

Hard-wired heterogeneity in blood stem cells revealed using a dynamic regulatory network model

Nicola Bonzanni^{1,2,3,4,†}, Abhishek Garg^{5,†}, K. Anton Feenstra^{1,2,4,†}, Judith Schütte^{6,7}, Sarah Kinston^{6,7}, Diego Miranda-Saavedra^{6,7}, Jaap Heringa^{1,2,4,*}, Ioannis Xenarios^{5,*} and Berthold Göttgens^{6,7,*}

¹BIVU Centre for Integrative Bioinformatics, VU University Amsterdam, ²AIMMS Amsterdam Institute for Molecules Medicines and Systems, VU University Amsterdam, De Boelelaan 1081, ³NKI-AVL The Netherlands Cancer Institute, Plesmanlaan 121, Amsterdam, ⁴NBIC Netherlands Bioinformatics Centre, Geert Grooteplein 28, Nijmegen, The Netherlands, ⁵Swiss Institute of Bioinformatics, Quartier Sorge—Batiment Genopode, Lausanne, Switzerland, ⁶Cambridge Institute for Medical Research and ⁷Wellcome Trust and MRC Cambridge Stem Cell Institute, University of Cambridge, Hills Road, Cambridge, UK

ABSTRACT

Motivation: Combinatorial interactions of transcription factors with *cis*-regulatory elements control the dynamic progression through successive cellular states and thus underpin all metazoan development. The construction of network models of *cis*-regulatory elements, therefore, has the potential to generate fundamental insights into cellular fate and differentiation. Haematopoiesis has long served as a model system to study mammalian differentiation, yet modelling based on experimentally informed *cis*-regulatory interactions has so far been restricted to pairs of interacting factors. Here, we have generated a Boolean network model based on detailed *cis*-regulatory functional data connecting 11 haematopoietic stem/progenitor cell (HSPC) regulator genes.

Results: Despite its apparent simplicity, the model exhibits surprisingly complex behaviour that we charted using strongly connected components and shortest-path analysis in its Boolean state space. This analysis of our model predicts that HSPCs display heterogeneous expression patterns and possess many intermediate states that can act as ‘stepping stones’ for the HSPC to achieve a final differentiated state. Importantly, an external perturbation or ‘trigger’ is required to exit the stem cell state, with distinct triggers characterizing maturation into the various different lineages. By focusing on intermediate states occurring during erythrocyte differentiation, from our model we predicted a novel negative regulation of Fli1 by Gata1, which we confirmed experimentally thus validating our model. In conclusion, we demonstrate that an advanced mammalian regulatory network model based on experimentally validated *cis*-regulatory interactions has allowed us to make novel, experimentally testable hypotheses about transcriptional mechanisms that control differentiation of mammalian stem cells.

Contact: j.heringa@vu.nl or ioannis.xenarios@isb-sib.ch or bg200@cam.ac.uk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

The remarkable power of small combinations of transcription factors to program and reprogram cellular phenotypes is exerted through their ability to modulate the expression levels of their target genes, typically in the range of a few hundred to a few thousand genes. Despite the power of single transcription factors to influence cell fate decisions, it is clear that the transcriptional state of any given cell type is the result of interactions within wider (transcriptional) regulatory networks. These regulatory networks are composed of both the transcription factors (TFs) and the *cis*-regulatory elements they are bound to (Davidson, 2006). Regulatory network reconstruction, therefore, requires the identification of *cis*-regulatory elements, as well as the upstream factors that bind them.

Haematopoiesis (blood formation) has long served as a model process for studying stem cells and represents the best characterized adult stem cell system with sophisticated purification strategies and functional stem cell assays. Transcriptional regulation is a key factor controlling haematopoiesis (Miranda-Saavedra and Göttgens, 2008), a fact underlined by the large number of TF genes that play key roles in normal haematopoiesis and/or the development of leukaemia (Göttgens, 2004). However, relatively little is known about the way key regulators interact with each other in forming the transcriptional networks controlling haematopoiesis.

Identification and subsequent characterization of gene regulatory elements is central to the reconstruction of transcriptional regulatory networks because these elements dictate the connectivity and topology of transcriptional regulatory networks (Davidson, 2006). Regulatory elements can be analysed using a variety of assays, such as transfection assays of luciferase reporter constructs or chromatin immunoprecipitation (ChIP) analysis to identify upstream regulators. However, the identification of true *in vivo* activities of mammalian regulatory elements requires the use of transgenic mouse systems. Regulatory elements from 11 gene loci active in haematopoietic stem/progenitor cells (HSPCs) have been validated using all the aforementioned assays, including transgenic mice (Donaldson *et al.*, 2005a; Göttgens *et al.*, 2002; Göttgens *et al.*, 2004; Kobayashi-Osaki *et al.*, 2005; Landry *et al.*, 2008; Nottingham *et al.*, 2007; Okuno *et al.*, 2005; Pimanda *et al.*, 2007; Vyas *et al.*, 1999;

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

Wilson *et al.*, 2009). This wealth of data, therefore, represents a unique opportunity to (re)construct transcriptional network models for developing blood stem cells.

Network modelling is increasingly recognized as a powerful approach to deal with the complexity of biological processes, including the intricate interactions between TFs (Georgescu *et al.*, 2008; Hu *et al.*, 2007; Karlebach and Shamir, 2008; Krumsiek *et al.*, 2011; Spooner *et al.*, 2009; Thoms *et al.*, 2011). Most of the current experimental data describing the function of haematopoietic TFs are of a qualitative nature (e.g. Gata1 and Scl together activate Scl expression), which limits the choice of possible modelling approaches. However, the accumulated knowledge of regulatory interactions (Foster *et al.*, 2009; Swiers *et al.*, 2006) contains experimentally validated information on the topology of regulatory sub-circuits, including positive and negative feedback loops, which are important for maintenance of both the stem cell phenotype (Pimanda *et al.*, 2007) and differentiation into different mature blood cell types (Sieweke and Graf, 1998). An important challenge for regulatory network reconstruction is to devise models that can represent the dynamic interactions between important sub-circuits and represent the changes in gene expression when cells are undergoing differentiation.

Importantly, experimentally defined regulatory hierarchies based on regulatory elements up to now largely represent a static view, which, in the case of blood stem cell formation is centred on a single time point in transgenic mouse assays (activity within the dorsal aorta region and foetal liver of the mid-gestation mouse embryo). Here, we have generated a network model based on extensive experimental data with the goal to better understand how core stem cell network circuits are incorporated into the wider dynamic system of blood stem cell development and differentiation. Through the modelling of steady-states and dynamic network behaviour, we were able to identify specific genes and feedback loops within the network that are likely key players in cellular decision making, such as the dynamic processes of stem cell maintenance and/or differentiation. Moreover, our analysis revealed heterogeneous gene expression states within undifferentiated blood stem cells, as well as accurately captured perturbations required to differentiate HSPCs along a specific lineage. Importantly, based on our modelling results, we made a hypothesis that Gata1 negatively regulates Fli1, which we validated experimentally using transcriptional assays, thus providing new insights into the dynamic nature of regulatory networks controlling differentiation of blood stem cells into erythroid cells.

2 METHODS

2.1 Experimental

A reporter construct carrying the luciferase gene driven by Fli1 enhancer was introduced into the HPC7 cells (a murine haematopoietic progenitor cell line) by electroporation, and luciferase activity measured as described previously (Göttgens *et al.*, 1997). The Fli1 enhancer reporter construct has been described previously (Donaldson *et al.*, 2005b). Results of transactivation experiments were analysed as described previously (Bockamp *et al.*, 1998). Individual experiments were performed in triplicates on at least two different days.

2.2 Boolean modelling

In Boolean modelling of a gene regulatory network (GRN), a gene can exist in only two expression states: active and inactive (represented by Boolean 1 and 0, respectively); and the interaction between genes/proteins is represented using Boolean logic functions, such as AND, OR, BUFF and NOT (Davidich and Bornholdt, 2008; Davidson *et al.*, 2002; Garg *et al.*, 2009; Li *et al.*, 2004; Smith *et al.*, 2007). Advantages of Boolean modelling approach over more traditional continuous modelling approaches based on ordinary differential equations is that kinetic parameters are not required to define interactions between the genes (or proteins). However, such a simplification comes at the cost of discretization of the gene expression (or gene activity) to only two expression levels, namely, present or absent. Nevertheless, Boolean modelling can efficiently capture the required dynamics of a GRN and has been successfully applied in the past to model various biological phenomena, such as cellular differentiation and embryo development (Davidich and Bornholdt, 2008; Davidson *et al.*, 2002; Li *et al.*, 2004; Smith *et al.*, 2007). We used Boolean logic functions AND, OR, BUFF, IAND and NOT as described previously (Garg *et al.*, 2008, 2009; Kauffman *et al.*, 2003; Klamt *et al.*, 2006; Mendoza and Xenarios, 2006) to represent the interactions between the genes in the GRN. A Boolean network corresponding to a sample GRN is shown in Supplementary Figure S1. We use the Boolean modelling toolbox GenYsis to compute attractors of Boolean functions mapped GRN and perform *in silico* gene perturbation experiments (Garg *et al.*, 2008). In this work, we use the fully asynchronous approach to model the time evolution and compute steady-states of the haematopoietic gene regulatory network (shown in Fig. 1A). The following three assumptions are made in the fully asynchronous modelling approach: (i) at most one gene can change its state (be updated) in a single step. (ii) At least one gene changes its state in a single step unless none of the genes can change their expression levels. (iii) Every gene is equally likely to change its state in a given step. With these constraints, every state can have potentially N successor states, N being the number of genes in the GRN, where each successor state differs from the present state in only one gene expression. The fully asynchronous models have been used often in the literature (Mangla *et al.*, 2010; Thomas, 1991; Thomas *et al.*, 1995).

2.3 State-space analysis

Strongly connected components SCCs in the state space were calculated using Tarjan's algorithm. Stable states were identified as terminal SCCs (TSCCs), i.e. an SCC with no outgoing edges.

Shortest traversable paths in the state space were calculated using Dijkstra's algorithm. A naïve brute force analysis (sufficient to analyse all ~900 000 paths found) was performed to find shortest traversable paths originating from the experimentally known HSPC expression pattern and leading to the experimentally known cell-type expression pattern states of interest. A similar analysis was performed for the state nearest to any of the sub-states in the HSPC TSCC leading to the cell type states.

Supplementary Figure S3 summarizes the procedure taken here in an overview of our analysis pipeline, starting from the experimental interactions included in the model.

3 RESULTS

3.1 A transcriptional regulatory network model for blood stem cells

Systematic curation of previously published results on haematopoietic regulatory elements allowed us to construct the first comprehensive regulatory network model based on 11 fully validated regulatory elements linking together 11 transcription factors, all of which are active in early HSPCs (see Supplementary Table S1

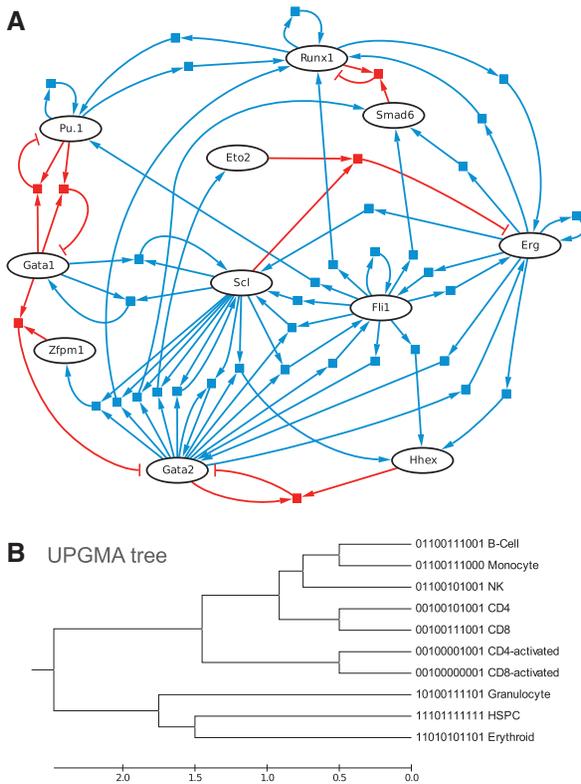


Fig. 1. A blood stem cell regulatory network model grounded on comprehensive *cis*-regulatory information. **(A)** Diagram of the haematopoietic gene regulatory network with logical functions between genes (ellipses) encoded directly by explicit transitions (squares). Activating interactions are shown as blue arrows, repressing interactions in red with 'flat heads'. All regulatory information encoded in this model can be found in Supplementary Table S2. **(B)** UPGMA (Unweighted Pair Group Method with Arithmetic mean) dendrogram based on hamming distance between cell-type-specific gene expression patterns shows that the 11 network model genes are sufficient to uniquely identify each of the cell types considered. Labels indicate the cell type names and the corresponding binary expression patterns. Note that this dendrogram should not be confused with the developmental tree; the latter is shown schematically in grey lines in Figure 3

for details on the 11 genes). Figure 1 shows the resulting 11 gene regulatory network. Importantly, as all 11 regulatory elements have been studied extensively using DNA/protein-binding assays, as well as reporter gene assays of wild-type and mutant elements, both the direction and value of each of the regulatory interactions is known with certainty. Moreover, protein-protein interactions curated from the literature were included, such as the well-characterized Gata1-Pu.1 interactions whenever their value (activatory/inhibitory) was known (see Supplementary Table S2 for details).

The resulting network was modelled as logical interactions encoding the activating and/or inhibitory links, including the specific combinations in which particular interactions occur (e.g. Gata2 and Scl together activate Eto2). This logical model was implemented in advanced Boolean notation, as described in Section 2 and shown in Figure 1 (see Supplementary Table S2 for a full network description). Several observations are noteworthy:

(i) a network of 11 genes with three types of possible interactions (activatory, inhibitory and none) could adopt in excess of 10^{50} possible network topologies. It would, therefore, simply be unfeasible to perform modelling analysis using all possible topologies and then work backwards to identify the likely correct topology. (ii) At the heart of the network lies the triad of Scl, Gata2 and Fli1, which is characterized by extensive positive feedback loops, but negative regulatory interactions are common outside this central triad. (iii) We have 11 genes connected by 47 links (an average degree of 4.3) forming a densely connected network. Within this network, we can identify an even more densely connected core consisting of Erg, Gata2, Scl and Fli1 with an average degree of 8.5. Furthermore, Gata2 and Scl connect out to most other genes, and nearly always operate together as a dimer.

3.2 Network genes are expressed dynamically during haematopoiesis

In order for a network model to be useable as a predictive tool, the behaviour of its component genes needs to be assessed using available experimental data. We, therefore, explored the expression patterns of the 11 component genes in primary haematopoietic cell types. To this end, we took advantage of two published datasets: a single-cell gene expression profiling study comparing haematopoietic stem with progenitor cells (Ramos *et al.*, 2006) and the haematopoietic fingerprints database, a collection of expression profiling data for HSPCs, as well as nine differentiated lineages (Chambers *et al.*, 2007). Based on the available literature, all our HSPC network genes except Gata1 should be expressed in the most immature stem cell population, which is precisely what we found when interrogating the two expression profiling datasets. Moreover, Gata1 expression was found in the immediate progeny of the most immature progenitors, e.g. the multipotent progenitor population. In contrast to the ubiquitous expression of our 11 genes in the stem/progenitor compartment, mature blood lineages only express subsets of the 11 genes that make up the HSPC network ranging from 2 of 11 in activated CD8 T-cells to 7 of 11 in granulocytes. Of note, different mature cell types express different subsets of genes, which prompted us to investigate whether this variability would be sufficient to at least partially reconstruct a haematopoietic differentiation tree. Indeed, clustering based on expression of these 11 genes was sufficient to capture key aspects of the haematopoietic differentiation tree (Fig. 1B). Our HSPC network model may, therefore, not only reveal properties of the stem cell state but also allow us to interrogate potential mechanisms and external stimuli that direct stem cell differentiation into specific mature lineages.

3.3 Dynamic modelling of the network predicts heterogeneous HSPC expression states

Having generated a complex vertebrate transcriptional regulatory network model based on comprehensive experimental evidence, we next performed dynamic modelling analysis to explore whether any predicted network behaviour would allow us to gain new insights into blood stem cell biology. Dynamic modelling revealed that the experimentally validated network topology allows for three stable states (Fig. 2A):

- (i) All genes are off (S-3-1),

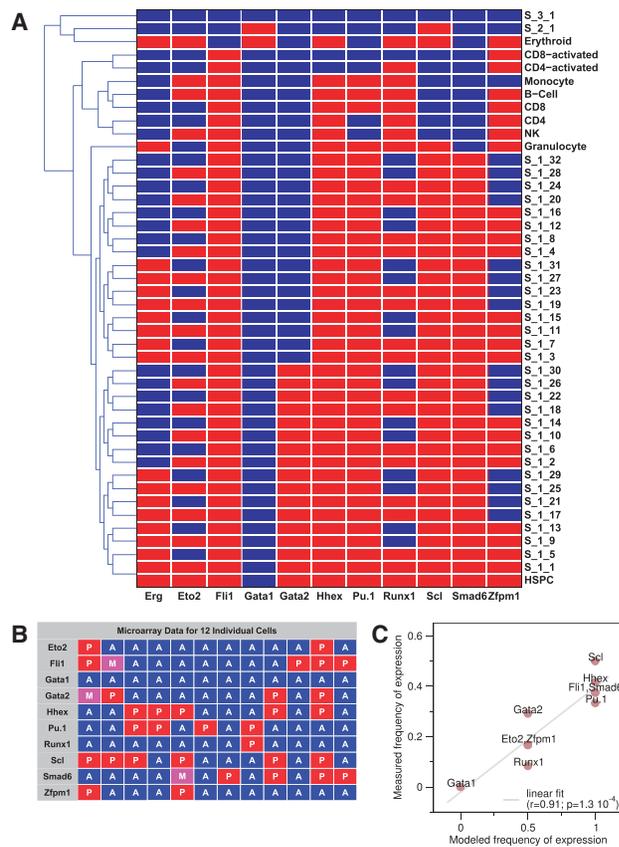


Fig. 2. Steady-state analysis and comparison with expression patterns in 10 haematopoietic cell types. (A) The relation between the expression patterns of the 10 major cell types and the steady-states from the network model is shown by means of hierarchical clustering. Cell types are identified by their names. Steady-states are labelled with ‘S’ and two numbers; the first indicates the steady-state (1, 2 or 3) and the second the sub-states within the steady-state (up to 32 for steady-state ‘S-1’). Red, expression present; blue, expression absent. (B) Heterogeneous gene expression observed in single-cell microarray experiments of 12 individual HSPCs for all genes in our network except Erg (from Ramos *et al.*, 2006). Red, expression present; blue, expression absent; magenta, marginal expression. (C) A near-linear correlation of averaged gene expression activity from the 12 single-cell profiles from (B) compared with average gene activity from the modelled HSPC steady-state individually for each of the 10 genes included in (B)

- (ii) Only Gata1 and Scl are expressed (S-2-1) and
- (iii) An interconnected set of 32 expression states with multiple genes active but Gata1 always repressed (S-1-1 to 32).

To explore whether these steady-states matched observed cell states, we next performed clustering of expression patterns from our stable states together with the expression patterns in the 10 haematopoietic cell types. The results shown in Figure 2A indicate that steady-state S-3-1 corresponds to a non-haematopoietic cell. The steady-state S-2-1 clusters with the erythroid cell profile (Fig. 2A), but they are not identical. Importantly, however, the hamming distance is much smaller if we take into account the fact that the benchmark expression data of Chambers *et al.* (2007) that we used here represents a mixture of mature and

immature erythrocytes. It has been shown that during final maturation, erythrocytes will downregulate Erg, Hhex and Runx1 (Lorsbach *et al.*, 2004; Merryweather-Clarke *et al.*, 2011; Seita *et al.*, 2012; Wong *et al.*, 2011).

Most interesting, however, is stable state S-1, which is composed of 32 interconnected internal states, including a state that matches the expected pattern for HSPCs. This suggests that the precursor HSPC is not a homogeneous cell population, but rather is composed of cells in different stages of activation. Furthermore, there is a striking correlation between gene expression profiling results from single HSPCs (Ramos *et al.*, 2006, summarized in Fig. 2B) and the heterogeneous states predicted by our network, as those genes predicted by our model to be stably present were consistently found expressed in a high proportion of single-cell profiling experiments, whereas genes predicted to be variable (or ‘oscillatory’) in this stable state by our model were consistently found expressed in fewer single cells (Fig. 2C). This analysis, therefore, not only demonstrates that our knowledge-driven network topology is compatible with expression patterns observed in HSPCs *in vivo* but also suggests that expression of genes, such as Gata2, Zfp1, Erg and Eto2 is heterogeneous in HSPCs and may define intermediate states within this cell population.

3.4 Modelling state transitions reveals possible differentiation triggers and a potential role for expression heterogeneity in stem cell function

Analysis of transitions between different steady-states in the model can be useful to predict experimental conditions for cells to differentiate out of the HSPC state. We analysed all possible state transitions in the context of our model. Most theoretically possible transitions cannot occur with our experimentally informed network topology; of all $2048^2 = 4\,194\,304$ possible paths between the 2048 states in our model, only 895 751 (21%) can be traversed within our network. This result is not unexpected, as cell types should be stable states, and network wiring would be expected to constrain flexibility of regulatory states and thus stabilize cell types. There are no paths out of the HSPC state, which is consistent with the HSPC being a stable cell type within the context of a regulatory network based on HSPC transcription factors.

To further classify the transitions, we next mapped all shortest paths onto the known paths of the haematopoietic hierarchy connecting the 10 cell types profiled by Chambers *et al.* (2007). This allowed us to classify these permitted transitions in our model into three categories:

- (i) There are 11 transition paths that follow the developmental tree to the mature cell types, and all start with the activation or repression of one or more genes by some external stimulus (i.e. not by any of the other genes in the network). We call these transitions ‘on path’, and they are shown in Figure 3. The external activation/repression out of the HSPC state we call the ‘initial trigger’ or ‘push’, with a ‘push distance’ indicating the number of genes that need this activation/repression; these are also shown in Figure 3.

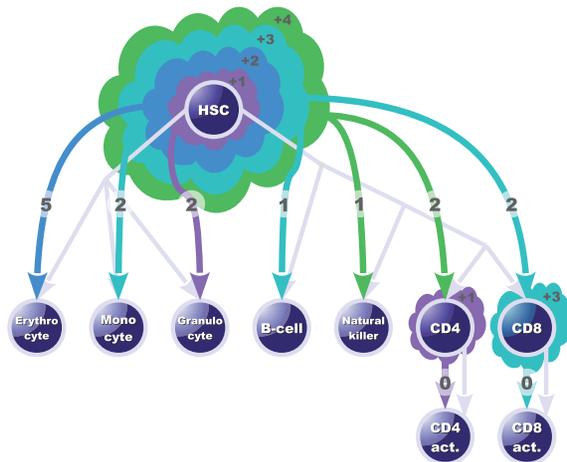


Fig. 3. Analysis of state transitions. Developmental routes (in grey) between the major cell types in the developmental tree, with corresponding ‘on path’ transitions (leading to mature cell types) observed in the modelled network state space indicated as arrows (in colours; numbers indicate path lengths). The ‘on path’ transitions all start with an external trigger from the HSPC cell-type state; this trigger, or ‘push’, changes the state of one (+1) or more (+2, +3 and +4) genes. Similar ‘pushes’ are needed for transitions out of the CD4 and CD8 cell type to their respective activated cell types

- (ii) There are a further 11 transition paths in the reverse direction, which we call ‘upstream’; these reach the HSPC state without requiring a ‘push’ (Supplementary Table S3).
- (iii) There are an additional 18 transition paths that make direct connections between differentiated cell types. These transition paths may provide a way to cross-differentiate between mature cell types without first having to de-differentiate into a stem cell as an intermediate step. We call this third category of transitions ‘cross-path’ (Supplementary Table S3).

This analysis, therefore, demonstrates that our network topology constrains the majority of transitions to be either ‘on path’ or ‘cross-path’; just over half of these transitions are between biologically similar/related cell types, such as monocytes and granulocytes.

We determined for our model, which states closest to the HSPC state connect to each of the mature cell types. For example, for the erythrocyte state, there is a state at a distance of two from HSPC that can differentiate into an erythroblast in another five steps (Fig. 3 and Supplementary Table S3). This observation corresponds to the notion that the transition from HSPC to erythrocyte would need a ‘push’ or ‘trigger’ of repressing *Fli1* and activating *Gata1*, thereby shifting the state two steps away from the HSPC, from which point the system can progress without further interventions into erythrocyte. Examining the other transitions in the development tree, it turns out that all transitions out of HSPC towards a mature cell type need a ‘push’ ranging from +1 (Granulocyte) to +4 (CD4 T-cells and NK cells) as is shown in Figure 3 (Supplementary Table S3 for details). We performed the same analysis for stable state S-1,

using as a starting point any of its 32 sub-states (including the HSPC; see also Fig. 2A). Interestingly, for each target mature cell-type state, we found there exists a transition path involving either a shorter ‘push’ distance, or a shorter transition path after the ‘push’ (see Supplementary Table S3 for details). The heterogeneous stem cell state in our model thus enables more efficient transitions towards different mature cell types. This observation is consistent with a role for expression heterogeneity in stem cells in terms of mediating multi-lineage differentiation potential.

Supplementary Table S3 shows that for specific differentiation directions different transition states can be involved, but at each ‘push distance’, +1, +2, +3 and +4 from the HSPC state, only a single-transition state is used. For example, monocyte, CD8 and B-cell share the same transition state (at +3 from the HSPC), as do NK and CD4 (+4). Taken together, this analysis supports the notion that the HSPC state is highly stabilized when encoded using our experimentally informed network topology, and that the system first needs to be perturbed to initiate differentiation into specific lineages. Intriguingly, this notion of destabilization has also been put forward in the experimental study of Pina *et al.* (2012), who observe that early exit of HSPCs from the self-renewal state is not yet linked with specific gene expression changes that would commit the cell to differentiate down a single trajectory. The analysis of dynamic transitions in a regulatory network model, such as ours, permits the formulation of hypotheses on how this destabilization may be triggered, and how it may lead to subsequent entry into commitment towards the different lineages.

3.5 Increasing the ‘power’ of *Gata1* results in a ‘one-step’ trigger for differentiation towards erythrocyte

We chose the differentiation pathway towards erythroid cells for further investigation because (i) the pathway is well characterized at the experimental level; (ii) it has been the subject of modelling approaches based on simple two-gene interactions (Chickarmane *et al.*, 2009; Roeder and Glauche, 2006); and (iii) it connects two stable states in our modelling based on our 11 gene network.

Of note, experimental evidence suggests that a single ‘trigger’ or ‘push’ (e.g. ectopic expression of *Gata1*) would be sufficient to drive immature blood progenitors towards an erythroid fate (Heyworth *et al.*, 2002; Kulesa *et al.*, 1995). However, as noted earlier in the text, our modelling results suggest that HSPC cells need to undergo two-state changes or ‘pushes’ as a trigger to differentiate into erythroid cells. We considered that there might be two possible explanations for this discrepancy with our model: (i) *Gata1* regulates a protein not present in our network, and this can generate this second ‘push’ or (ii) there is a missing link in our wiring diagram, which when introduced would increase the ‘power’ of *Gata1* so that its ectopic expression would become a single push differentiation trigger. Interrogating the first of these two possibilities is potentially rather speculative, but the second could be readily explored.

We therefore considered potentially missing network links from our current topology. In particular, we extended our model by introducing the possible repression of *Fli1* by *Gata1* based on the rationale that the *Fli1* regulatory element is

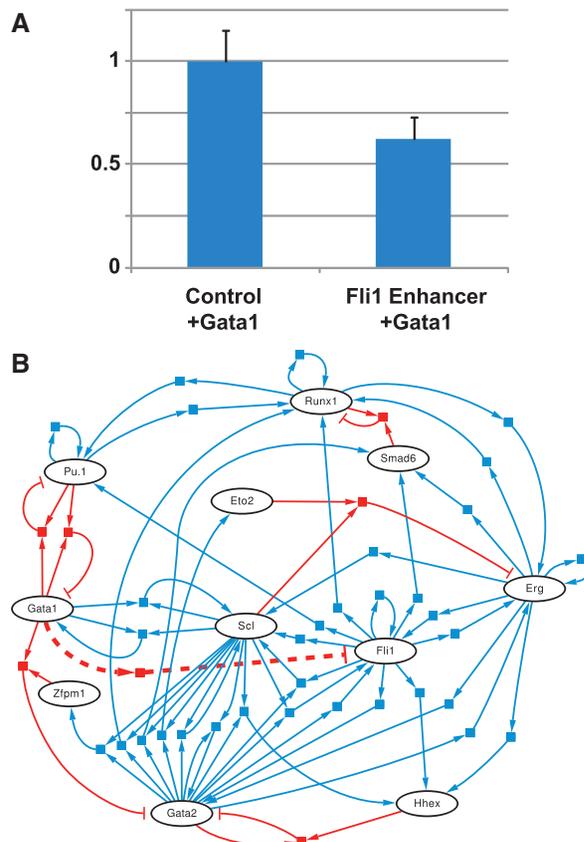


Fig. 4. Gata1 inhibits activity of the Fli1 HSPC enhancer. **(A)** Co-transfection of the Fli1 enhancer construct with a Gata1 expression vector results in significant reduction of the Fli1 enhancer activity. Co-transfection studies were performed in the HSPC cell line HPC7. The data shown represent the average fold change of four individual experiments, each performed in triplicate. **(B)** Diagram of the gene regulatory network, compare Figure 1 showing the predicted and experimentally validated inhibition of Fli1 by Gata1 (dashed lines)

structurally similar to the Gata2 element, which is known to be repressed by Gata1 (Grass *et al.*, 2003). Interestingly, just introducing this single additional repressive link elevated Gata1 to a ‘single push’ trigger for erythroid differentiation. Following on from this modelling result, we investigated whether Gata1 was indeed able to repress activity of the Fli1 enhancer in blood stem/progenitor cells. To test this, the haematopoietic progenitor cell line HPC7 was electroporated with a luciferase reporter construct containing the Fli1 enhancer together with either an empty control plasmid or a Gata1 overexpression construct. As shown in Figure 4A, co-transfection of the Gata1 expression plasmid resulted in significant repression of the activity of the Fli1 enhancer construct, thus demonstrating that Gata1 is indeed able to negatively regulate expression of Fli1. Furthermore, analysis of Gata1 ChIP-Seq data in primary fetal liver blood progenitor cells co-expressing Gata1 and Fli1 (NCBI GEO Accession GSM923586) showed direct binding of Gata1 to the +12kb HSPC enhancer region, as well as the Fli1 promoter (Supplementary Fig. S2).

Network transition modelling, therefore, allowed us to predict a previously unrecognized network link, which we were able to validate experimentally. The revised network diagram is shown in Figure 4B with the new repressive link indicated by dashed lines. Interestingly, including repression of Fli1 by Gata1 did not alter the steady-states of our model, illustrating how some network links specifically influence transitions between states rather than the states themselves.

4 DISCUSSION

The construction of accurate regulatory network models is an essential prerequisite towards gaining a systems-level understanding of the transcriptional control of complex cellular behaviour. Here, we have generated a regulatory network model for HSPCs based on comprehensive experimental data, which represents the most complex mammalian network model to date anchored on *cis*-regulatory functional data. This experimentally validated network topology generated three stable states, one of which was composed of 32 interconnected internal states, including the one that matched the stem cell expression pattern. Binary on/off expression of an 11 gene network could theoretically generate 2048 possible expression states. The fact that we identify only 34 states thus highlights how network modelling based on experimental data can serve to reduce the complexity of analysing multi-gene interactions. Analysis of state-space transitions identified potential triggers that might mediate exit from the stem cell state and highlighted a previously unrecognized inhibition of Fli1 by Gata1, which was subsequently validated experimentally.

4.1 Experimentally validated network models—insights and open questions

Regulatory network topology determines the nature of possible regulatory states, as well as the possible transitions paths between them. The experimental evidence used for model construction is, therefore, critical. Previous studies in lower model organisms have made extensive use of comprehensive gene regulatory experimental data anchored on the interactions of upstream regulators with specific gene regulatory sequences (Davidson, 2006; Davidson *et al.*, 2002; Smith *et al.*, 2007). By contrast, recent network models for mammalian systems, including blood (Krumsiek *et al.*, 2011) relied on less explicit experimental data. Direct experimental knowledge of the interactions within our network model not only provides high confidence in the modelling but also offers an opportunity to consider the possible consequences if our experimental knowledge was more limited. For example, without the repression of Erg by Scf there would only be 16 rather than 32 internal sub-states in steady-state 1. Importantly, introducing the novel interaction generates internal states that are closer to some of the differentiated states. Consequently, the number of internal states that a stem cell can ‘explore’ increases with a concurrent decrease in the number of external triggers required to move out of the HSPC state to differentiate.

Another notable observation is that most repressive interactions in the network (Fig. 1) arise from pairs of genes. A common theme here is that co-regulators, such as Eto2 and Zfp1, are thought to bind DNA indirectly through interactions

with conventional transcription factors, such as *Sc1* and *Gata1*, and by doing so convert the latter from activators to repressors. Interestingly, in our network, these negative co-regulators are themselves activated by the conventional TFs, thus generating an abundance of incoherent feed-forward loops within the wider network. Simple negative feedback loops have previously been proposed to result in oscillatory expression of important cell fate regulators (Hirata *et al.*, 2002; Lahav *et al.*, 2004). To better understand the potential for oscillatory behaviour in increasingly complex networks, future developments might need to include building more fine-grained models, such as the use of Petri nets, which can be readily adapted to move from a Boolean range of values towards discrete multi-valued expression levels (Bonzanni *et al.*, 2009a, b).

Within the context of our 11 gene HSPC network topology, several expression states that correspond to the differentiated cell types shown in Figure 3 can automatically revert to the stem cell state, suggesting a potential for spontaneous reversion of differentiated cells to the immature stem cell phenotype (details in Supplementary Table S3). In a sense, this may merely be a reflection of the fact that our experimentally informed HSPC network topology generated a stable HSPC attractor. However, it also suggests that ‘commitment features’ that would block these regressions, may be missing from our network. A recent model of the myeloid lineage (Krumsiek *et al.*, 2011), which did not include the stem cell state, found the mature cell types (erythroid, megakaryocyte, monocyte and granulocyte) to be attractor states. A likely explanation for the contrast between this study and our findings may be that rather than excluding the stem cell state, we explicitly focused on regulatory interactions within HSPCs. Multiple positive feedback loops, therefore, stabilize the HSPC state in our model, whereas the external triggers that ‘break’ some of these feedback loops and thus induce differentiation remain unknown. It is likely that some of these commitment events will transmit extracellular signals to the nucleus, to modulate epigenetic processes that regulate the availability of regulatory regions for transcription factor binding. For example, epigenetic silencing of a given regulatory element could prevent access of upstream factors with the consequence of ‘locking in’ the differentiated state.

4.2 The ‘stem cell state’—a moving target?

Comprehensive exploration of the state space dictated by our experimentally validated HSPC network topology resulted in a set of 32 interconnected states, which together constitute a stable state with a gene expression pattern consistent with HSPCs. However, only a single internal state in the HSPC attractor matched expression levels of all HSPC associated genes, whereas all others expressed different subsets of genes, suggesting possible heterogeneity between discrete expression states. The heterogeneous steady-state predicted by our model might at first have been considered an artefact because of either the unavoidably partial knowledge we have about the system, or introduced by the high level of discretization used (i.e. from potentially continuous expression levels to Boolean values). However, we believe that on the contrary, our results may provide potentially important new insights into the nature of transcriptional control of stem cells and differentiation as outlined below: first, the striking

correlation between gene expression profiling results from single HSPCs and the heterogeneous states predicted by our network (Fig. 2C). Moreover, single-cell analysis of highly purified murine HSPCs using digital polymerase chain reaction (PCR) assays (Warren *et al.*, 2006) also showed heterogeneous transcription factor expression in individual HSPCs. Taken together, these observations suggest that the stem cell state is composed of a discrete set of sub-states with a substantial degree of oscillations in gene expression, which includes genes thought of as central regulators of stem cell fate. Of note, this concept is largely consistent with the recently introduced theory of non-genetic micro-heterogeneity in multi-potential stem cell populations (Huang, 2009).

It might at first glance seem difficult to reconcile such oscillations and the resultant transcriptional heterogeneity with the model of multi-lineage priming. This latter concept was founded on the observation that some HSPCs display low-level co-expression of cytokine receptor genes affiliated with divergent differentiation pathways (Hu *et al.*, 1997). Consequently, HSPCs have widely been thought of as highly promiscuous with widespread co-expression rather than only expressing subsets of genes. However, in addition to demonstrating the potential for multi-lineage priming, the original article in 1997 (Hu *et al.*) also found heterogeneous expression of stem cell affiliated genes when analysed at the single-cell level. Both multi-lineage priming of cytokine receptor genes and expression of HSPC affiliated transcription factors, therefore, show cellular heterogeneity consistent with oscillating expression in individual HSPCs. Based on the results presented in this article, cellular heterogeneity of multi-lineage priming may, therefore, be hardwired into HSPC regulatory networks rather than being a consequence of low-level, non-specific gene expression noise as had been speculated previously. This in turn would suggest that characterization of the underlying mechanisms will provide novel insights into the functional role of multi-lineage priming as a key mediator of differentiation. Rather than there being a ‘stem cell continuum’, the regulatory space within which a stem cell can move may be constrained where a given differentiation trajectory requires passage through a number of specific intermediate states.

Other recent work also challenges the notion of a stem cell continuum in multi-potential stem cell populations and multi-lineage priming, but instead offers a scenario with multiple ‘discordant’ entries into lineages and subsequent ‘coalescence’ into mature expression patterns (Pina *et al.*, 2012). In analogy to this, we see a heterogeneous stem cell state that offers several routes into distinct lineage-specific transition states, which would be consistent with the notion of ‘multiple discordant entries’. Our model also suggests the possibility of triggering cross-lineage transitions, which may be exceedingly rare in normal cells but have been observed experimentally (Di Tullio *et al.*, 2011) and in leukaemias (van Wering *et al.*, 1995). For example, a leukaemia may be of myeloid phenotype when a patient first presents, but of lymphoid phenotype at relapse (Chucrallah *et al.*, 1995; Stass *et al.*, 1984). A better understanding of cross-lineage transition paths may, therefore, aid to develop therapies for relapsed patients, who currently have a poor prognosis. Cross-lineage transitions may also be exploited in the field of regenerative medicine, where protocols are being developed to for example make macrophages out of B-cells (Bussmann *et al.*, 2009).

The lack of explicit commitment in the mature cell types in our model, as discussed earlier in the text, is consistent with the notion that entry into a lineage may at first be reversible. This is in line with findings from a recent model of the myeloid lineage that exhibits a heterogeneous entry into mature cell type attractor states (Krumstiek *et al.*, 2011). In many cases, the particular order of external triggers applied in our model to exit the HSPC state seems not to be critical. That is, along the ‘pushes’ to different distances we do not observe overlap, except for where both lead to the same intermediate state (Supplementary Table S3). Similarly multiple transition paths to mature cell type states show order-independence of individual genes switched, consistent with the notion of network coalescence (Tipping *et al.*, 2009). Thus, the emerging picture seems to be that, starting from a heterogeneous HSPC stable state, external stimuli may trigger different initial responses within individual cells in a heterogeneous stem cell population, but ultimately resolve into a clearly demarcated mature cell state.

4.3 Discrete stem cell states and differentiation triggers

As the stem cell state space is composed of a set of regulatory states with inter-conversions between them dictated by the network topology, the question arises to what extent knowledge of network wiring may increase our ability to manipulate stem cell fate choices. In this study, we show that specific differentiation triggers can be modelled successfully and inform specific hypotheses for subsequent experimental testing. Importantly, specific sub-states within the stem cell state are closer to certain downstream cellular fates than others; indeed fewer activating triggers (‘pushes’) are needed and shorter transition paths exist when starting from these sub-states. This in turn suggests that the distribution of stem cell internal states has the potential to influence the propensity of a stem cell to choose between divergent differentiation choices. A mechanistic understanding of the underlying processes would have important scientific and clinical implications. For example, altering the levels of Gata2 has recently been shown to affect the ratio between cycling and quiescent HSPCs (Tipping *et al.*, 2009), providing direct experimental evidence that levels for one of the factors shown to be oscillating in our network model are associated with phenotypically identifiable sub-states of HSPCs. From a translational point of view, *in vitro* production of specific blood cell types from HSPCs has the potential to provide safer and cheaper alternatives to blood transfusions. However, directed differentiation *in vitro* remains disappointingly inefficient, suggesting that knowledge of the underlying regulatory networks is critical for the development of new protocols. Finally, treatment responses for patients carrying the same leukaemogenic mutations can be different. As many leukaemia oncogenes cause a differentiation block of early progenitors, it is possible that this block may occur in different ‘sub-states’ of the stem cell compartment in different patients, suggesting that a deeper understanding of these sub-states may provide novel treatment options.

ACKNOWLEDGEMENT

The authors would like to thank Mark Ibberson for reading the manuscript and providing helpful comments.

Funding: European Commission FP6 project ENFIN (Experimental Network for Functional INtegration). Computations were performed at the Vital-IT Center of the Swiss Institute of Bioinformatics and the Centre for Integrative Bioinformatics (IBIVU) at VU University Amsterdam. Experimental work was supported by Leukaemia & Lymphoma Research (UK), and the Leukemia & Lymphoma Fund (USA).

Conflict of Interest: none declared

REFERENCES

- Bockamp, E.O. *et al.* (1998) Transcriptional regulation of the stem cell leukemia gene by PU.1 and Elf-1. *J. Biol. Chem.*, **273**, 29032–29042.
- Bonzanni, N. *et al.* (2009a) Executing multicellular differentiation: quantitative predictive modelling of *C.elegans* vulval development. *Bioinformatics*, **25**, 2049–2056.
- Bonzanni, N. *et al.* (2009b) What can formal methods bring to systems biology?. In: *FM 2009: Formal Methods, Lecture Notes in Computer Science*. Vol. 5850, Berlin/Heidelberg, Springer, pp. 16–22.
- Bussmann, L.H. *et al.* (2009) A robust and highly efficient immune cell reprogramming system. *Cell Stem Cell*, **5**, 554–566.
- Chambers, S.M. *et al.* (2007) Hematopoietic fingerprints: an expression database of stem cells and their progeny. *Cell Stem Cell*, **1**, 578–591.
- Chickarmane, V. *et al.* (2009) Computational modeling of the hematopoietic erythroid-myeloid switch reveals insights into cooperativity, priming, and irreversibility. *PLoS Comput. Biol.*, **5**, e1000268.
- Chucrallah, A.E. *et al.* (1995) Adult acute lymphoblastic leukemia at relapse. Cytogenetic, immunophenotypic, and molecular changes. *Cancer*, **76**, 985–991.
- Davidich, M.I. and Bornholdt, S. (2008) Boolean network model predicts cell cycle sequence of fission yeast. *PLoS One*, **3**, e1672.
- Davidson, E. (2006) *The Regulatory Genome: Gene Regulatory Networks In Development And Evolution*. Academic Press, San Diego.
- Davidson, E.H. *et al.* (2002) A genomic regulatory network for development. *Science*, **295**, 1669–1678.
- Di Tullio, A. *et al.* (2011) CCAAT/enhancer binding protein alpha (C/EBP(alpha))-induced transdifferentiation of pre-B cells into macrophages involves no overt retrodifferentiation. *Proc. Natl Acad. Sci. USA*, **108**, 17016–17021.
- Donaldson, I.J. *et al.* (2005a) Genome-wide identification of cis-regulatory sequences controlling blood and endothelial development. *Hum. Mol. Genet.*, **14**, 595–601.
- Donaldson, I.J. *et al.* (2005b) Genome-wide identification of cis-regulatory sequences controlling blood and endothelial development. *Hum. Mol. Genet.*, **14**, 595–601.
- Foster, S.D. *et al.* (2009) From genes to cells to tissues—modelling the haematopoietic system. *Mol. Biosyst.*, **5**, 1413–1420.
- Garg, A. *et al.* (2008) Synchronous versus asynchronous modeling of gene regulatory networks. *Bioinformatics*, **24**, 1917–1925.
- Garg, A. *et al.* (2009) Modeling stochasticity and robustness in gene regulatory networks. *Bioinformatics*, **25**, i101–i109.
- Georgescu, C. *et al.* (2008) A gene regulatory network armature for T lymphocyte specification. *Proc. Natl Acad. Sci. USA*, **105**, 20100–20105.
- Göttgens, B. (2004) Transcriptional regulation of haematopoiesis. *Vox Sang.*, **87** (Suppl. 1), 15–19.
- Göttgens, B. *et al.* (1997) Transcription of the SCL gene in erythroid and CD34 positive primitive myeloid cells is controlled by a complex network of lineage-restricted chromatin-dependent and chromatin-independent regulatory elements. *Oncogene*, **15**, 2419–2428.
- Göttgens, B. *et al.* (2002) Establishing the transcriptional programme for blood: the SCL stem cell enhancer is regulated by a multiprotein complex containing Ets and GATA factors. *EMBO J.*, **21**, 3039–3050.
- Göttgens, B. *et al.* (2004) The scl +18/19 stem cell enhancer is not required for hematopoiesis: identification of a 5′ bifunctional hematopoietic-endothelial enhancer bound by Fli-1 and Elf-1. *Mol. Cell. Biol.*, **24**, 1870–1883.
- Grass, J.A. *et al.* (2003) GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. *Proc. Natl Acad. Sci. USA*, **100**, 8811–8816.
- Heyworth, C. *et al.* (2002) Transcription factor-mediated lineage switching reveals plasticity in primary committed progenitor cells. *EMBO J.*, **21**, 3770–3781.

- Hirata, H. et al. (2002) Oscillatory expression of the bHLH factor Hes1 regulated by a negative feedback loop. *Science*, **298**, 840–843.
- Hu, M. et al. (1997) Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev.*, **11**, 774–785.
- Hu, Z. et al. (2007) Genetic reconstruction of a functional transcriptional regulatory network. *Nat. Genet.*, **39**, 683–687.
- Huang, S. (2009) Non-genetic heterogeneity of cells in development: more than just noise. *Development*, **136**, 3853–3862.
- Karlebach, G. and Shamir, R. (2008) Modelling and analysis of gene regulatory networks. *Nat. Rev. Mol. Cell Biol.*, **9**, 770–780.
- Kauffman, S. et al. (2003) Random Boolean network models and the yeast transcriptional network. *Proc. Natl Acad. Sci. USA*, **100**, 14796–14799.
- Klamt, S. et al. (2006) A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics*, **7**, 56.
- Kobayashi-Osaki, M. et al. (2005) GATA motifs regulate early hematopoietic lineage-specific expression of the Gata2 gene. *Mol. Cell. Biol.*, **25**, 7005–7020.
- Krumsiek, J. et al. (2011) Hierarchical differentiation of myeloid progenitors is encoded in the transcription factor network. *PLoS One*, **6**, e22649.
- Kulesa, H. et al. (1995) GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblats, and erythroblats. *Genes Dev.*, **9**, 1250–1262.
- Lahav, G. et al. (2004) Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nat. Genet.*, **36**, 147–150.
- Landry, J.R. et al. (2008) Runx genes are direct targets of Scl/Tall in the yolk sac and fetal liver. *Blood*, **111**, 3005–3014.
- Li, F. et al. (2004) The yeast cell-cycle network is robustly designed. *Proc. Natl Acad. Sci. USA*, **101**, 4781–4786.
- Lorsbach, R.B. et al. (2004) Role of RUNX1 in adult hematopoiesis: analysis of RUNX1-IRES-GFP knock-in mice reveals differential lineage expression. *Blood*, **103**, 2522–2529.
- Mangla, K. et al. (2010) Timing robustness in the budding and fission yeast cell cycles. *PLoS One*, **5**, e8906.
- Mendoza, L. and Xenarios, I. (2006) A method for the generation of standardized qualitative dynamical systems of regulatory networks. *Theor. Biol. Med. Model.*, **3**, 13.
- Merryweather-Clarke, A.T. et al. (2011) Global gene expression analysis of human erythroid progenitors. *Blood*, **117**, 96–108.
- Miranda-Saavedra, D. and Göttgens, B. (2008) Transcriptional regulatory networks in haematopoiesis. *Curr. Opin. Genet. Dev.*, **18**, 530–535.
- Nottingham, W.T. et al. (2007) Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer. *Blood*, **110**, 4188–4197.
- Okuno, Y. et al. (2005) Potential autoregulation of transcription factor PU.1 by an upstream regulatory element. *Mol. Cell. Biol.*, **25**, 2832–2845.
- Pimanda, J.E. et al. (2007) Gata2, Fli1, and Scl form a recursively wired gene-regulatory circuit during early hematopoietic development. *Proc. Natl Acad. Sci. USA*, **104**, 17692–17697.
- Pina, C. et al. (2012) Inferring rules of lineage commitment in haematopoiesis. *Nat. Cell Biol.*, **14**, 287–294.
- Ramos, C.A. et al. (2006) Evidence for diversity in transcriptional profiles of single hematopoietic stem cells. *PLoS Genet.*, **2**, e159.
- 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.
- Seita, J. et al. (2012) Gene Expression Commons: an open platform for absolute gene expression profiling. *PLoS One*, **7**, e40321.
- Sieweke, M.H. and Graf, T. (1998) A transcription factor party during blood cell differentiation. *Curr. Opin. Genet. Dev.*, **8**, 545–551.
- Smith, J. et al. (2007) A gene regulatory network subcircuit drives a dynamic pattern of gene expression. *Science*, **318**, 794–797.
- Spooner, C.J. et al. (2009) A recurrent network involving the transcription factors PU.1 and Gfi1 orchestrates innate and adaptive immune cell fates. *Immunity*, **31**, 576–586.
- Stass, S. et al. (1984) Lineage switch in acute leukemia. *Blood*, **64**, 701–706.
- Swiers, G. et al. (2006) Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. *Dev. Biol.*, **294**, 525–540.
- Thomas, R. (1991) Regulatory networks seen as asynchronous automata: a logical description. *J. Theor. Biol.*, **153**, 1–23.
- Thomas, R. et al. (1995) Dynamical behaviour of biological regulatory networks—I. Biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bull. Math. Biol.*, **57**, 247–276.
- Thoms, J.A. et al. (2011) ERG promotes T-acute lymphoblastic leukemia and is transcriptionally regulated in leukemic cells by a stem cell enhancer. *Blood*, **117**, 7079–7089.
- Tipping, A.J. et al. (2009) High GATA-2 expression inhibits human hematopoietic stem and progenitor cell function by effects on cell cycle. *Blood*, **113**, 2661–2672.
- van Wering, E.R. et al. (1995) Immunophenotypic changes between diagnosis and relapse in childhood acute lymphoblastic leukemia. *Leukemia*, **9**, 1523–1533.
- Vyas, P. et al. (1999) Different sequence requirements for expression in erythroid and megakaryocytic cells within a regulatory element upstream of the GATA-1 gene. *Development*, **126**, 2799–2811.
- Warren, L. et al. (2006) Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. *Proc. Natl Acad. Sci. USA*, **103**, 17807–17812.
- Wilson, N.K. et al. (2009) The transcriptional program controlled by the stem cell leukemia gene Scl/Tall during early embryonic hematopoietic development. *Blood*, **113**, 5456–5465.
- Wong, P. et al. (2011) Gene induction and repression during terminal erythropoiesis are mediated by distinct epigenetic changes. *Blood*, **118**, e128–138.