 genes not included in C_i
$time_points$: 30 sampled differentiation time points, $time_points \in \{0, 2, 4, ..., 96, 120, 168\}$
$x_{ery}(i, g, t) \leftarrow$ gene expression in erythrocyte condition for gene combination i, gene g, and time point t
$x_{neu}(i, g, t) \leftarrow$ gene expression in neutrophil condition for gene combination i, gene g, and time point t
$x_{eryW}(i, g, t)$: empty $66 \times 12 \times 30$ array for storing swapped and normal gene expression in erythrocyte condition for gene combination i, gene g, and time point t
$x_{neuW}(i, g, t)$: empty $66 \times 12 \times 30$ array for storing swapped and normal gene expression in neutrophil condition for gene combination i, gene g, and time point t
$file_i$: output file for writing swapped and normal expression for gene combination i
for i in 1:$\binom{12}{10}$ **do**
  for gene g in C_i **do**
    for t in $time_points$ **do**
      if $t < 96$ **then**
        $x_{eryW}(i, g, t) \leftarrow x_{ery}(i, g, t)$
        $x_{neuW}(i, g, t) \leftarrow x_{neu}(i, g, t)$
      else
        $x_{eryW}(i, g, t) \leftarrow x_{neu}(i, g, t)$
        $x_{neuW}(i, g, t) \leftarrow x_{ery}(i, g, t)$
      end if
    end for
  end for
  for gene g in c_i **do**
    for t in $time_points$ **do**
      $x_{eryW}(i, g, t) \leftarrow x_{ery}(i, g, t)$
      $x_{neutW}(i, g, t) \leftarrow x_{neu}(i, g, t)$
    end for
  end for
  for gene g in (C_i and c_i) **do**
    for t in $time_points$ **do**
      WRITE($file_i$, $x_{eryW}(i, g, t)$, $x_{neuW}(i, g, t)$)
    end for
  end for
end for

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

4.4 KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
FDCP-mix cells microarray	[25]	GEO: GSE49991
Mice bone marrow scRNA-Seq	[11]	GEO: GSE49991
Estimated gene circuit parameters	This manuscript	Table S1
Software and Algorithms		
Gene circuit source code	`https://github.com/mlekkha/EryNeu`	
Matlab	https://www.mathworks.com/	v2020a
R	https://www.r-project.org/	v4.0.0
Perl	https://www.perl.org/	v5.16.3
Cytoscape	https://cytoscape.org/index.html	v3.2.1

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

5 Supplemental information

Figures S1-S2, Tables S1-S2.

References

[1] Orkin SH, Zon LI. Hematopoiesis: an evolving paradigm for stem cell biology. Cell. 2008 Feb;132(4):631–44.

[2] Laslo P, Pongubala JMR, Lancki DW, Singh H. Gene regulatory networks directing myeloid and lymphoid cell fates within the immune system. Semin Immunol. 2008 Aug;20(4):228–35.

[3] Laslo P, Spooner CJ, Warmflash A, Lancki DW, Lee HJ, Sciammas R, et al. Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. Cell. 2006 Aug;126(4):755–66.

[4] Huang S, Guo Y, May G, Enver T. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. Developmental Biology. 2007;305:695–713.

[5] Enver T, Pera M, Peterson C, Andrews PW. Stem cell states, fates, and the rules of attraction. Cell Stem Cell. 2009 May;4(5):387–97.

[6] Graf T, Enver T. Forcing cells to change lineages. Nature. 2009 Dec;462(7273):587–94.

[7] Novershtern N, Subramanian A, Lawton LN, Mak RH, Haining WN, McConkey ME, et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell. 2011 Jan;144(2):296–309.

[8] Wilson NK, Foster SD, Wang X, Knezevic K, Schütte J, Kaimakis P, et al. Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. Cell Stem Cell. 2010 Oct;7(4):532–44.

[9] Nègre N, Brown CD, Ma L, Bristow CA, Miller SW, Wagner U, et al. A cis-regulatory map of the Drosophila genome. Nature. 2011 Mar;471(7339):527–31.

[10] 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 Apr;19(4):271–281.

[11] Tusi BK, Wolock SL, Weinreb C, Hwang Y, Hidalgo D, Zilionis R, et al. Population snapshots predict early haematopoietic and erythroid hierarchies. Nature. 2018 Mar;555(7694):54–60. 29466336[pmid]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29466336.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[12] Hoppe PS, Schwarzfischer M, Loeffler D, Kokkaliaris KD, Hilsenbeck O, Moritz N, et al. Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. Nature. 2016 Jul;535(7611):299–302. Available from: https://doi.org/10.1038/nature18320.

[13] Mossadegh-Keller N, Sarrazin S, Kandalla PK, Espinosa L, Stanley ER, Nutt SL, et al. M-CSF instructs myeloid lineage fate in single haematopoietic stem cells. Nature. 2013 May;497(7448):239–43.

[14] Rieger MA, Hoppe PS, Smejkal BM, Eitelhuber AC, Schroeder T. Hematopoietic cytokines can instruct lineage choice. Science. 2009 Jul;325(5937):217–8.

[15] Reinitz J, Sharp DH. Mechanism of *eve* stripe formation. Mechanisms of Development. 1995;49:133–158.

[16] Fehr David A JE Handzlik, Manu, Loh YL. Classification-Based Inference of Dynamical Models of Gene Regulatory Networks. G3: Genes, Genomes, Genetics. 2019;9(12):4183–4195. Available from: https://www.g3journal.org/content/9/12/4183.

[17] Chu KW, Deng Y, Reinitz J. Parallel simulated annealing by mixing of states. The Journal of Computational Physics. 1999;148:646–662.

[18] Kozlov K, Surkova S, Myasnikova E, Reinitz J, Samsonova M. Modeling of gap gene expression in Drosophila Kruppel mutants. PLoS Comput Biol. 2012;8(8):e1002635.

[19] Gursky VV, Jaeger J, Kozlov KN, Reinitz J, Samsonova AM. Pattern formation and nuclear divisions are uncoupled in *Drosophila* segmentation: comparison of spatially discrete and continuous models. Physica D. 2004;197:286–302.

[20] Abdol AM, Cicin-Sain D, Kaandorp JA, Crombach A. Scatter Search Applied to the Inference of a Development Gene Network. Computation. 2017;5(2). Available from: https://www.mdpi.com/2079-3197/5/2/22.

[21] Jaeger J, Surkova S, Blagov M, Janssens H, Kosman D, Kozlov KN, et al. Dynamic control of positional information in the early *Drosophila* embryo. Nature. 2004;430:368–371.

[22] Manu, Surkova S, Spirov AV, Gursky V, Janssens H, Kim A, et al. Canalization of gene expression in the *Drosophila* blastoderm by gap gene cross regulation. PLoS Biology. 2009;7:e1000049. Doi:10.371/journal.pbio.1000049.

[23] Manu, Surkova S, Spirov AV, Gursky V, Janssens H, Kim A, et al. Canalization of gene expression and domain shifts in the *Drosophila* blastoderm by dynamical attractors. PLoS Computational Biology. 2009;5:e1000303. Doi:10.1371/journal.pcbi.1000303.

[24] Wu H, Manu, Jiao R, Ma J. Temporal and spatial dynamics of scaling-specific features of a gene regulatory network in Drosophila. Nat Commun. 2015;6:10031.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[25] May G, Soneji S, Tipping AJ, Teles J, McGowan SJ, Wu M, et al. Dynamic analysis of gene expression and genome-wide transcription factor binding during lineage specification of multipotent progenitors. Cell Stem Cell. 2013 Dec;13(6):754–68.

[26] Cantor AB, Orkin SH. Transcriptional regulation of erythropoiesis: an affair involving multiple partners. Oncogene. 2002 May;21(21):3368–76.

[27] Doré LC, Chlon TM, Brown CD, White KP, Crispino JD. Chromatin occupancy analysis reveals genome-wide GATA factor switching during hematopoiesis. Blood. 2012 Apr;119(16):3724–33.

[28] Vicente C, Conchillo A, García-Sánchez MA, Odero MD. The role of the GATA2 transcription factor in normal and malignant hematopoiesis. Critical Reviews in Oncology/Hematology. 2012;82(1):1 – 17. Available from: http://www.sciencedirect.com/science/article/pii/S1040842811001168.

[29] Shivdasani RA, Mayer EL, Orkin SH. Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. Nature. 1995 Feb;373(6513):432–4.

[30] Mikkola HKA, Klintman J, Yang H, Hock H, Schlaeger TM, Fujiwara Y, et al. Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene. Nature. 2003 Jan;421(6922):547–51.

[31] Huang Z, Dore LC, Li Z, Orkin SH, Feng G, Lin S, et al. GATA-2 reinforces megakaryocyte development in the absence of GATA-1. Mol Cell Biol. 2009 Sep;29(18):5168–80.

[32] Cantor AB, Orkin SH. Hematopoietic development: a balancing act. Curr Opin Genet Dev. 2001 Oct;11(5):513–9.

[33] Starck J, Cohet N, Gonnet C, Sarrazin S, Doubeikovskaia Z, Doubeikovski A, et al. Functional cross-antagonism between transcription factors FLI-1 and EKLF. Mol Cell Biol. 2003 Feb;23(4):1390–402.

[34] Stachura DL, Chou ST, Weiss MJ. Early block to erythromegakaryocytic development conferred by loss of transcription factor GATA-1. Blood. 2006 Jan;107(1):87–97.

[35] Porcher C, Swat W, Rockwell K, Fujiwara Y, Alt FW, Orkin SH. The T Cell Leukemia Oncoprotein SCL/tal-1 Is Essential for Development of All Hematopoietic Lineages. Cell. 1996;86(1):47 – 57. Available from: http://www.sciencedirect.com/science/article/pii/S0092867400800768.

[36] Mancini E, Sanjuan-Pla A, Luciani L, Moore S, Grover A, Zay A, et al. FOG-1 and GATA-1 act sequentially to specify definitive megakaryocytic and erythroid progenitors. The EMBO Journal. 2012;31(2):351–365. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1038/emboj.2011.390.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[37] Siatecka M, Bieker JJ. The multifunctional role of EKLF/KLF1 during erythropoiesis. Blood. 2011;118(8):2044–2054. Available from: http://www.bloodjournal.org/content/118/8/2044.

[38] Scott EW, Simon MC, Anastasi J, Singh H. Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. Science. 1994 Sep;265(5178):1573–7.

[39] Zhang DE, Zhang P, Wang Nd, Hetherington CJ, Darlington GJ, Tenen DG. Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein Îś-deficient mice. Proceedings of the National Academy of Sciences. 1997;94(2):569–574. Available from: https://www.pnas.org/content/94/2/569.

[40] Tian SS, Tapley P, Sincich C, Stein RB, Rosen J, Lamb P. Multiple signaling pathways induced by granulocyte colony-stimulating factor involving activation of JAKs, STAT5, and/or STAT3 are required for regulation of three distinct classes of immediate early genes. Blood. 1996 Dec;88(12):4435–44.

[41] Palani S, Sarkar CA. Positive receptor feedback during lineage commitment can generate ultrasensitivity to ligand and confer robustness to a bistable switch. Biophys J. 2008 Aug;95(4):1575–89.

[42] Robb L. Cytokine receptors and hematopoietic differentiation. Oncogene. 2007 Oct;26(47):6715–23.

[43] Walsh JC, DeKoter RP, Lee HJ, Smith ED, Lancki DW, Gurish MF, et al. Cooperative and Antagonistic Interplay between PU.1 and GATA-2 in the Specification of Myeloid Cell Fates. Immunity. 2002 Nov;17(5):665–676. Available from: https://doi.org/10.1016/S1074-7613(02)00452-1.

[44] Dahl R, Walsh JC, Lancki D, Laslo P, Iyer SR, Singh H, et al. Regulation of macrophage and neutrophil cell fates by the PU.1:C/EBPalpha ratio and granulocyte colony-stimulating factor. Nat Immunol. 2003 Oct;4(10):1029–36.

[45] Iwasaki-Arai J, Iwasaki H, Miyamoto T, Watanabe S, Akashi K. Enforced granulocyte/macrophage colony-stimulating factor signals do not support lymphopoiesis, but instruct lymphoid to myelomonocytic lineage conversion. J Exp Med. 2003 May;197(10):1311–22.

[46] Weiss MJ, Keller G, Orkin SH. Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. Genes & Development. 1994;8(10):1184–1197. Available from: http://genesdev.cshlp.org/content/8/10/1184.abstract.

[47] Kitajima K, Zheng J, Yen H, Sugiyama D, Nakano T. Multipotential differentiation ability of GATA-1-null erythroid-committed cells. Genes & Development. 2006;20(6):654–659. Available from: http://genesdev.cshlp.org/content/20/6/654.abstract.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[48] Cooper S, Guo H, Friedman AD. The +37 kb Cebpa Enhancer Is Critical for Cebpa Myeloid Gene Expression and Contains Functional Sites that Bind SCL, GATA2, C/EBPÎś, PU.1, and Additional Ets Factors. PLoS One. 2015;10(5):e0126385.

[49] Bertolino E, Reinitz J, Manu. The analysis of novel distal Cebpa enhancers and silencers using a transcriptional model reveals the complex regulatory logic of hematopoietic lineage specification. Dev Biol. 2016 May;413(1):128–44.

[50] Repele A, Krueger S, Bhattacharyya T, Tuineau MY, Manu. The regulatory control of Cebpa enhancers and silencers in the myeloid and red-blood cell lineages. PLoS One. 2019;14(6):e0217580.

[51] Wilson NK, Timms RT, Kinston SJ, Cheng YH, Oram SH, Landry JR, et al. Gfi1 expression is controlled by five distinct regulatory regions spread over 100 kilobases, with Scl/Tal1, Gata2, PU.1, Erg, Meis1, and Runx1 acting as upstream regulators in early hematopoietic cells. Mol Cell Biol. 2010 Aug;30(15):3853–63.

[52] Moignard V, Macaulay IC, Swiers G, Buettner F, SchΓÎŇtte J, Calero-Nieto FJ, et al. Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. Nature Cell Biology. 2013 Apr;15(4):363–372. Available from: https://doi.org/10.1038/ncb2709.

[53] van der Meer LT, Jansen JH, van der Reijden BA. Gfi1 and Gfi1b: key regulators of hematopoiesis. Leukemia. 2010 Nov;24(11):1834–1843. Available from: https://doi.org/10.1038/leu.2010.195.

[54] Khanna-Gupta A, Zibello T, Idone V, Sun H, Lekstrom-Himes J, Berliner N. Human neutrophil collagenase expression is C/EBP-dependent during myeloid development. Experimental Hematology. 2005;33(1):42–52. Available from: https://www.sciencedirect.com/science/article/pii/S0301472X04003236.

[55] Hock H, Hamblen MJ, Rooke HM, Traver D, Bronson RT, Cameron S, et al. Intrinsic Requirement for Zinc Finger Transcription Factor Gfi-1 in Neutrophil Differentiation. Immunity. 2003 Jan;18(1):109–120. Available from: https://doi.org/10.1016/S1074-7613(02)00501-0.

[56] Spooner CJ, Cheng JX, Pujadas E, Laslo P, Singh H. A recurrent network involving the transcription factors PU.1 and Gfi1 orchestrates innate and adaptive immune cell fates. Immunity. 2009 Oct;31(4):576–86.

[57] Kim W, Klarmann KD, Keller JR. Gfi-1 regulates the erythroid transcription factor network through Id2 repression in murine hematopoietic progenitor cells. Blood. 2014 09;124(10):1586–1596. Available from: https://doi.org/10.1182/blood-2014-02-556522.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[58] Hock H, Hamblen MJ, Rooke HM, Traver D, Bronson RT, Cameron S, et al. Intrinsic Requirement for Zinc Finger Transcription Factor Gfi-1 in Neutrophil Differentiation. Immunity. 2003 Jan;18(1):109–120. Available from: https://doi.org/10.1016/S1074-7613(02)00501-0.

[59] Iwasaki H, Mizuno Si, Arinobu Y, Ozawa H, Mori Y, Shigematsu H, et al. The order of expression of transcription factors directs hierarchical specification of hematopoietic lineages. Genes Dev. 2006 Nov;20(21):3010–21.

[60] Chickarmane V, Enver T, Peterson C. Computational modeling of the hematopoietic erythroid-myeloid switch reveals insights into cooperativity, priming, and irreversibility. PLoS Comput Biol. 2009 Jan;5(1):e1000268.

[61] Paul F, Arkin Y, Giladi A, Jaitin DA, Kenigsberg E, Keren-Shaul H, et al. Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. Cell. 2015 Dec;163(7):1663–77.

[62] Hock H, Hamblen MJ, Rooke HM, Traver D, Bronson RT, Cameron S, et al. Intrinsic Requirement for Zinc Finger Transcription Factor Gfi-1 in Neutrophil Differentiation. Immunity. 2003 Jan;18(1):109–120. Available from: https://doi.org/10.1016/S1074-7613(02)00501-0.

[63] Dahl R, Walsh JC, Lancki D, Laslo P, Iyer SR, Singh H, et al. Regulation of macrophage and neutrophil cell fates by the PU.1 C/EBPa ratio and granulocyte colony-stimulating factor. Nature Immunology. 2003 Oct;4(10):1029–1036. Available from: https://doi.org/10.1038/ni973.

[64] Leddin M, Perrod C, Hoogenkamp M, Ghani S, Assi S, Heinz S, et al. Two distinct autoregulatory loops operate at the PU.1 locus in B cells and myeloid cells. Blood. 2011 Mar;117(10):2827–38.

[65] Ma O, Hong S, Guo H, Ghiaur G, Friedman AD. Granulopoiesis Requires Increased C/EBPα Compared to Monopoiesis, Correlated with Elevated Cebpa in Immature G-CSF Receptor versus M-CSF Receptor Expressing Cells. PLOS ONE. 2014 04;9(4):1–14. Available from: https://doi.org/10.1371/journal.pone.0095784.

[66] Legraverend C, Antonson P, Flodby P, Xanthopoulos KG. High level activity of the mouse CCAAT/enhancer binding protein (C/EBP alpha) gene promoter involves autoregulation and several ubiquitous transcription factors. Nucleic Acids Res. 1993 Apr;21(8):1735–42.

[67] Zhang DE, Zhang P, Wang ND, Hetherington CJ, Darlington GJ, Tenen DG. Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. Proc Natl Acad Sci U S A. 1997 Jan;94(2):569–74.

[68] Smith L, Hohaus S, Gonzalez D, Dziennis S, Tenen D. PU.1 (Spi-1) and C/EBP alpha regulate the granulocyte colony- stimulating factor receptor promoter in myeloid cells.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Blood. 1996;88(4):1234–1247. Available from: http://www.bloodjournal.org/content/88/4/1234.

[69] Li H, Ji M, Klarmann KD, Keller JR. Repression of Id2 expression by Gfi-1 is required for B-cell and myeloid development. Blood. 2010 Aug;116(7):1060–9.

[70] Weinreb C, Wolock S, Tusi BK, Socolovsky M, Klein AM. Fundamental limits on dynamic inference from single-cell snapshots. Proceedings of the National Academy of Sciences. 2018;115(10):E2467–E2476. Available from: https://www.pnas.org/content/115/10/E2467.

[71] Li C, Wang J. Quantifying Cell Fate Decisions for Differentiation and Reprogramming of a Human Stem Cell Network: Landscape and Biological Paths. PLoS Comput Biol. 2013 08;9(8):e1003165 EP –. Available from: http://dx.doi.org/10.1371%2Fjournal.pcbi.1003165.

[72] Hong T, Xing J, Li L, Tyson JJ. A simple theoretical framework for understanding heterogeneous differentiation of CD4+ T cells. BMC Syst Biol. 2012 Jun;6:66.

[73] 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. Bioinformatics. 2013 Jul;29(13):i80–8.

[74] Collombet S, van Oevelen C, Sardina Ortega JL, Abou-Jaoudé W, Di Stefano B, Thomas-Chollier M, et al. Logical modeling of lymphoid and myeloid cell specification and transdifferentiation. Proc Natl Acad Sci U S A. 2017 06;114(23):5792–5799.

[75] Zhang P, Zhang X, Iwama A, Yu C, Smith KA, Mueller BU, et al. PU.1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding. Blood. 2000 Oct;96(8):2641–8.

[76] Nerlov C, Querfurth E, Kulessa H, Graf T. GATA-1 interacts with the myeloid PU.1 transcription factor and represses PU.1-dependent transcription. Blood. 2000 Apr;95(8):2543–51.

[77] Scott L, Civin C, Rorth P, Friedman A. A novel temporal expression pattern of three C/EBP family members in differentiating myelomonocytic cells. Blood. 1992 10;80(7):1725–1735. Available from: https://doi.org/10.1182/blood.V80.7.1725.1725.

[78] Jack GD, Zhang L, Friedman AD. M-CSF elevates c-Fos and phospho-C/EBPalpha(S21) via ERK whereas G-CSF stimulates SHP2 phosphorylation in marrow progenitors to contribute to myeloid lineage specification. Blood. 2009 Sep;114(10):2172–80.

[79] Hu M, Krause D, Greaves M, Sharkis S, Dexter M, Heyworth C, et al. Multilineage gene expression precedes commitment in the hemopoietic system. Genes Dev. 1997 Mar;11(6):774–85.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[80] Yoshida T, Ng SYM, Georgopoulos K. Awakening lineage potential by Ikaros-mediated transcriptional priming. Curr Opin Immunol. 2010 Apr;22(2):154–60.

[81] Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature. 2008 May;453(7194):544–7.

[82] Zhao W, Kitidis C, Fleming MD, Lodish HF, Ghaffari S. Erythropoietin stimulates phosphorylation and activation of GATA-1 via the PI3-kinase/AKT signaling pathway. Blood. 2006 Feb;107(3):907–15.

[83] Rogers HM, Yu X, Wen J, Smith R, Fibach E, Noguchi CT. Hypoxia alters progression of the erythroid program. Experimental Hematology. 2008 Jan;36(1):17–27. Available from: https://doi.org/10.1016/j.exphem.2007.08.014.

[84] Deindl P, Klar M, Drews D, Cremer M, Gammella E, Gassmann M, et al. Mice overexpressing human erythropoietin indicate that erythropoietin enhances expression of its receptor via up-regulated Gata1 and Tal1. Haematologica. 2014 Oct;99(10):e205–e207. Available from: https://haematologica.org/article/view/7185.

[85] Chiba T, Nagata Y, Kishi A, Sakamaki K, Miyajima A, Yamamoto M, et al. Induction of erythroid-specific gene expression in lymphoid cells. Proceedings of the National Academy of Sciences. 1993;90(24):11593–11597. Available from: https://www.pnas.org/content/90/24/11593.

[86] Rogers H, Wang L, Yu X, Alnaeeli M, Cui K, Zhao K, et al. T-cell Acute Leukemia 1 (TAL1) Regulation of Erythropoietin Receptor and Association with Excessive Erythrocytosis. Journal of Biological Chemistry. 2012 Oct;287(44):36720–36731. Available from: https://doi.org/10.1074/jbc.M112.378398.

[87] Hirsch MW, Smale S, Devaney R. Differential Equations, Dynamical Systems, and an Introduction to Chaos. Boston: Academic Press; 2004.

[88] Surkova S, Kosman D, Kozlov K, Manu, Myasnikova E, Samsonova A, et al. Characterization of the *Drosophila* Segment Determination Morphome. Developmental Biology. 2008;313(2):844–862.

[89] Muir P, Li S, Lou S, Wang D, Spakowicz DJ, Salichos L, et al. The real cost of sequencing: scaling computation to keep pace with data generation. Genome Biol. 2016 Mar;17:53.

[90] Weinreb C, Wolock S, Klein AM. SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. Bioinformatics. 2018 04;34(7):1246–1248.

[91] Lam J, Delosme JM. An efficient simulated annealing schedule: Derivation. New Haven, CT: Yale Electrical Engineering Department; 1988. 8816.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

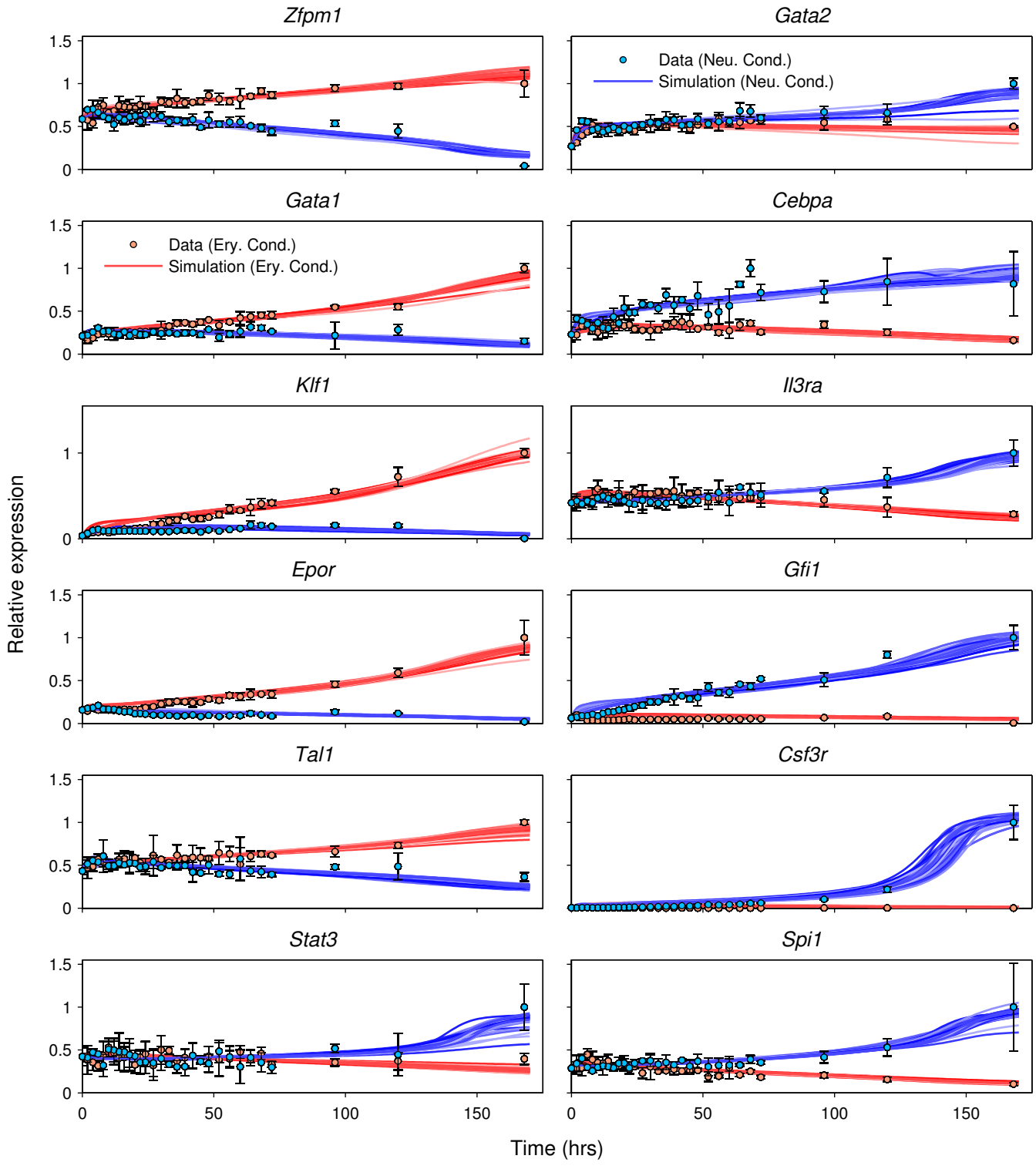

Figure 1:

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

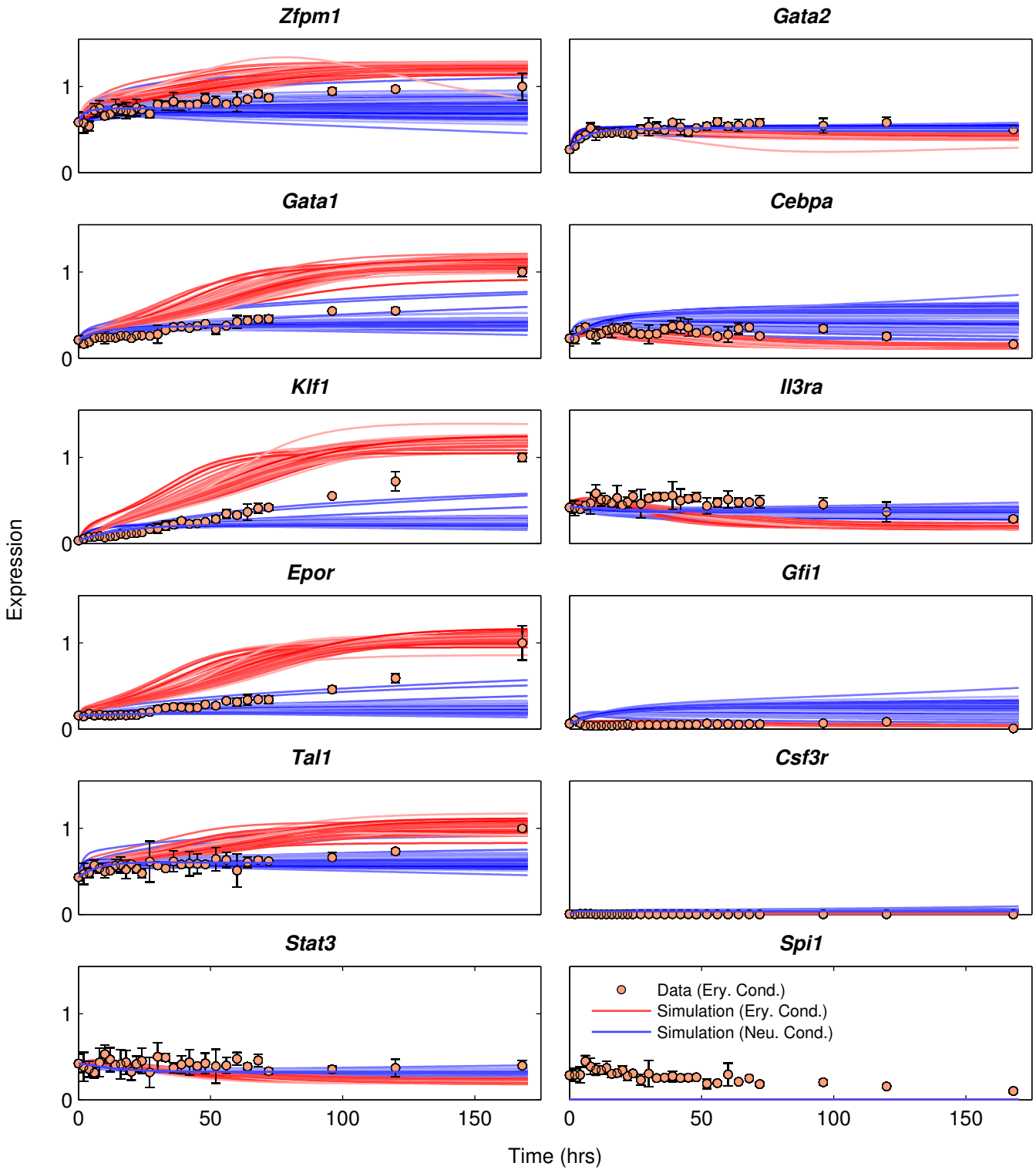

Figure 2:

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

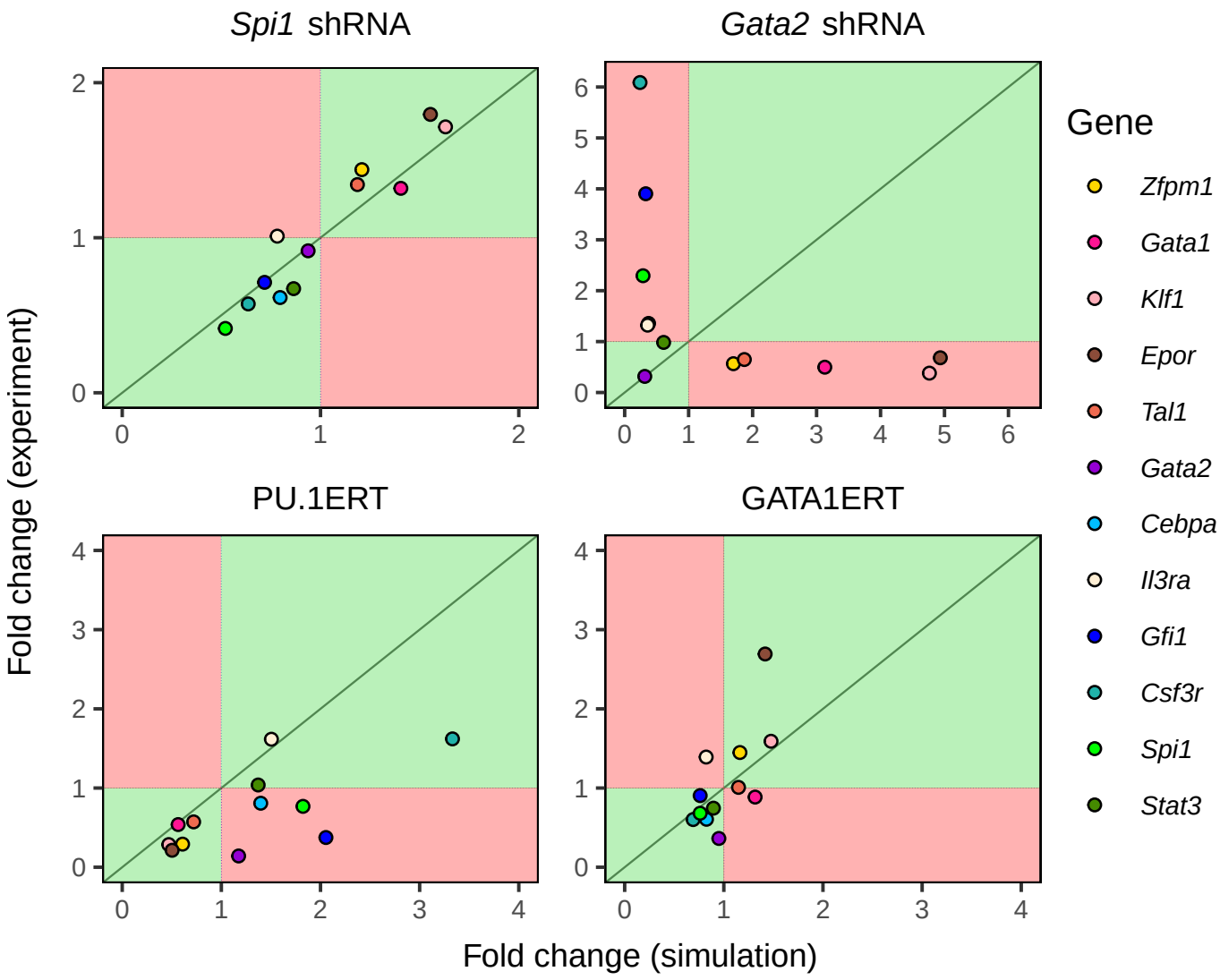

Figure 3:

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

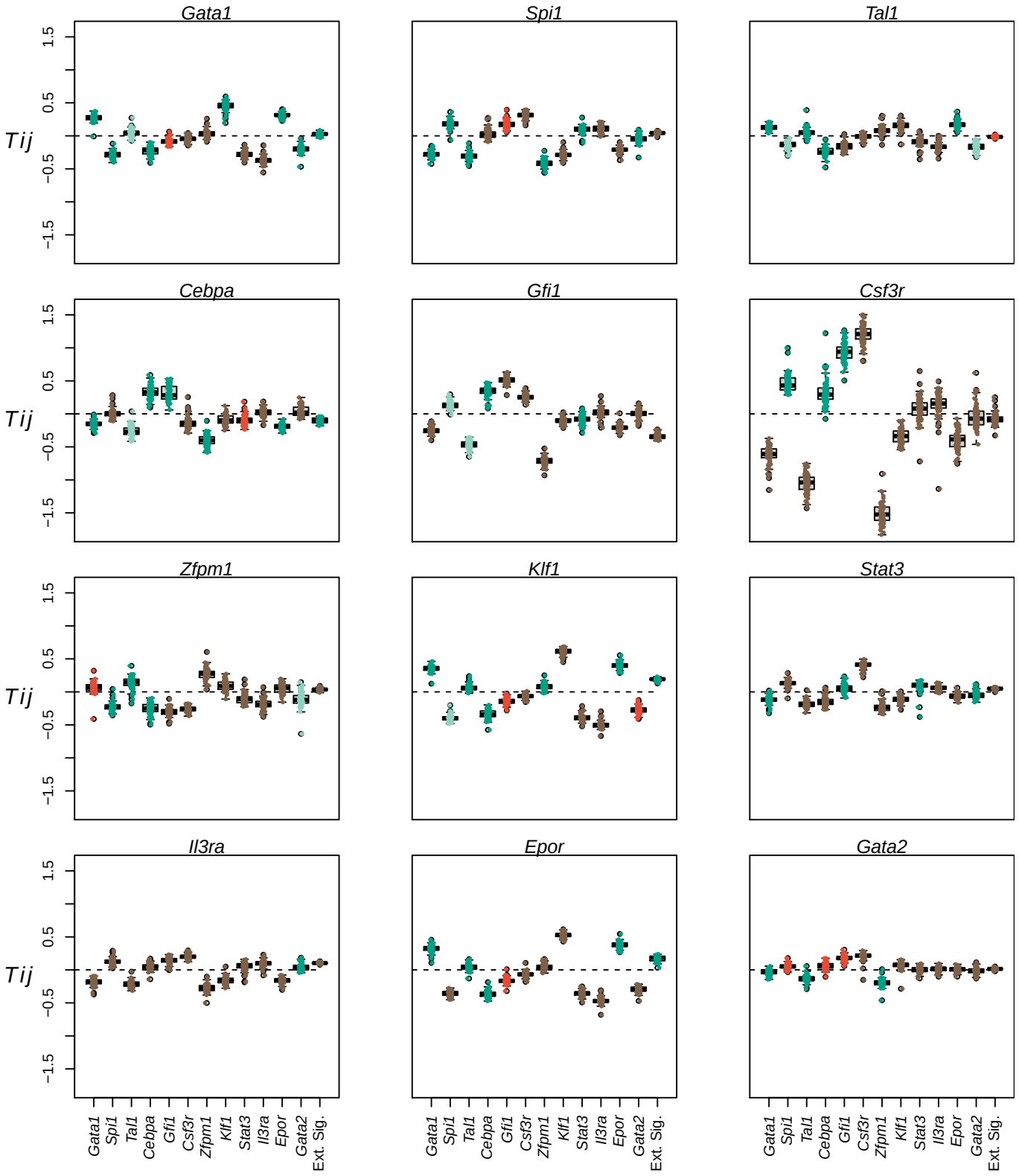

Figure 4:

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

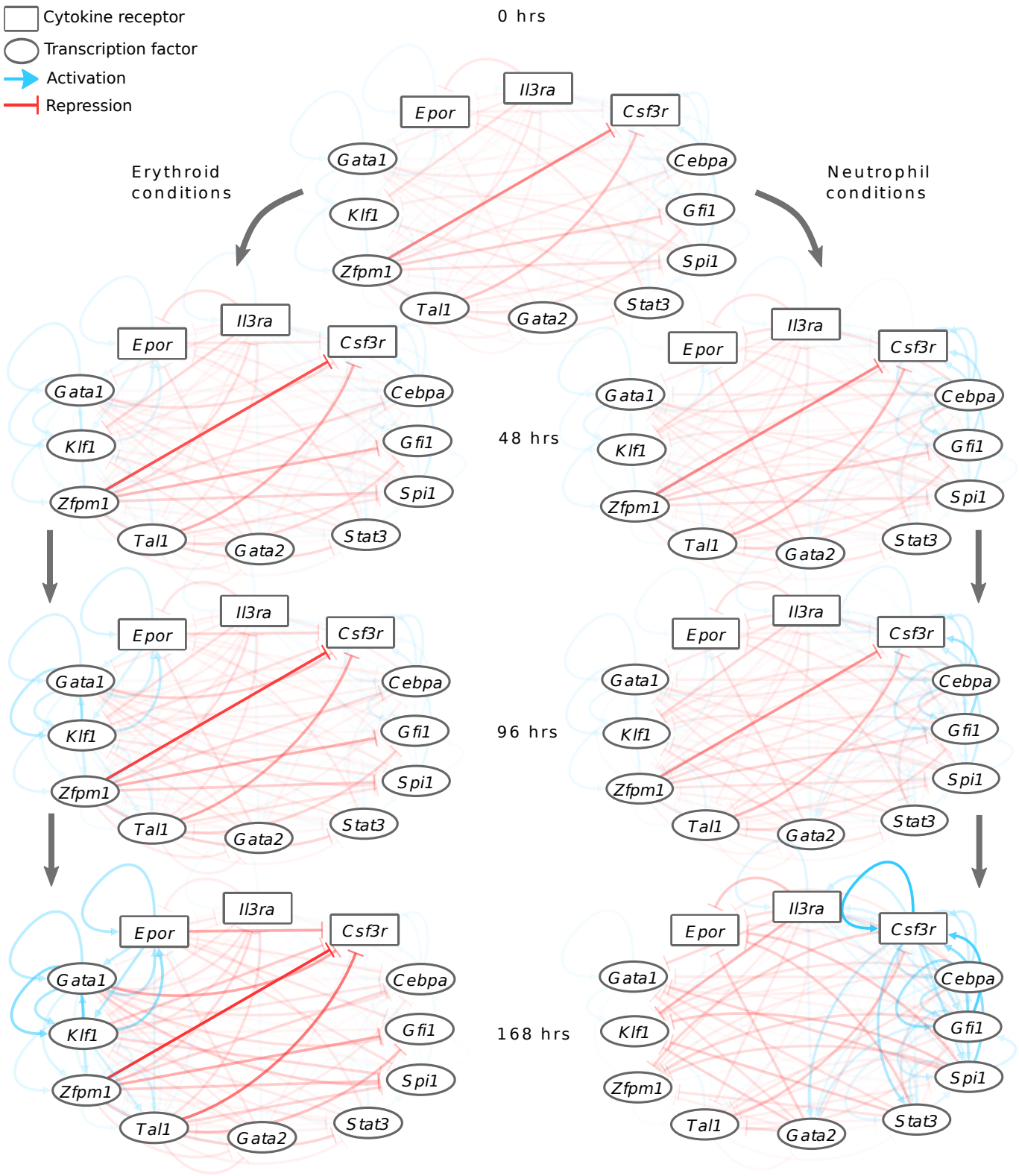

Figure 5:

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

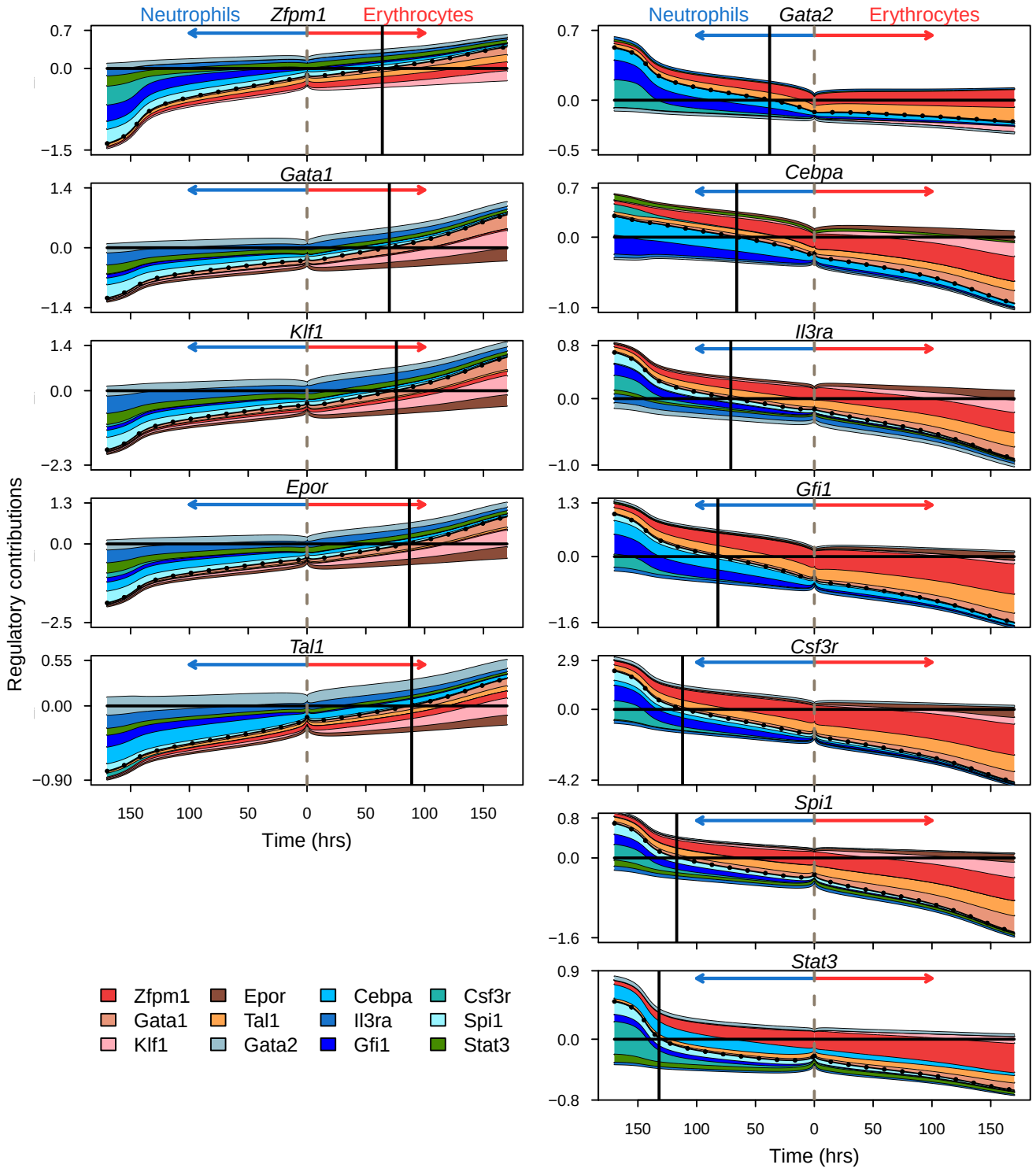

Figure 6:

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443501; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

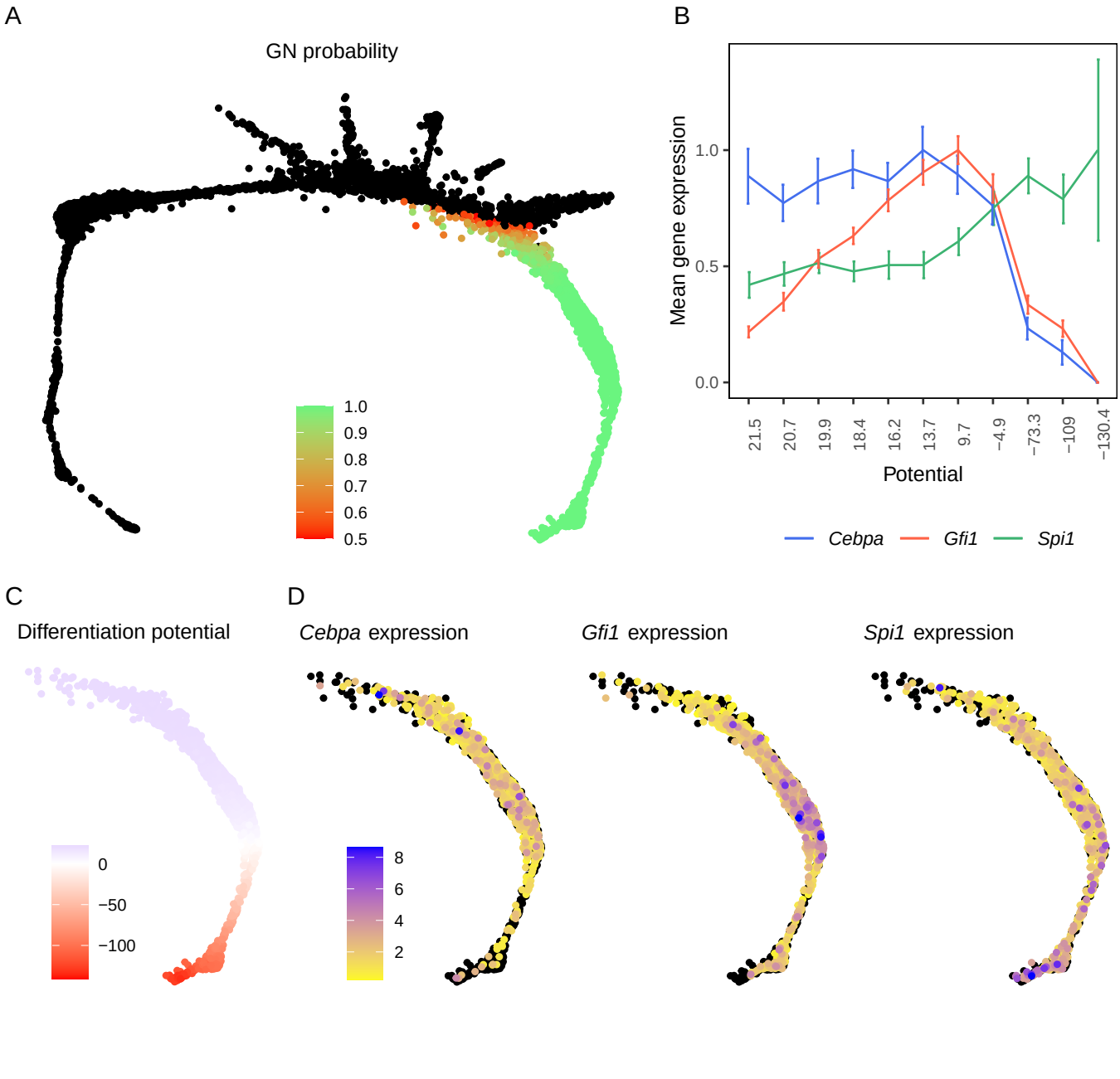

Figure 7: