Stem Cell Reports
Report

ISSCR

OPEN ACCESS

A Generalized Gene-Regulatory Network Model of Stem Cell Differentiation for Predicting Lineage Specifiers

Satoshi Okawa,[1] Sarah Nicklas,[2] Sascha Zickenrott,[1] Jens C. Schwamborn,[2] and Antonio del Sol[1,*]
[1]Computational Biology Group
[2]Developmental and Cellular Biology Group
Luxembourg Centre for Systems Biomedicine, University of Luxembourg, 7 Avenue des Hauts Fourneaux, 4362 Esch-sur-Alzette, Luxembourg
*Correspondence: antonio.delsol@uni.lu
http://dx.doi.org/10.1016/j.stemcr.2016.07.014

SUMMARY

Identification of cell-fate determinants for directing stem cell differentiation remains a challenge. Moreover, little is known about how cell-fate determinants are regulated in functionally important subnetworks in large gene-regulatory networks (i.e., GRN motifs). Here we propose a model of stem cell differentiation in which cell-fate determinants work synergistically to determine different cellular identities, and reside in a class of GRN motifs known as feedback loops. Based on this model, we develop a computational method that can systematically predict cell-fate determinants and their GRN motifs. The method was able to recapitulate experimentally validated cell-fate determinants, and validation of two predicted cell-fate determinants confirmed that overexpression of ESR1 and RUNX2 in mouse neural stem cells induces neuronal and astrocyte differentiation, respectively. Thus, the presented GRN-based model of stem cell differentiation and computational method can guide differentiation experiments in stem cell research and regenerative medicine.

INTRODUCTION

Cellular phenotypes are characterized by stable gene-expression states determined by underlying gene-regulatory networks (GRNs), particularly by subnetworks that appear frequently and are functionally important (i.e., GRN motifs). A classical GRN motif, the toggle switch, constitutes a molecular mechanism that determines cell-fate decisions, and provides stability to transcriptional programs of binary cell-fate choices. Overexpression of each transcription factor (TF) corresponds to one of the two mutually exclusive cell fates, whereas a "balanced" expression of both TFs maintains the stem/progenitor state (Huang et al., 2007; Jacob and Monod, 1961; Roeder and Glauche, 2006). The toggle switch has been experimentally shown to play an important role in binary cell-fate control of stem/progenitor cells (Graf, 2002; Lin et al., 2008; Ralston and Rossant, 2005). A well-known example is the one consisting of an erythroid determinant *Gata1* (Pevny et al., 1991) and a myeloid determinant *Spi1* (Voso et al., 1994) in the hematopoietic stem cell (HSC) system.

Interestingly, a different GRN motif has been more recently proposed for explaining mesendodermal and ectodermal specification of mouse embryonic stem cells (mESCs) (Shu et al., 2013). In this motif, the balanced expression (i.e., similar expression levels) of a mesendodermal and an ectodermal cell-fate determinant, POU5F1 (Niwa et al., 2000; Zeineddine et al., 2006) and SOX2 (Kopp et al., 2008), respectively, maintains the pluripotent state, whereas significant up- or downregulation of either of these genes induces differentiation into the respective lineage. Moreover, replacing POU5F1 with other mesendodermal determinants was able to induce reprogramming of fibroblasts to pluripotency in both mouse and human (Montserrat et al., 2013; Shu et al., 2013). These observations suggest that stem/progenitor cell states in general seem to be maintained by a balance between differentiation forces exerted by groups of opposing cell-fate determinants, and that the underlying GRN motifs do not necessarily comprise toggle switches. Indeed, a toggle switch belongs to a more general class of network motifs, known as feedback loops (Thomas, 1978; Siebert, 2009; Zañudo and Albert, 2013).

Taking these facts together, here we propose a computational model that generalizes binary-fate stem cell differentiation events (Figure 1), according to which stem/progenitor cells correspond to stable gene-expression states maintained by the balanced expression of cell-fate determinants residing in clusters of interconnected feedback loops (strongly connected components). Furthermore, these strongly connected components consist of differentially expressed TFs between two daughter cell types from the stem/progenitor cells, and stabilize the two stable gene-expression states corresponding to these two daughter cell types. Upregulated TFs in one of the daughter cells cooperate among themselves and compete with those upregulated in the other daughter cell.

Based on this model, we further propose a Boolean network-based computational method that can systematically predict cell-fate determinants and the GRN motifs to which these genes belong. This method is general, since it can be applied to any stem cell differentiation system for which gene-expression data of the stem/progenitor and the two daughter cells are available. We selected five

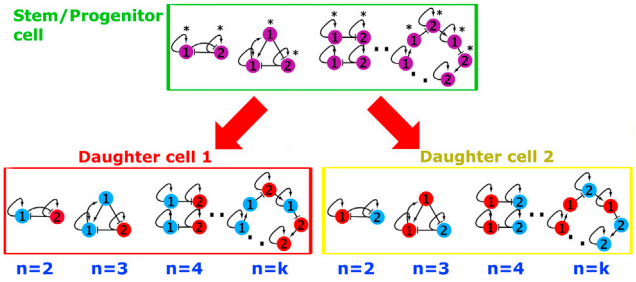


Figure 1. Proposed Model of Binary-Fate Stem Cell Differentiation Governed by GRN Motifs

In this model two different daughter cell types (daughter 1 and daughter 2) from a common stem/progenitor cell correspond to two stable steady states, which are stabilized by strongly connected components of any number of genes consisting of differentially expressed TFs between two daughter cells. The same strongly connected components are used for maintaining the stem/progenitor state, in which pair(s) of TFs exhibit a more balanced expression pattern in comparison with that in two daughter cells (indicated by asterisks). TFs that do not show this balanced expression pattern are still necessary for stabilizing the expression balance of TFs marked with asterisks. The classical toggle switch that consists of two TFs (n = 2) is the simplest case of this model. Red nodes are TFs upregulated in daughter 1. Blue nodes are TFs upregulated in daughter 2. Purple nodes indicate TF expression in the stem/progenitor cell. Pointed arrows indicate activation and blunted arrows indicate inhibition. Note that motifs shown in this figure are examples of each n. Motifs with different topologies (not shown) are possible.

stem cell systems to assess the validity of the method: mESCs, mHSCs, mouse neural stem cells (mNSCs), mouse mesenchymal stem cells (mMSCs), and hESCs differentiating into MESP1$^+$ or MESP1$^-$ (pre-)cardiac progenitor cells (hCPCs) (den Hartogh et al., 2015). Our predictions were able to recapitulate experimentally validated cell-fate determinants in these systems. In particular, the method predicted known cell-fate determinants in the hCPC system, where the differentiation is incomplete and phenotypic differences between the two daughter cells are relatively small. Finally, we experimentally validated predicted cell-fate determinants in the mNSC system, which confirmed that ESR1 and RUNX2 induce neuronal and astrocyte differentiation, respectively.

Thus, this study presents a general GRN-based computational model that can identify GRN motifs crucial for both maintenance and differentiation of stem/progenitor cells. From a systems biology point of view, identification of functionally important subnetworks is important for extracting biologically meaningful information from a large GRN. Finally, the method solely requires transcriptome data and literature knowledge of TF interactions, while not requiring prior knowledge of potential candidate genes; neither are pathways or gene ontology necessary. Therefore, our approach offers practical guidance to experiments in stem cell biology and regenerative medicine.

RESULTS

The Method Recapitulates Known Toggle Switches

The overview of our computational method is shown in Figures 2 and S4. In brief, a Boolean GRN among differentially expressed TFs between two daughter cell types is reconstructed using database knowledge and our network-pruning algorithm. In parallel, pairs of TFs whose expression patterns are significantly disrupted upon differentiation in comparison with the stem/progenitor cell are identified. In each of these significant TF pairs, if the two TFs are directly connected to each other in their respective most frequent strongly connected component, they are considered candidate cell-fate determinants (see Experimental Procedures for details). Application of this method was able to recapitulate the well-characterized toggle switches for the *Gata1-Spi1* pair (Graf, 2002) and the *Runx2-Pparg* pair (Lin et al., 2008) in the mHSC and mMSC systems, respectively (Figures 3B and 3E). Note that the statistical metric we devised in this study, the normalized ratio difference (NRD), was intended to identify pairs of TFs whose expression ratios showed a significant change in daughter cells in comparison with the stem/progenitor cells. We suggest that the NRD is biologically more relevant than the absolute ratio within each cell type, since the basal/effective level of expression differs among different TFs. Indeed, the expression values of the two TFs in well-known pairs, such as *Gata1-Spi1* and *Runx2-Pparg*, were very different in the progenitor cells (Table 1). The TF pairs with significant NRD (significant NRD TF pairs) in each system are listed in Table S1.

The Predicted GRN Motif Explains Previous Experimental Evidence in mESCs

It has been shown that induced pluripotent stem cells (iPSCs) could be derived by expressing KLF4 and POU5F1 in SOX2-expressing mouse neural progenitor cells (Blelloch et al., 2006), or in mouse embryonic fibroblasts in combination with small compounds that can substitute SOX2 (Shi et al., 2008). These previous observations suggest that KLF4 acts similarly to POU5F1 and antagonistically to SOX2. In addition, *PAX6* is a known ectoderm determinant in human (Zhang et al., 2010). Therefore, the mESC GRN motif predicted for the *Pou5f1-Sox2* pair in this study (consisting of *Pou5f1*, *Sox2*, *Klf4*, and *Pax6*) (Figure 3A and Table 1) can mechanistically explain these previous experimental observations. In addition, this motif resembles the one previously proposed (Shu et al., 2013), which consisted of two nodes representing *Pou5f1* and

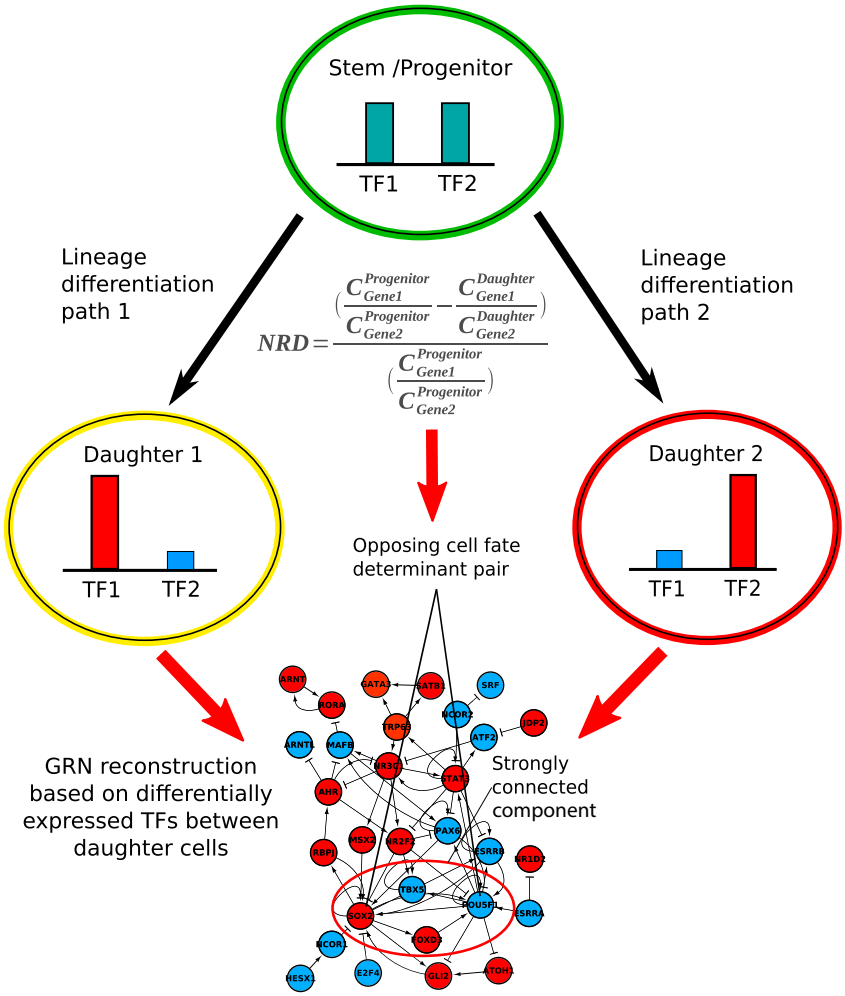


Figure 2. Schematic View of Proposed Method

Differentially expressed genes are computed between two daughter cells and Boolean GRNs are reconstructed from differentially expressed TFs by first retrieving literature-based interactions and then pruning this network by removing interactions incompatible with Booleanized gene-expression data of two daughter cells. In parallel, statistically significant NRD TF pairs are computed. Finally, for each significant NRD TF pair, the most frequent strongly connected component is identified among the best GRN solutions. If two paired TFs are directly connected to each other in that strongly connected component, they are considered predicted opposing cell-fate determinants together with their GRN motif.

Sox2 and two hyper-nodes (i.e., collections of unknown genes) representing the ectoderm and mesoderm, further supporting our proposed general differentiation model. Importantly, our method does not use hyper-nodes, so it can explicitly describe key interactions among cell-fate determinants that collectively maintain different cellular identities.

The Method Predicted Known Lineage Specifiers Even when Differentiation Is Incomplete

The method was also applied to the dataset of CPC differentiation, in which hESCs were differentiated into MESP1$^+$ and MESP1$^-$ (pre-)CPCs (den Hartogh et al., 2015), where only the former was able to differentiate further into cardiomyocytes. Our predictions of cell-fate determinants for the MESP1$^+$ lineage included *GATA4*, a well-known inducer of cardiac differentiation (Kuo et al., 1997) (Figure 3F), and *MYC*, which has recently been shown to play a critical role in long-term expansion of CPCs (Birket et al., 2015) (Table 1). Hence, our method was able to predict important cardiac cell-fate determinants even when the differentiation is not terminal and the two daughter cell types are close to each other. This aspect of the method can be useful when differentiation into a not well-defined particular subtype of a cell lineage is desired.

Deterministic Continuous Simulation Reproduces Expected Differentiation Dynamics upon Perturbations

An advantage of our method is that it is solely based on a simple Boolean network model for its predictions. However, because the Boolean model might oversimplify the quantitative nature of real biology, such as the inability to represent an intermediate steady state for stem/progenitor cells, we next investigated whether predicted GRN motifs could capture expected cell-fate decisions in a more realistic continuous model (see Supplemental Experimental Procedures for details). The mHSC system was used for this purpose, as it is the most well studied both experimentally and theoretically. The result indicates that our

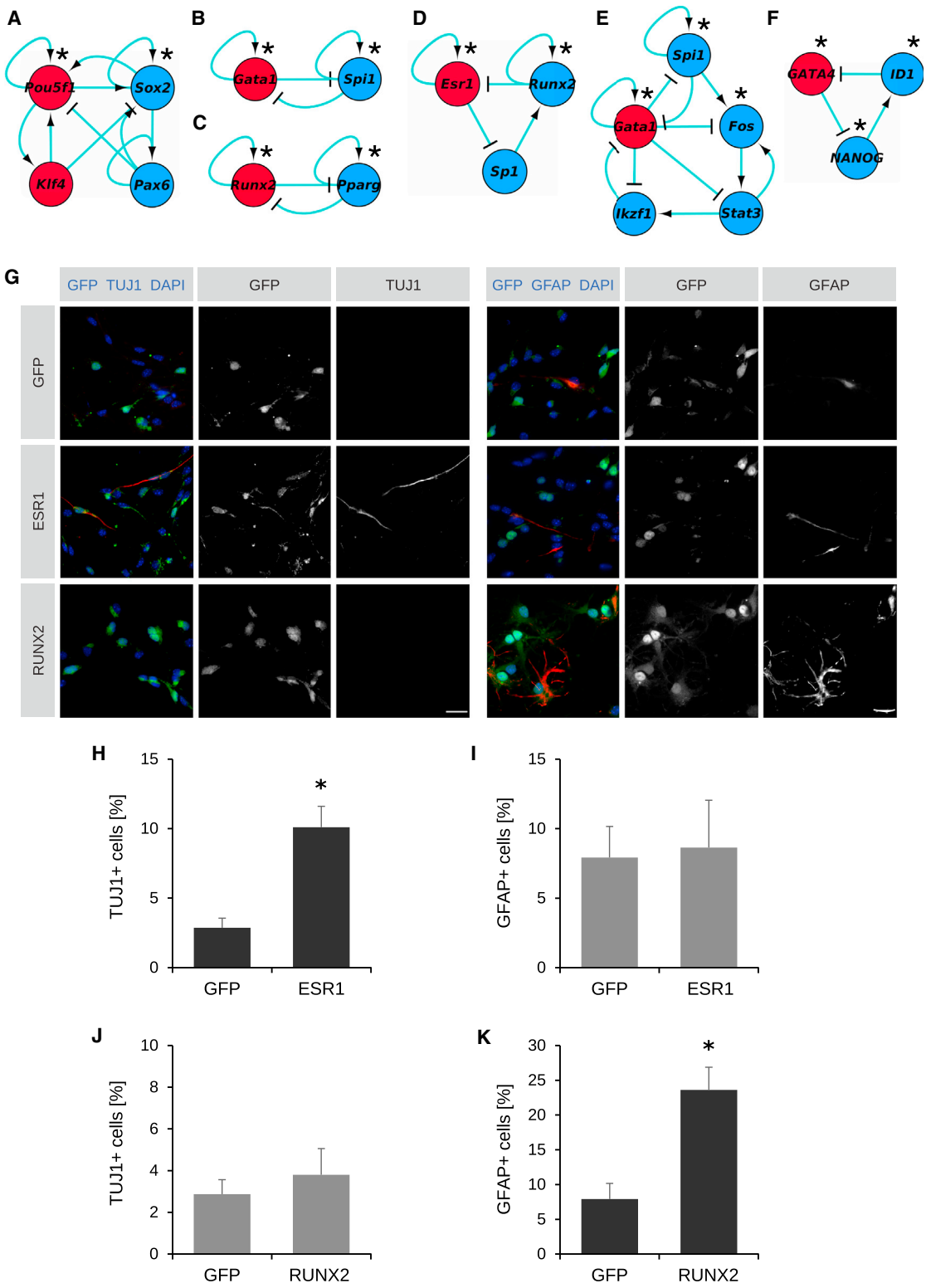


Figure 3. Predicted Opposing Cell-Fate Determinant Pairs, Their GRN Motifs, and Their Experimental Validation in mNSCs

(A–F) Red nodes are TFs upregulated in daughter 1 (mesoderm, erythroid, neuron, osteoblast, and MESP1^{+} CPC). Blue nodes are TFs upregulated in daughter 2 (ectoderm, myeloid, astrocyte, adipocyte and MESP1^{-} CPC). Pointed arrows indicate activation, blunted arrows

(legend continued on next page)

Table 1. Predicted Opposing Cell-Fate Determinant Pairs, Their Minimum Out-Degree Interface, Number of TFs in Their GRN Motifs, Microarray Expression Values, and Figure Locations

Best Candidate Opposing Cell-Fate Determinant Pair	Minimum Out-Degree Interface	No. of TFs in Strongly Connected Component	Log_2(Expression) Value in Progenitor	Log_2(Expression) Value in Daughter 1	Log_2(Expression) Value in Daughter 2	Figure
1. mESC						
Pou5f1-Sox2	8	4	10.63__11.05	4.26__3.13	2.58__7.85	3A
Sox17-Sox2	20	8	9.75__11.05	0.74__7.85	4.59__3.13	S1A
2. mHSC						
Gata1-Fos	16	5	12.57__7.12	14.87__6.24	12.21__10.75	3E
Gata1-Cebpa	16	4	12.57__10.44	14.87__10.03	12.21__12.37	S1B
Gata1-Spi1	12	2	12.57__8.26	14.87__6.77	12.21__10.04	3B
Gata1-Gata2	10	5	12.57__14.79	14.87__15.69	12.21__16.88	S1C
Cux1-Irf1	4	8	10.40__13.38	10.56__11.95	9.51__14.64	S1D
3. mNSC						
Esr1-Runx2	6	3	4.03__3.74	7.24__5.20	4.97__7.30	3D
Esr1-Stat5a	5	5	4.03__4.74	7.24__5.37	4.97__6.70	S1E
Mef2c-Hey1	1	6	6.82__5.67	13.14__9.52	8.67__11.05	S1F
4. mMSC						
Runx2-Pparg	4	2	5.27__7.11	10.13__4.25	5.25__11.96	3C
5. hCPC						
MYC-PBX1	6	4	10.11__9.03	10.24__7.48	9.48__9.02	S1G
MYC-NANOG	6	4	10.11__8.93	10.24__7.07	9.48__9.29	S1G
GATA4-NANOG	3	2	6.70__8.93	8.25__7.07	7.45__9.29	3F
GATA4-ID1	2	3	6.70__12.48	8.25__10.06	7.45__12.28	3F

Columns 4–6 indicate normalized log_2 microarray expression values. "Progenitor" indicates hESCs, mHSCs, mNSCs, mMSCs, and hESCs, "daughter 1" indicates ectoderms, erythroids, neurons, osteoblasts, and MESP1$^+$ CPCs, and "daughter 2" indicates mesoderms, myeloids, astrocytes, adipocytes and MESP1$^-$ CPCs, respectively. Left TF and right TF in each pair are predicted cell-fate determinants for daughter 1 and daughter 2, respectively, when overexpressed. See Experimental Procedures for the definition of minimum out-degree interface.

continuous simulation recapitulated the known dynamics of the *Gata1-Spi1* toggle switch (Figure 3B), in which the progenitor state remained stable over time but reached the erythroid state when either GATA1 was upregulated or SPI1 was downregulated (Huang et al., 2007) (Figure S2). On the other hand, the opposite myeloid differentiation dynamics was also reproduced (Figure S2). In addition, the more complex five-gene motif for the *Gata1-Fos* pair, which includes the *Gata1-Spi1* toggle switch (Figure 3E), also exhibited the tristability corresponding to the three cell types, and reached the expected erythroid or myeloid state upon appropriate perturbation of any gene in the

indicate inhibition. Asterisks indicate TFs that showed a significant NRD. GRN motifs of (A) *Pou5f1-Sox2* pair in mESCs, (B) *Gata1-Spi1* pair in mHSCs, (C) *Runx2-Pparg* pair in mMSCs, (D) *Esr1-Runx2* pair in mNSCs, (E) *Gata1-Fos* pairs in mHSCs, and (F) *GATA4-NANOG* and *GATA4-NANOG* pairs in hCPCs.

(G) Lineage marker (TUJ1 and glial fibrillary acidic protein [GFAP]) immunostaining of cells cultivated under maintenance conditions for 5 days after transduction with lentiviruses encoding GFP (negative control), ESR1, or RUNX2. Scale bar, 20 μm.

(H–K) Diagrams showing the percentage of TUJ1-positive (H, J) and GFAP-positive (I, K) cells transduced with lentiviruses encoding GFP, ESR1, or RUNX2 (mean ± SEM; $n \geq 420$ cells, N = 3 independent mNSC cultures; *$p < 0.05$, t test).

motif (Figure S3). Hence, this continuous simulation study demonstrates that, although our Boolean network-based method does not consider the intermediate stem/progenitor attractor state, the predicted GRN motif exhibited the appropriate differentiation dynamics upon perturbation of its genes.

ESR1 Induces Neuronal Differentiation and RUNX2 Induces Astrocyte Differentiation

To our knowledge, no well-defined cell-fate determinant pair and their GRN motif are known for the mNSC system. Our method predicted the *Esr1-Runx2* pair as the top candidate (Table 1), which stabilizes the three cell types via a three-gene GRN motif (Figure 3D). To validate this prediction, we performed the lentiviral transduction experiment under mNSC maintenance conditions (Conti et al., 2005). We used these conditions rather than differentiation-inducing conditions, since the latter will have a mixture of influences (i.e., both transduced TF and differentiation signals coming from the media) and it is difficult to draw clear conclusions under these conditions. The result confirmed a significant increase in the amount of neuron-specific class III β-tubulin (TUJ1)-positive cells upon overexpression of ESR1 in mNSCs (from 2.9% to 10%) (Figures 3G and 3H). Importantly, it did not induce astrocyte differentiation (Figures 3G and 3I). Conversely, overexpression of RUNX2 strongly induced astrocyte differentiation (from 8% to 24%) (Figures 3G and 3K), but this effect was restricted to the astrocyte lineage only (no increase in TUJ1-positive cells) (Figures 3G and 3J). These data demonstrate that the predicted function of TF pairs to induce lineage specifications can be validated experimentally, indicating the applicability of the method described here to stem cell differentiation experiments in general. Although our method has been shown to predict cell-fate determinants for different cell lineages, including neurons and astrocytes, it can also be applied for more specific cell subtypes, such as dopaminergic neurons or subventricular zone astrocytes. In these cases additional cell type specific markers would be required.

DISCUSSION

The interest in directed cell-fate determination in stem cell biology and regenerative medicine has been increasing over the years. However, due to the complexity of GRNs, identification of cell-fate determinants and their functionally important subnetworks (GRN motifs) that determine stem cell maintenance and differentiation remains a challenge. Indeed, there have been a few attempts to model cellular conversions by means of network biology (Crespo and del Sol, 2013; Cahan et al., 2014; Zañudo and Albert, 2015; Rackham et al., 2016; Jo et al., 2016) using bulk transcriptome data. These previous studies deal with transitions between two well-defined cell types, such as reprogramming. As a complementary approach to these studies, the present study provides a generalized network model of stem/progenitor differentiation, providing insights into how pluri-/multipotent stem/progenitor cells capable of differentiating into multiple distinct lineages are maintained by the balanced gene expression of cell-fate determinants. In addition, the applicability of the aforementioned previous methods appears to be limited to cell types, for which not only transcriptome data but also other types of data, such as gene ontology and curated GRNs, are already available. The computational method presented in this study requires only bulk transcriptome data and literature knowledge of TF interactions.

Experimental validation of predicted cell-fate determinants confirmed that overexpression of ESR1 and RUNX2 in mNSCs induces neuronal and astrocyte differentiation, respectively. Indeed, it has been previously shown that overexpression of ESR1 was able to induce differentiation in neuroblastoma cells (Loven et al., 2010; Ma et al., 1993). In addition, embryonic rat NSCs have been shown to undergo neuronal differentiation in response to an ESR1 ligand, estradiol (Brannvall et al., 2002). However, until the present study evidence that overexpression of ESR1 is able to induce differentiation of mNSCs into neurons has been lacking. Furthermore, RUNX2 is well known for inducing differentiation of MSCs into osteoblasts (Banerjee et al., 1997; Ducy et al., 1997; Komori et al., 1997; Otto et al., 1997) but it has not been previously shown to induce astrocyte differentiation, which demonstrates how a same gene can have different lineage specification roles depending on the biological context characterized by GRN motifs. Interestingly, the induction of neuronal differentiation by ESR1 is not as strong as the induction of astrocytes by RUNX2. Most probably this is because the utilized mNSC system mimics the developmental stage of late radial glia cells, which are more primed toward the astrocyte fate (Conti and Cattaneo, 2010; Conti et al., 2005); consequently, their induction into this fate is easier than neuronal induction.

In sum, here we have proposed a generalized GRN-based computational model of stem cell differentiation and a computational method that systematically predicts cell-fate determinants and their GRN motifs. The generality and simplicity of this method makes it easy to apply to new cellular differentiation events, and therefore can assist in guiding experiments in stem cell biology and regenerative medicine.

EXPERIMENTAL PROCEDURES

Detection of TF Pairs Whose Expression Ratio Is Significantly Changed upon Differentiation

The test statistic NRD, determining whether a pair of genes is equally expressed in the parental cell in comparison with the daughter, cell is defined by

$$NRD = \frac{\left(\frac{C_{gene1}^{Progenitor}}{C_{gene2}^{Progenitor}} - \frac{C_{gene1}^{Daughter}}{C_{gene1}^{Daughter}}\right)}{\left(\frac{C_{gene1}^{Progenitor}}{C_{gene2}^{Progenitor}}\right)},$$

where $C_{gene1}^{Progenitor}$, $C_{gene2}^{Progenitor}$, $C_{gene1}^{Daughter}$, and $C_{gene2}^{Daughter}$ are the expression values of gene 1 and gene 2 in the progenitor cell and a daughter cell, respectively. This value was calculated for all pairs of TFs annotated in AnimalTFDB (http://www.bioguo.org/AnimalTFDB/download/gene_list_of_Mus_musculus.txt). Since the distribution of NRDs was not Gaussian, they were then normalized by median absolute deviation (MAD) normalization defined by

$$\widehat{X}_j = \frac{X_j - median}{MAD},$$

where X_j is the NRD of gene pair j and $\widehat{X}_j$ is the normalized NRD, and the median and MAD are computed based on all NRDs. This normalized NRD was computed for each microarray replicate, and the moderated t significance test of this statistic was performed using the limma R package. The Benjamini-Hochberg multiple test correction was then applied with a false discovery rate cutoff of 0.05. This procedure was applied to each of the two cell lineages (two daughter cell types) separately, and TF pairs that had a significant NRD in both lineages in the opposite ratio directions were taken as the final set. We call these gene pairs "significant NRD TF pairs." Note that this p-value cutoff was set arbitrarily; however, the stringency can be adjusted by this p-value cutoff as well as that for the initial differential gene-expression test.

Prediction of Cell-Fate Determinant Pairs and Their GRN Motifs

The flowchart of this part of the method is shown in Figure S4. Our model of stem cell differentiation states that stem/progenitor cells correspond to stable states maintained by the balanced expression of cell-fate determinants residing in clusters of interconnected feedback loops (strongly connected components), Therefore, our aim here is to identify strongly connected components that contain significant NRD TF pairs and stabilize the Boolean stable steady states of the two daughter cell types. However, if one is interested in how these GRN motifs are connected to other genes, the entire GRN can be looked up.

In each of the best GRN solutions the largest strongly connected component was first identified using the Graph::Directed Perl module (http://search.cpan.org/dist/Graph/lib/Graph.pod). Each strongly connected component was then decomposed into smaller strongly connected components by first finding the first 300 shortest path-elementary circuits from each node using the graphkshortestpaths.m program (http://www.mathworks.com/matlabcentral/fileexchange/35397-k-shortest-paths-in-a-graph-represented-by-a-sparse-matrix–yen-s-algorithm-/content/graphkshortestpaths.m). Here we employed shortest paths, since we later consider directly connected genes as candidate cell-fate determinants and paths longer than shortest paths are not necessary for this step. For each of these decomposed strongly connected components, the attractor states were computed from the Booleanized microarray expression data of the two daughter cell types. If these two attractors were mutually exclusive and 100% identical to the attractors of the original GRN and to their starting microarray data, the motif was kept for subsequent analyses. We discarded strongly connected components whose attractors are either all 0 or 1, since our target motifs need to contain at least one upregulated TF for both attractors as potential candidate lineage specifiers. For each significant NRD TF pair, the most frequent strongly connected component was searched among the best GRN solutions and if the two TFs in the pair were directly connected to each other in that strongly connected component, the pair was considered the final candidate opposing cell-fate determinant pair together with its GRN motif. Note that these criteria were stringently set in the present study to demonstrate the proof of concept of the method. However, they can be easily relaxed and a longer list of candidate pairs and motifs can be assessed. In each stem cell system, candidate opposing cell-fate determinant pairs were ranked by their minimum out-degree interface (i.e., the smaller number of genes regulated by one of the two genes within the pair), since a pair with a higher number of out-degree interface is more likely to have a higher regulatory influence on the GRN.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and four figures and can be found with this article online at http://dx.doi.org/10.1016/j.stemcr.2016.07.014.

ACKNOWLEDGMENTS

The authors would like to thank Maria Pavlou and Thea van Wüllen for their experimental contributions to the project, and Robert Passier and Sabine den Hartogh for valuable discussion about their cardiac differentiation data. S.O. is supported by an FNR AFR Postdoctoral grant (7682104/PDR). S.Z. is supported by an FNR CORE grant (C13/BM/5810227). The J.C.S. laboratory is supported by a University of Luxembourg Internal Research project grant (MidNSCs).

Received: May 3, 2016
Revised: July 15, 2016
Accepted: July 16, 2016
Published: August 18, 2016

REFERENCES

Banerjee, C., McCabe, L.R., Choi, J.Y., Hiebert, S.W., Stein, J.L., Stein, G.S., and Lian, J.B. (1997). Runt homology domain proteins in osteoblast differentiation: AML3/CBFA1 is a major component of a bone-specific complex. J. Cell Biochem. *66*, 1–8.

Birket, M.J., Ribeiro, M.C., Verkerk, A.O., Ward, D., Leitoguinho, A.R., den Hartogh, S.C., Orlova, V.V., Devalla, H.D., Schwach, V., Bellin, M., et al. (2015). Expansion and patterning of cardiovascular

progenitors derived from human pluripotent stem cells. Nat. Biotechnol. *33*, 970–979.

Blelloch, R., Wang, Z., Meissner, A., Pollard, S., Smith, A., and Jaenisch, R. (2006). Reprogramming efficiency following somatic cell nuclear transfer is influenced by the differentiation and methylation state of the donor nucleus. Stem Cells *24*, 2007–2013.

Brannvall, K., Korhonen, L., and Lindholm, D. (2002). Estrogen-receptor-dependent regulation of neural stem cell proliferation and differentiation. Mol. Cell Neurosci. *21*, 512–520.

Cahan, P., Li, H., Morris, S.A., Lummertz da Rocha, E., Daley, G.Q., and Collins, J.J. (2014). CellNet: network biology applied to stem cell engineering. Cell *158*, 903–915.

Conti, L., and Cattaneo, E. (2010). Neural stem cell systems: physiological players or in vitro entities? Nat. Rev. Neurosci. *11*, 176–187.

Conti, L., Pollard, S.M., Gorba, T., Reitano, E., Toselli, M., Biella, G., Sun, Y., Sanzone, S., Ying, Q.L., Cattaneo, E., et al. (2005). Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. PLoS Biol. *3*, e283.

Crespo, I., and del Sol, A. (2013). A general strategy for cellular reprogramming: the importance of transcription factor cross-repression. Stem Cells *31*, 2127–2135.

den Hartogh, S.C., Schreurs, C., Monshouwer-Kloots, J.J., Davis, R.P., Elliott, D.A., Mummery, C.L., and Passier, R. (2015). Dual reporter MESP1 mCherry/w-NKX2-5 eGFP/w hESCs enable studying early human cardiac differentiation. Stem Cells *33*, 56–67.

Ducy, P., Zhang, R., Geoffroy, V., Ridall, A.L., and Karsenty, G. (1997). Osf2/Cbfa1: a transcriptional activator of osteoblast differentiation. Cell *89*, 747–754.

Graf, T. (2002). Differentiation plasticity of hematopoietic cells. Blood *99*, 3089–3101.

Huang, S., Guo, Y.P., May, G., and Enver, T. (2007). Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. Dev. Biol. *305*, 695–713.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. *3*, 318–356.

Jo, J., Hwang, S., Kim, H.J., Hong, S., Lee, J.E., Lee, S.G., Baek, A., Han, H., Lee, J.I., Lee, I., et al. (2016). An integrated systems biology approach identifies positive cofactor 4 as a factor that increases reprogramming efficiency. Nucleic Acids Res. *18*, 44.

Komori, T., Yagi, H., Nomura, S., Yamaguchi, A., Sasaki, K., Deguchi, K., Shimizu, Y., Bronson, R.T., Gao, Y.H., Inada, M., et al. (1997). Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts. Cell *89*, 755–764.

Kopp, J.L., Ormsbee, B.D., Desler, M., and Rizzino, A. (2008). Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. Stem Cells *26*, 903–911.

Kuo, C.T., Morrisey, E.E., Anandappa, R., Sigrist, K., Lu, M.M., Parmacek, M.S., Soudais, C., and Leiden, J.M. (1997). GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. Genes Dev. *11*, 1048–1060.

Lin, Y.F., Jing, W., Wu, L., Li, X.Y., Wu, Y., Liu, L., Tang, W., Long, J., Tian, W.D., and Mo, X.M. (2008). Identification of osteo-adipo progenitor cells in fat tissue. Cell Prolif. *41*, 803–812.

Loven, J., Zinin, N., Wahlstrom, T., Muller, I., Brodin, P., Fredlund, E., Ribacke, U., Pivarcsi, A., Pahlman, S., and Henriksson, M. (2010). MYCN-regulated microRNAs repress estrogen receptor-alpha (ESR1) expression and neuronal differentiation in human neuroblastoma. Proc. Natl. Acad. Sci. USA *107*, 1553–1558.

Ma, Z.Q., Spreafico, E., Pollio, G., Santagati, S., Conti, E., Cattaneo, E., and Maggi, A. (1993). Activated estrogen receptor mediates growth arrest and differentiation of a neuroblastoma cell line. Proc. Natl. Acad. Sci. USA *90*, 3740–3744.

Montserrat, N., Nivet, E., Sancho-Martinez, I., Hishida, T., Kumar, S., Miquel, L., Cortina, C., Hishida, Y., Xia, Y., Esteban, C.R., et al. (2013). Reprogramming of human fibroblasts to pluripotency with lineage specifiers. Cell Stem Cell *13*, 341–350.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat. Genet. *24*, 372–376.

Otto, F., Thornell, A.P., Crompton, T., Denzel, A., Gilmour, K.C., Rosewell, I.R., Stamp, G.W., Beddington, R.S., Mundlos, S., Olsen, B.R., et al. (1997). Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. Cell *89*, 765–771.

Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.F., D'Agati, V., Orkin, S.H., and Costantini, F. (1991). Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature *349*, 257–260.

Rackham, O.J., Firas, J., Fang, H., Oates, M.E., Holmes, M.L., Knaupp, A.S., FANTOM Consortium, Suzuki, H., Nefzger, C.M., Daub, C.O., et al. (2016). A predictive computational framework for direct reprogramming between human cell types. Nat. Genet. *48*, 331–335.

Ralston, A., and Rossant, J. (2005). Genetic regulation of stem cell origins in the mouse embryo. Clin. Genet. *68*, 106–112.

Roeder, I., and Glauche, I. (2006). Towards an understanding of lineage specification in hematopoietic stem cells: a mathematical model for the interaction of transcription factors GATA-1 and PU.1. J. Theor. Biol. *241*, 852–865.

Shi, Y., Desponts, C., Do, J.T., Hahm, H.S., Scholer, H.R., and Ding, S. (2008). Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds. Cell Stem Cell *3*, 568–574.

Shu, J., Wu, C., Wu, Y., Li, Z., Shao, S., Zhao, W., Tang, X., Yang, H., Shen, L., Zuo, X., et al. (2013). Induction of pluripotency in mouse somatic cells with lineage specifiers. Cell *153*, 963–975.

Siebert, H. (2009). Deriving behavior of Boolean bioregulatory networks from subnetwork dynamics. Math. Comput. Sci. *2*, 421–442.

Thomas, R. (1978). Logical analysis of systems comprising feedback loops. J. Theor. Biol. *73*, 631–656.

Voso, M.T., Burn, T.C., Wulf, G., Lim, B., Leone, G., and Tenen, D.G. (1994). Inhibition of hematopoiesis by competitive binding of transcription factor PU.1. Proc. Natl. Acad. Sci. USA *91*, 7932–7936.

Zañudo, J.G.T., and Albert, R. (2013). An effective network reduction approach to find the dynamical repertoire of discrete dynamic networks. Chaos *23*, 025111.

Zañudo, J.G.T., and Albert, R. (2015). Cell fate reprogramming by control of intracellular network dynamics. PLoS Comput. Biol. *11*, e1004193.

Zeineddine, D., Papadimou, E., Chebli, K., Gineste, M., Liu, J., Grey, C., Thurig, S., Behfar, A., Wallace, V.A., Skerjanc, I.S., et al. (2006). Oct-3/4 dose dependently regulates specification of embryonic stem cells toward a cardiac lineage and early heart development. Dev. Cell *11*, 535–546.

Zhang, X., Huang, C.T., Chen, J., Pankratz, M.T., Xi, J., Li, J., Yang, Y., Lavaute, T.M., Li, X.J., Ayala, M., et al. (2010). Pax6 is a human neuroectoderm cell fate determinant. Cell Stem Cell *7*, 90–100.

Stem Cell Reports, Volume *7*

Supplemental Information

A Generalized Gene-Regulatory Network Model of Stem Cell Differentiation for Predicting Lineage Specifiers

Satoshi Okawa, Sarah Nicklas, Sascha Zickenrott, Jens C. Schwamborn, and Antonio del Sol

Figure S1

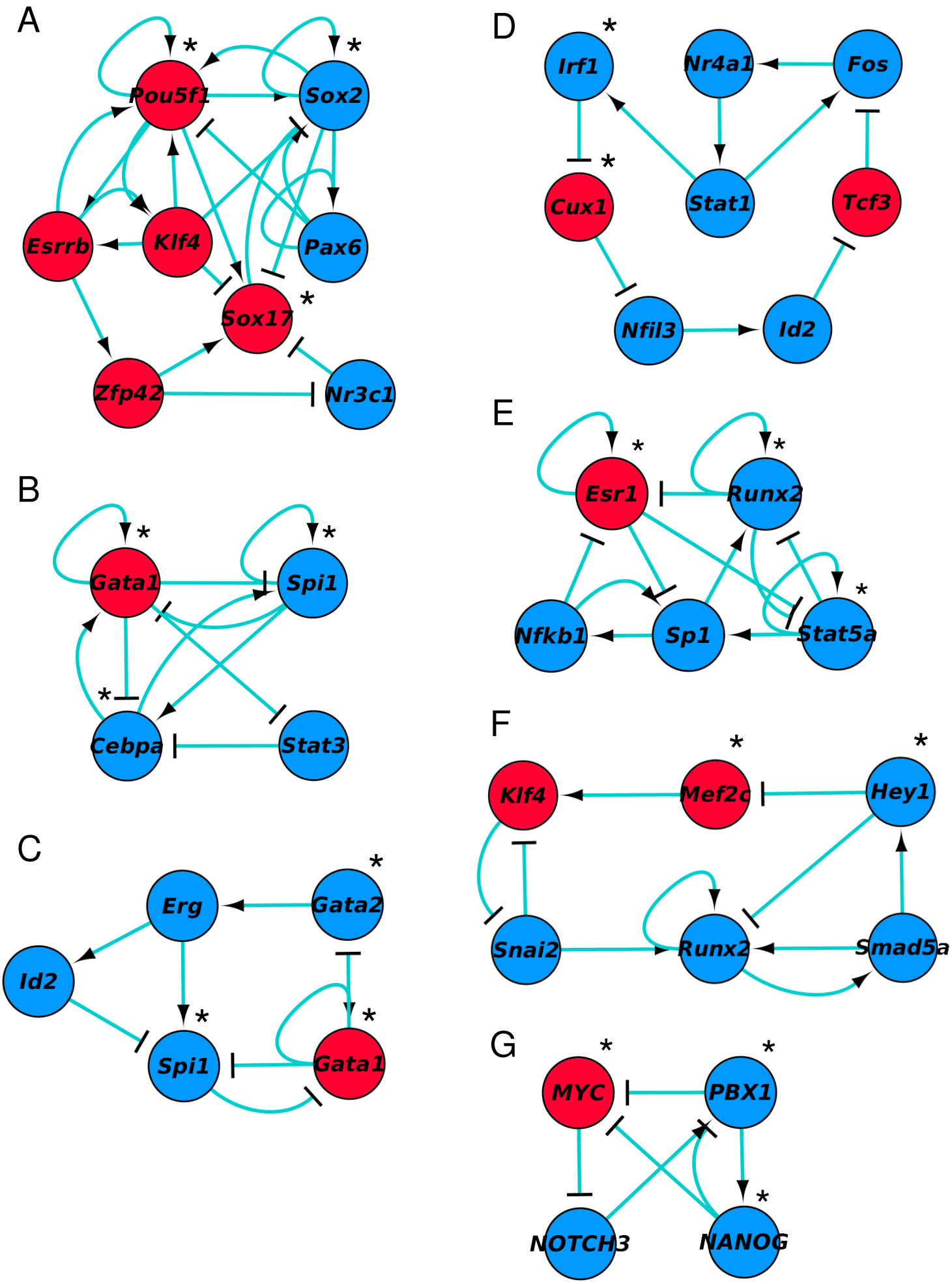

A
Pou5f1
Sox2
Esrrb
Klf4
Pax6
Sox17
Zfp42
Nr3c1
B
Gata1
Spi1
Cebpa
Stat3
C
Erg
Gata2
Id2
Spi1
Gata1
D
Irf1
Nr4a1
Fos
Cux1
Stat1
Tcf3
Nfil3
Id2
E
Esr1
Runx2
Nfkb1
Sp1
Stat5a
F
Klf4
Mef2c
Hey1
Snai2
Runx2
Smad5a
G
MYC
PBX1
NOTCH3
NANOG

Figure S2

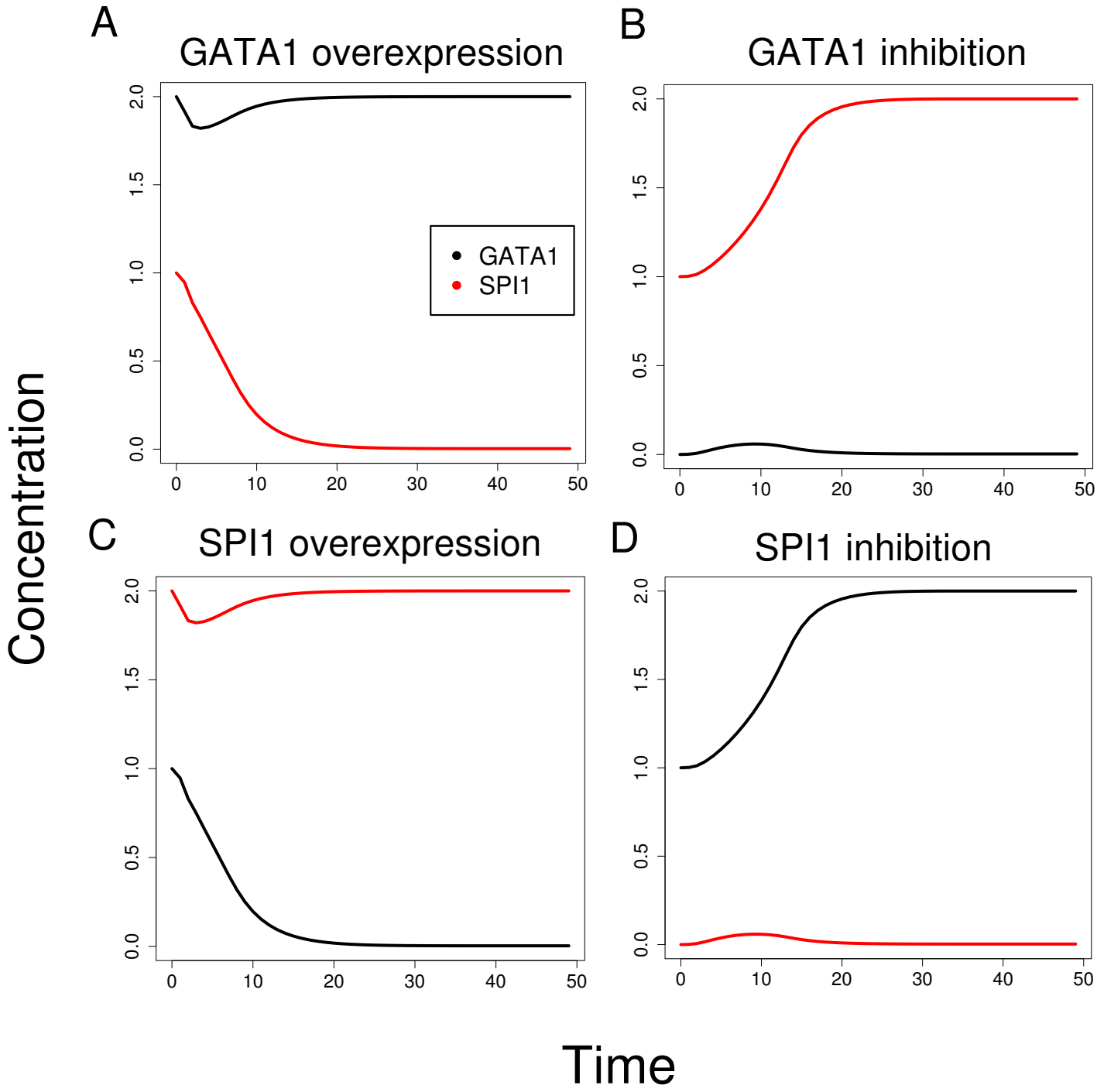

A
GATA1 overexpression
B
GATA1 inhibition
C
SPI1 overexpression
D
SPI1 inhibition
GATA1
SPI1
Concentration
Time
0.0
0.5
1.0
1.5
2.0
0
10
20
30
40
50

Figure S3

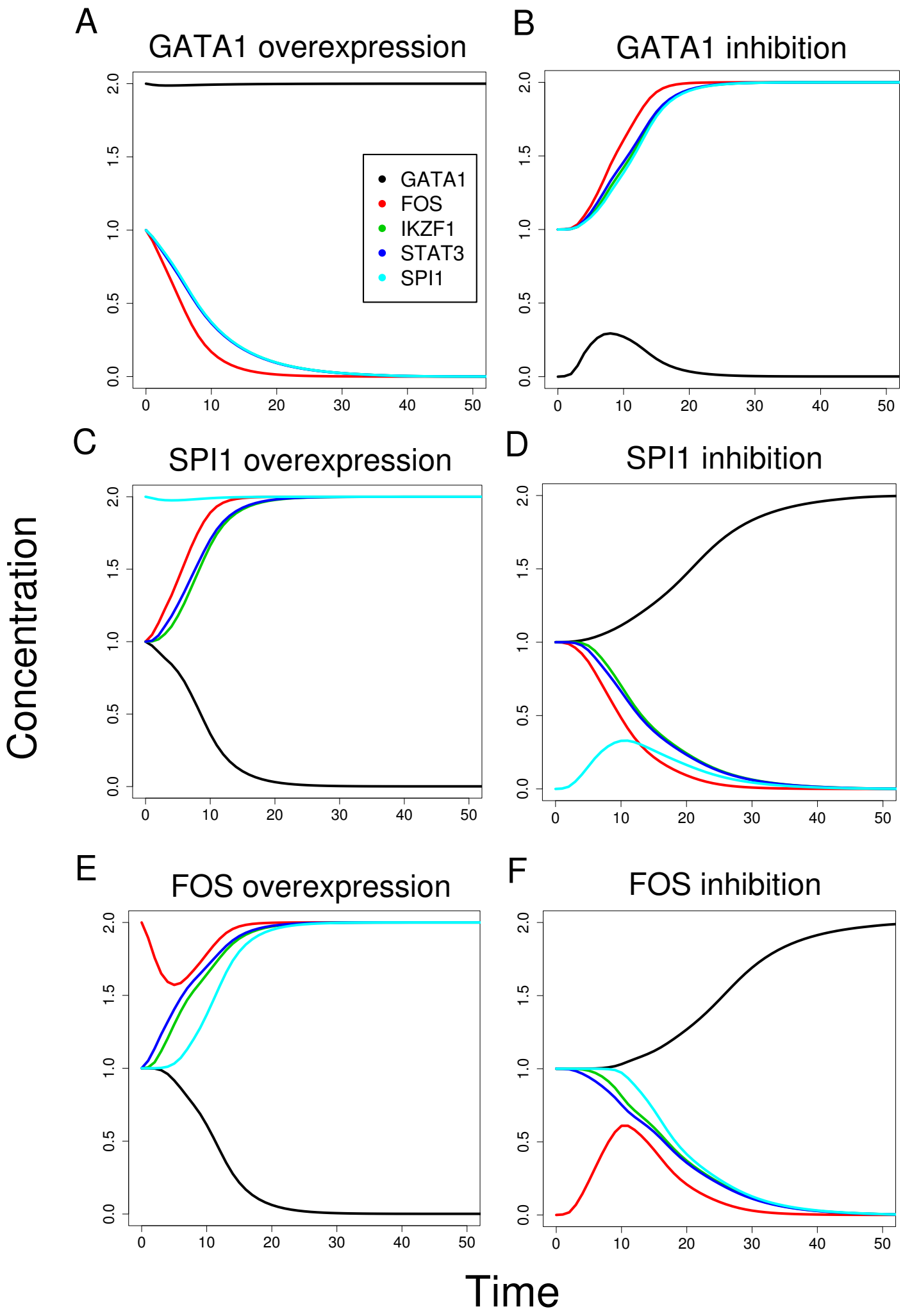
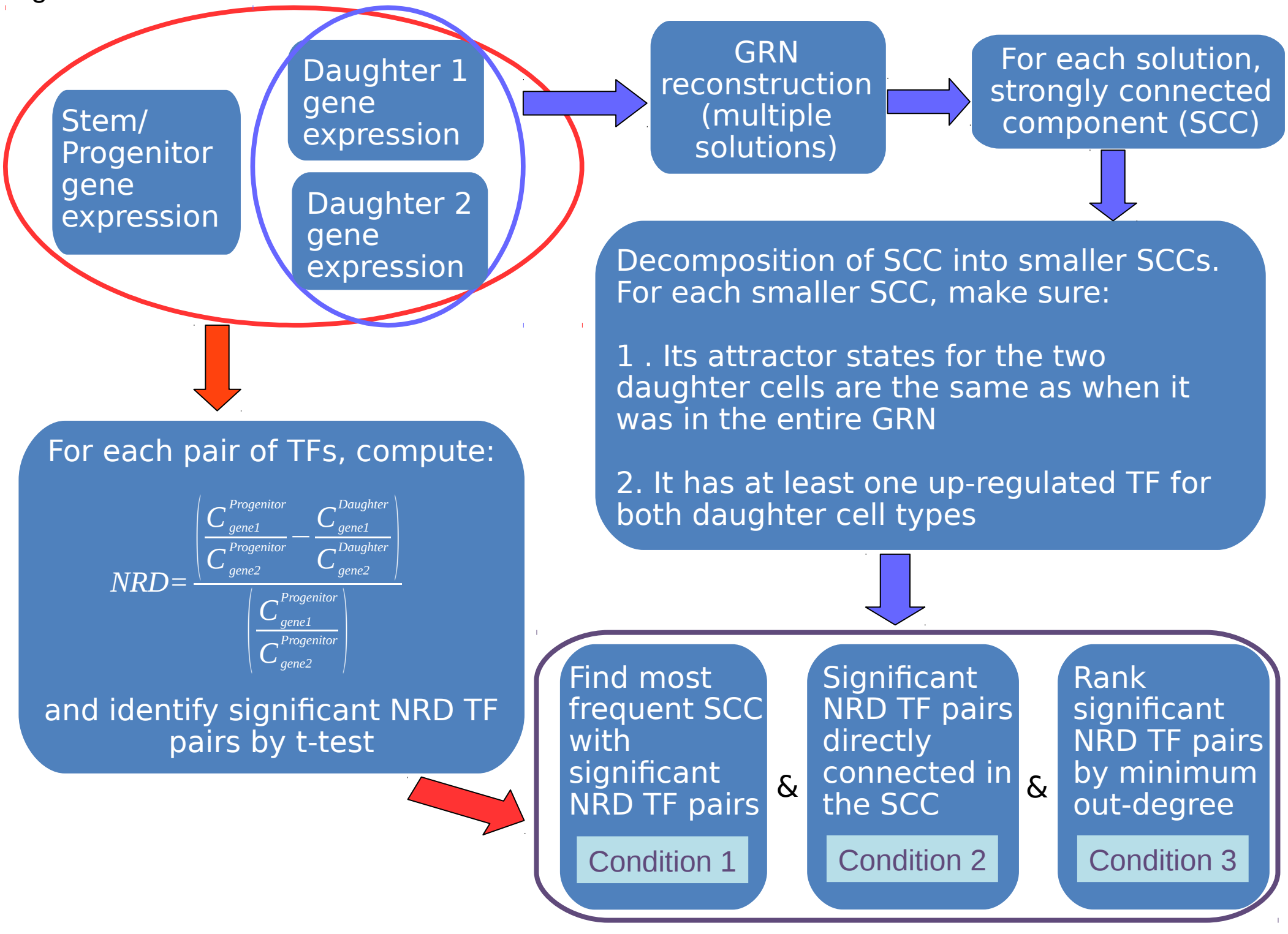

A
GATA1 overexpression
B
GATA1 inhibition
C
SPI1 overexpression
D
SPI1 inhibition
E
FOS overexpression
F
FOS inhibition
GATA1
FOS
IKZF1
STAT3
SPI1
Concentration
Time
0.0
0.5
1.0
1.5
2.0
0
10
20
30
40
50

Figure S4


Stem/ Progenitor gene expression
Daughter 1 gene expression
Daughter 2 gene expression
GRN reconstruction (multiple solutions)
For each solution, strongly connected component (SCC)
Decomposition of SCC into smaller SCCs. For each smaller SCC, make sure:
1 . Its attractor states for the two daughter cells are the same as when it was in the entire GRN
2. It has at least one up-regulated TF for both daughter cell types
For each pair of TFs, compute:
$$NRD = \frac{\left(\frac{C_{gene1}^{Progenitor}}{C_{gene2}^{Progenitor}} - \frac{C_{gene1}^{Daughter}}{C_{gene2}^{Daughter}} \right)}{\left(\frac{C_{gene1}^{Progenitor}}{C_{gene2}^{Progenitor}} \right)}$$
and identify significant NRD TF pairs by t-test
Find most frequent SCC with significant NRD TF pairs
Condition 1
&
Significant NRD TF pairs directly connected in the SCC
Condition 2
&
Rank significant NRD TF pairs by minimum out-degree
Condition 3

Supplementary Table and Figure legends

Figure S1. Predicted opposing cell fate determinant pairs and their strongly connected components in mESC, mHSC, mNSC, mMSC and hCPC systems.

Red nodes are genes upregulated in daughter 1 (mesoderm, erythroid, neuron, osteoblast and MESP1+ CPC, respectively). Blue nodes are genes upregulated in daughter 2 (ectoderm, myeloid, astrocyte, adipocyte and MESP1- CPC, respectively). Pointed arrows indicate activation, blunted arrows indicate inhibition. Network motifs of (**A**) *Sox17—Sox2* pair in mESCs, (**B**) *Gata1—Cebpa* pair in mHSCs, (**C**) *Gata1—Gata2* pair in mMSCs, (**D**) *Cux1—Irf1* pair in mHSCs, (**E**) *Esr1—Stat5a* pairs in mNSCs, and (**F**) *Hey1—Mef2c* pair in mNSCs, and (**G**) *MYC—NANOG* and *MYC—PBX1* pairs in hCPCs. * indicates predicted cell fate determinants.

Figure S2. Time trajectories of deterministic continuous simulation of *Gata1—Spi1* toggle switch (Figure 3B) upon perturbation.

Expression of GATA1 and SPI1 at stable steady states were defined as [1, 1], [2, 0] and [0, 2] in arbitrary unit for mHSC progenitors, erythroids and myeloids, respectively (see Methods). Initial conditions are (**A**) GATA1 overexpression, (**B**) GATA1 inhibition, (**C**) SPI1 overexpression, and (**D**) SPI1 inhibition. (**A**) and (**D**) reached erythroid stable steady state, whereas (**B**) and (**C**) reached myeloid stable steady state. ODEs and estimated parameters are shown in (Supplementary methods).

Figure S3. Time trajectories of continuous model for 5-gene motif containing *Gata1*, *Fos*, *Ikzf1*, *Stat3* and *Spi1* shown in Figure 3E.

Expression of GATA1, FOS, IKZF1, STAT3 and SPI1 at stable steady states were defined as [1, 1, 1, 1, 1], [2, 0, 0, 0, 0] and [0, 2, 2, 2, 2] in arbitrary unit for HSC progenitors, erythroids and myeloids, respectively (see Methods). Initial conditions are (**A**) GATA1 overexpression, (**B**) GATA1 inhibition, (**C**) SPI1 overexpression, (**D**) SPI1 inhibition, (**E**) FOS overexpression, and (**F**) FOS inhibition. (**A**), (**D**) and (**F**) reached erythroid stable steady state, whereas (**B**), (**C**) and (**E**) reached myeloid stable steady state. The progenitor state remained stable at [1, 1, 1, 1, 1] (not shown). ODEs and estimated parameters are shown in (Supplemental methods).

Figure S4. Flowchart of the method.

Three transcriptome data for stem/progenitor cell type and two daughter cell types are used for computing significant NRD TF pairs. In parallel, GRNs are reconstructed using transcriptome data for two daufhter cell types. Each GRN is then decomposed into a strongly connected component (SCC), which is then further decomposed into smaller SCCs. Finally, SCCs with significant NRD TF pairs, which satisfy the presented criteria are considered thd final predictions.

Supplemental Experimental Procedures

Microarray data processing and analysis

Micorarray data of five stem cell systems (mESC, mHSC, mNSC, mMSC and hCPC) were obtained from the following sources. For the mESC system, mESCs (GSM720404, GSM720405, GSM720406, GSM720407, GSM720408, GSM720409, GSM720410, GSM720412), ectoderms (GSM338146, GSM338150, GSM338152, GSM338156), and mesoderms (GSM747049, GSM747050, GSM747051). The data for the mHSC system was taken from (May et al., 2013), including mFDCPs (GSM1211192, GSM1211193, GSM1211194), erythroids (GSM1211279, GSM1211280, GSM1211281), and myeloids (GSM1211366, GSM1211367, GSM1211368). The data for the mNSC system consists of mNSCs (Palm et al., 2013), neurons (GSM241896, GSM241897, GSM241899, GSM241901, GSM241903, GSM241904 and GSM241922), and astrocytes (GSM241905, GSM241906, GSM241907, GSM241909, GSM241910, GSM241911, GSM241913, GSM241923 and GSM241924) (Cahoy et al., 2008). The data for mMSCs were obtained from GSM1180589, GSM1180590 and GSM1180591, osteoblasts from GSM234794, GSM234795, GSM234796, and GSM234797 (Schroeder et al., 2007) and adipocytes from GSM1254880, GSM1254881, GSM1254882 and GSM1254883 (Ralston et al., 2014). The data for the differentiation of hESCs (Day 0) into MESP1+ CPCs and MESP1- CPCs (Day 3) were taken from (Den Hartogh et al., 2015).

Raw intensity values were normalized by variance stabilising normalization using the vsn R package (Huber et al., 2002). Quantile normalization was performed when the platforms within each stem cell system were different (i.e., mESC, mNSC and mMSC systems). The differential expression analysis was performed by a moderated t-test using the limma R package (Smyth, 2004) between the ectoderm and mesoderm, erythroids and myeloids, neurons and astrocytes, osteaoblasts and adipocytes, and MESP1+ CPCs and MESP1- CPCs. Genes were binned into 30 bins by intensity and the moderated t-test was applied to each bin. The Benjamini-Hochberg multiple test correction was applied with the false discovery rate (FDR) cutoff 0.05. In all the cases genes with mean log2 fold-change less than 1 were discarded. When a gene had more than one microarray probe, the one with the highest variance across samples was used for subsequent analysis.

GRN reconstruction

Direct gene interactions between the two daughter cells in each differentiation system were retrieved from MetaCore (GeneGo Inc. (Nikolsky et al., 2005)) using differentially expressed TFs. The dates of download were between March and August of 2014. The interaction types "Transcriptional regulation" and "Binding" identified in both mouse and human were kept for the subsequent analyses. In addition, genes with node degree less than seven were discarded to focus on genes with high degrees, since these genes were not forming strongly connected

components in the initial GRN and therefore would not be in the final GRNs. The network edge (interaction) pruning was performed using the modified version of the method proposed by (Crespo et al., 2013) re-implemented in MATLAB using the genetic algorithm (ga) function. Briefly, this algorithm assumes that each cellular phenotype is a Boolean stable steady state attractor of a given network, and removes edges that are inconsistent with the Booleanized mRNA expression data. This pruning was conducted between the two daughter cell types, which resulted in GRNs whose Boolean attractor states correspond to the gene expression states of both daughter cell types. The genetic algorithm was run between 1000-1500 populations and 100 iterations. Although the current version of our algorithm does not regularize potential overfitting, we alleviate this issue by considering all best GRN solutions for subsequent analyses, although this might not fully resolve potential overfitting. The Boolean simulation was carried out using the pbn-matlab-toolbox (http://code.google.com/p/pbn-matlab-toolbox/downloads/list) using the synchronous updating scheme. The node weights were set to all 1. The logic rule was defined, so that the number of activating edges and inhibiting edges acting on a gene were compared and the one with a higher number dominates (i.e., the threshold rule). If both numbers are the same, the state was set to 0 (i.e., inhibition dominant). During this process, "unassigned" interactions (i.e., interactions without knowledge of activation or inhibition) were randomly assigned "activation" or "inhibition" and the one that yielded a better result was taken for the next generation. GRN motifs were visualized in Cytoscape (version 2.7.0) (Shannon et al., 2003).

Deterministic continuous simulation

The dynamics of relative protein abundance was modelled using the ODEs with the "OR" logic described in (Huang et al., 2007), which draw on the Michaelis-Menten formalism with Hill coefficients. The microarray expression value of each gene in the motif was ordered among the three different cell types (stem/progenitor and two daughter cells) and assigned three integers, 0, 1 or 2, for the low, intermediate and high expression values. For example, if a gene has log2 gene expression values 6, 8 and 10 for the daughter cell 1, progenitor, and daughter cell 2, then the integers 0, 1 and 2 were assigned to each stable steady state, respectively. Then the parameters were estimated by equating the ODEs to 0 at these three stable steady states. Note, we did not impose any constraint on the dynamics of the system, since the purpose of this simulation study is to illustrate that the predicted GRN motif can, upon perturbation of its genes, exhibit the expected binary differentiation dynamics from the steady state corresponding to the stem/progenitor cell type, Since this problem is intractable and it is infeasible to explore the entire parameter space, we used MATLAB's "fmincon" function (interior-point method), which was combined with the "GlobalSearch" function. The initial parameters were all set to 1 and the parameter boundary was set between 0.01 and 20. The solutions to ODEs were approximated by the 1st Taylor series using the MATLAB "taylor" function. The simulation was carried out using the Systems Biology Toolbox for MATLAB (Schmidt and Jirstrand, 2006). The "ode23s" function was used for solving ODEs.

The set of ODEs for the *Gata1-Spi1* toggle switch (Figure 3B) is,

$$\frac{d[GATA1]}{dt}=\frac{a_{11}[GATA1]^n}{K_1^n+[GATA1]^n}+\frac{a_{12}K_2^n}{K_2^n+[SPI1]}-\gamma_1[GATA1]$$

$$\frac{d[SPI1]}{dt}=\frac{a_{22}[SPI1]^n}{K_2^n+[SPI1]^n}+\frac{a_{21}K_1^n}{K_1^n+[GATA1]}-\gamma_2[SPI1]$$

where n = 7.4207, a_{11} = 4.824, a_{12} = 4.8712, a_{22} = 4.824, a_{21} = 4.8712, K_1 = 0.91729, K_1 = 0.91729, γ_1 = 4.8403, γ_2 = 4.8403. The set of ODEs for the motif shown in (Figure 3E) is,

$$\frac{d[GATA1]}{dt}=\frac{a_{11}[GATA1]^n}{K_{11}^n+[GATA1]^n}+\frac{a_{13}K_{13}^n}{K_{13}^n+[IKZF1]}+\frac{a_{15}K_{15}^n}{K_{15}^n+[SPI1]}-\gamma_1[GATA1]$$

$$\frac{d[FOS]}{dt}=\frac{a_{21}K_{21}^n}{K_{21}^n+[GATA1]^n}+\frac{a_{24}[STAT3]^n}{K_{24}^n+[STAT3]^n}+\frac{a_{25}[SPI1]^n}{K_{25}^n+[SPI1]^n}-\gamma_2[FOS]$$

$$\frac{d[IKZF1]}{dt}=\frac{a_{31}K_{31}^n}{K_{31}^n+[GATA1]^n}+\frac{a_{34}[STAT3]^n}{K_{34}^n+[STAT3]^n}-\gamma_3[IKZF1]$$

$$\frac{d[STAT3]}{dt}=\frac{a_{41}K_{41}^n}{K_{41}^n+[GATA1]^n}+\frac{a_{42}[FOS]^n}{K_{42}^n+[FOS]^n}-\gamma_4[STAT3]$$

$$\frac{d[SPI1]}{dt}=\frac{a_{51}K_{51}^n}{K_{51}^n+[GATA1]^n}+\frac{a_{55}[SPI1]^n}{K_{55}^n+[SPI1]^n}-\gamma_5[SPI1]$$

where n = 12.311, a_{11} = 4.5282, a_{13} = 2.8374, a_{15} = 2.8374, a_{21} = 7.1558, a_{24} = 6.5507, a_{25} = 6.5507, a_{31} = 4.2685, a_{34} = 6.8057, a_{41} = 4.2684, a_{42} = 6.8064, a_{51} = 4.2683, a_{55} = 6.805, K_{11} = 1.2849, K_{13} = 1.1608, K_{15} = 1.1608, K_{21} = 1.082, K_{24} = 1.0417, K_{25} = 1.0417, K_{31} = 1.1068, K_{34} = 1.0608, K_{41} = 1.1046, K_{42} = 1.0597, K_{51} = 1.1027, K_{55} = 1.0588, γ_1 = 5.0918, γ_2 = 10.126, γ_3 = 5.5357, γ_4 = 5.5361, γ_5 = 5.5353. All the models described above remained in the stable steady states corresponding to the three cell types. It is worth noting that there may exist different other sets of parameters, apart from the ones we showed here, that also reproduce the three stable states and the correct dynamics. Thus, our model provides only a qualitative analysis of the state-space of the motifs.

Pseudo-code for identification of key GRN motifs

1. Decompose strongly connected components (SCCs) into smaller SCCs

Best GRN solutions are the result from the previous step of GRN pruning, which gives rise to

```
# multiple equally best solutions
get all best GRN solutions;

loop through each best GRN solution
    find largest SCC;

    loop through each node in the SCC
        find n shortest paths starting from the node and coming back to it (i.e., feedback loops);
        from the GRN solution, extract all edges among the nodes in each shortest path (i.e.,
        decomposed SCC;

        save the decomposed SCCs;

    end loop

    # Discard duplicated topologically identical, decomposed SCCs
    get unique decomposed SCCs;

    loop through each decomposed SCC
        if (the SCC contains at least one node whose Booleanized gene expression state is up for
        both daughter cell types):

            compute Boolean attractors of the SCC with the initial states being the Booleanized gene
            expression states of the two daughter cell types;

            if (the computed two attractor states of the nodes in the SCC are identical to the
            computed two attractor states of the entire GRN solution):
                if (the computed two attractor states of the nodes in the SCC are identical to the
                Booleanized gene expression states of the two daughter cell types):

                    keep the SCC;

                end if
            end if

        end if
    end loop

end loop


## 2. Compute the frequency of SCCs containing each significant NRD TF pair ##

get all significant NRD TF pairs;

loop through each significant NRD TF pair
    loop through decomposed SCCs from 1.
        if (the SCC contains both TFs):

            compute the frequency of the SCC among all the best GRN solutions;

        end if
```

end loop
end loop

Reagents and plasmids

For immunolabelling the antibodies anti-TUJ1 (BioLegend, #801201) and anti-GFAP (Millipore, #MAB3402) were used. Alexa-fluorophore-conjugated antibodies (Invitrogen, #A11031) were used as secondary antibodies. DNA was counterstained using Hoechst 33258 (Invitrogen, #62249). The following plasmids were used: pCMV-VSV-G, psPAX2 (lentiviral packaging plasmids) (Addgene), pLenti-Runx2-C-mGFP, pLenti-Esr1-C-mGFP (Origene) and pGIPZ pMB049 (Marc Buehler).

Cell culture

Primary NSCs were isolated from C57BL/6N mouse brains at embryonic day 12.5-14.5 and cultivated as described previously (Conti and Cattaneo, 2010; Conti et al., 2005). Briefly, primary NSCs were kept on poly-D-Lysine (Sigma)-coated 10-cm polystyrene tissue culture dishes in DMEM/Ham's F12 medium (PAA) supplemented with 10 ng/mL EGF (Peprotech), 10 ng/mL bFGF-2 (Peprotech), 1 x N2 (Invitrogen), L-Glutamine (PAA), and Penicillin/Streptomycin (PAA). HEK293T cells were cultivated on uncoated 10-cm polystyrene tissue culture dishes in DMEM (Sigma) supplemented with 10% heat-inactivated FCS (PAA), L-Glutamine (PAA) and Penicillin/Streptomycin (PAA).

Lentivirus production

Lentiviruses were produced using a three-plasmid transfection protocol. One day prior to transfection, HEK293T cells were seeded in 10-cm polystyrene tissue culture dishes. The next day, the lentiviral packaging plasmids pCMV-VSV-G and psPAX2 were mixed with either pGIPZ pMB049, pLenti-Runx2-C-mGFP or pLenti-Esr1-C-mGFP and the HEK293T cells were transfected with these plasmids using Fugene6 (Promega) according to manufacturer`s instructions. Three days post transfection, the supernatants were harvested and cleared from remaining cells by centrifuging for 10 min at 3,000 x g and 4 °C. The supernatant was mixed with 1/5 volume of 40% PEG and incubated overnight at 4 °C. The next day, the lentivirus was concentrated by centrifugation for 30 min at 1,500 x g and 4 °C. After removal of the supernatant, the pellet was centrifuged again for 5 min at 1,500 x g and 4 °C. The remaining supernatant was removed and the pellet was resuspended in an appropriate amount of DMEM without supplements and stored at -80 °C.

Viral transductions

For viral transduction, primary NSCs were seeded onto poly-D-Lysine-coated coverslips at a density of 25,000 cells / well. One day after seeding the virus was diluted in growth medium and added to the cells. At two and four days post transduction, half of the growth medium was

exchanged by fresh growth medium.

Immunocytochemistry

For immunocytochemical staining, cells were fixed with 4% paraformaldehyde in 120 mM PBS, pH 7.4 (4% PFA/PBS) followed by permeabilisation for 3 min at 4 °C using 0.05% Triton X-100 in PBS. Next, cells were blocked with 10% FCS in PBS for 1 h at RT and subjected to immunofluorescence staining with primary and secondary antibodies diluted in blocking solution. Images were collected with a Zeiss epifluorescence microscope and image analysis was conducted using ZEN lite (Zeiss) and Adobe Photoshop softwares.

Supplemental References

Cahoy, J.D., Emery, B., Kaushal, A., Foo, L.C., Zamanian, J.L., Christopherson, K.S., Xing, Y., Lubischer, J.L., Krieg, P.A., Krupenko, S.A.*, et al.* (2008). A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. The Journal of neuroscience : the official journal of the Society for Neuroscience *28*, 264-278.

Crespo, I., Krishna, A., Le Bechec, A., and del Sol, A. (2013) Predicting missing expression values in gene regulatory networks using a discrete logic modeling optimization guided by network stable states. *Nucleic acids research 41*, e8.

Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A., and Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics *18 Suppl 1*, S96-104.

May, G., Soneji, S., Tipping, A.J., Teles, J., McGowan, S.J., Wu, M., Guo, Y., Fugazza, C., Brown, J., Karlsson, G.*, et al.* (2013). Dynamic analysis of gene expression and genome-wide transcription factor binding during lineage specification of multipotent progenitors. Cell stem cell *13*, 754-768.

Nikolsky, Y., Ekins, S., Nikolskaya, T., and Bugrim, A. (2005) A novel method for generation of signature networks as biomarkers from complex high throughput data. *Toxicology letters 158*, 20-29.

Palm, T., Hemmer, K., Winter, J., Fricke, I.B., Tarbashevich, K., Sadeghi Shakib, F., Rudolph, I.M., Hillje, A.L., De Luca, P., Bahnassawy, L.*, et al.* (2013). A systemic transcriptome analysis reveals the regulation of neural stem cell maintenance by an E2F1-miRNA feedback loop. Nucleic acids research *41*, 3699-3712.

Ralston, J.C., Badoud, F., Cattrysse, B., McNicholas, P.D., and Mutch, D.M. (2014). Inhibition of stearoyl-CoA desaturase-1 in differentiating 3T3-L1 preadipocytes upregulates elongase 6 and downregulates genes affecting triacylglycerol synthesis. Int J Obes (Lond) *38*, 1449-1456.

Schmidt, H., and Jirstrand, M. (2006). Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology. Bioinformatics *22*, 514-515.

Schroeder, T.M., Nair, A.K., Staggs, R., Lamblin, A.F., and Westendorf, J.J. (2007). Gene profile analysis of osteoblast genes differentially regulated by histone deacetylase inhibitors. BMC genomics *8*, 362.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research 13*, 2498-2504.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol *3*, Article3.