

HEMATOPOIETIC STEM CELLS

Regulatory network control of blood stem cells

Berthold Göttgens

Department of Haematology, Cambridge Institute for Medical Research and Wellcome Trust & Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, Cambridge, United Kingdom

Hematopoietic stem cells (HSCs) are characterized by their ability to execute a wide range of cell fate choices, including self-renewal, quiescence, and differentiation into the many different mature blood lineages. Cell fate decision making in HSCs, as indeed in other cell types, is driven by the interplay of external stimuli and intracellular regulatory programs. Given the pivotal nature of HSC decision making for both normal and aberrant hematopoiesis,

substantial research efforts have been invested over the last few decades into deciphering some of the underlying mechanisms. Central to the intracellular decision making processes are transcription factor proteins and their interactions within gene regulatory networks. More than 50 transcription factors have been shown to affect the functionality of HSCs. However, much remains to be learned about the way in which individual factors are connected within wider

regulatory networks, and how the topology of HSC regulatory networks might affect HSC function. Nevertheless, important progress has been made in recent years, and new emerging technologies suggest that the pace of progress is likely to accelerate. This review will introduce key concepts, provide an integrated view of selected recent studies, and conclude with an outlook on possible future directions for this field. (*Blood*. 2015;125(17):2614-2620)

Building blocks of transcriptional regulatory networks

The primary components of transcriptional regulatory networks are transcription factor (TF) proteins and the gene regulatory DNA sequences that they bind to.¹ By binding to specific DNA sequence motifs within gene regulatory regions, TF proteins are central players for this primary step of decoding gene regulatory instructions. TF proteins typically contain a number of distinct modules, such as DNA binding, transcriptional activation, and protein/protein interaction domains, with the latter 2 being essential for the recruitment of the basal transcriptional machinery and the assembly of higher-order TF complexes. Based on sequence similarity within the DNA binding domain, TF proteins can be categorized into distinct families, such as homeobox, basic helix-loop-helix, or zinc finger TFs. Members of a given TF family often bind to similar DNA sequences, and protein-protein interactions are common both within and between the different TF families.

Individual TFs bind short sequence motifs that are often no longer than 4 to 6 bp. Any given 6-bp sequence will occur on average approximately once every 4000 bp. Consequently, there will be ~750 000 occurrences just by chance within the 3 000 000 000 bp of the human genome. The number of possible binding sites therefore far exceeds the 20 000 or so human genes, and it has long been assumed that only a small minority of all motif instances play a role in transcriptional regulation. Functional gene regulatory regions should therefore display characteristics that go beyond the mere occurrence of a specific sequence motif. One such characteristic is the presence of clusters of binding sites, typically for up to 5 different TFs. Clustering of binding sites and the prevalence of protein/protein interaction domains within TF proteins facilitate the assembly of higher-order complexes, the formation of which is thought to represent a key step in the decoding of gene regulatory information.

Binding of multiple TFs to individual gene regulatory elements is consistent with the notion of complex networks of interactions, where the locus of each gene participating in the network receives inputs from multiple upstream regulators. Moreover, binding of multiple TFs to a given regulatory element commonly involves synergistic or antagonistic interactions mediated through higher-order complexes. In terms of network structure therefore, inputs from an upstream regulator to its downstream target are rarely simple linear connections, but instead entail synergistic or antagonistic cross-talk between 2 or more upstream regulators. Experimental evidence suggests that some TF complexes exist in solution in the nucleoplasm, and therefore presumably preform (at least partially) before binding to DNA.² There is also evidence for sequential TF binding, where binding of a so-called “settler” TF requires prior binding of “pioneer” TFs.³ Importantly, a given TF may act as a pioneer TF only on some, but not all, the regulatory elements that it binds to. Rather than being based on generalized assumptions, reconstruction of regulatory network models therefore will require mechanistic knowledge of the transcriptional mechanisms operating at individual gene regulatory elements (Figure 1).

Cooperative binding of TFs to gene regulatory sequences represents the first major information processing event during the regulation of gene expression.¹ This in turn is followed by the recruitment of accessory proteins such as chromatin-modifying enzymes and the multiple components of the RNA polymerase holo-complex.⁴ Additional points of potential regulation on the route toward the production of functional proteins include transcriptional elongation, RNA processing and nuclear export, translation, and post-translational modifications. In the context of transcriptional regulatory networks, the production of transcription factor proteins is of particular interest, as these will feed back directly into the activity of the network.

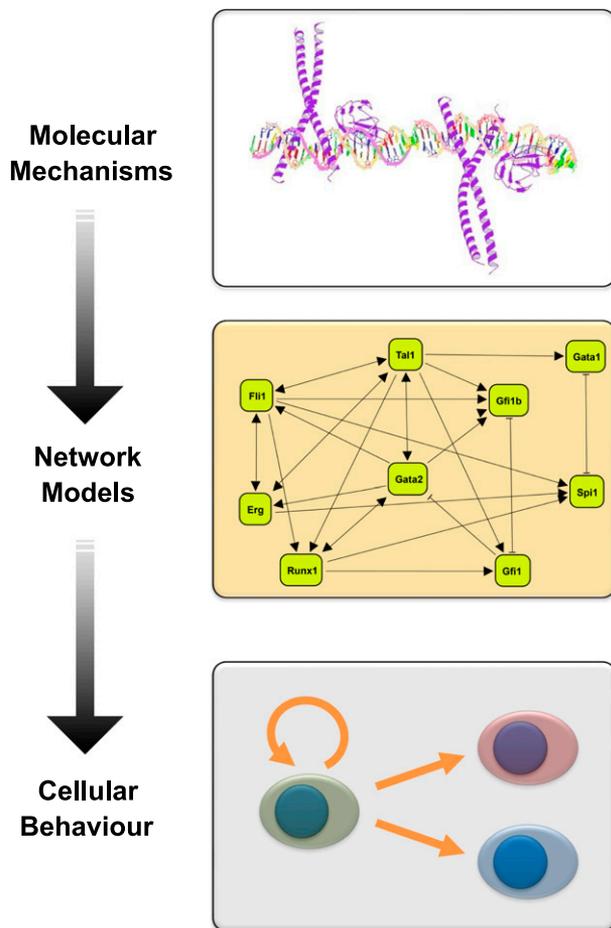


Figure 1. Regulatory network models based on mechanistic insights can advance our understanding of cellular behavior. This diagram illustrates the workflow from experimentally determined molecular mechanisms to regulatory network models that can advance our understanding of cellular decision making processes. Shown are, from top to bottom, a model of the Runx1/CEBP transcription factor complex bound to DNA, a literature curated regulatory network model connecting key HSPC transcription factors, and a diagram illustrating HSC fate decisions including the balance between self-renewal and multilineage differentiation.

Literature-curated hematopoietic transcriptional network models

Given the power of TFs to drive cell fate choice decision making, researchers recognized early on that the construction of models capturing TF interactions might provide a useful tool to improve our understanding of hematopoietic stem and progenitor cell (HSPC) differentiation. However, the generation of detailed functional knowledge on upstream regulator/downstream target gene relationships has traditionally been a laborious task, and therefore no single study would have generated sufficient experimental data to construct comprehensive regulatory network models. In light of these difficulties, several groups recognized the potential power of integrating experimental data from multiple studies to construct hematopoietic regulatory network models, although initial attempts focused on simple 2-factor or 3-factor models.^{5,6} Although these early efforts might have captured some key properties of core regulatory circuits, they clearly oversimplified the complexities of multifactor regulatory interactions, thus limiting their utility to model regulatory network control of cellular behavior. More recently, a qualitative network model connecting 11 transcription

factors controlling the differentiation of common myeloid progenitors (CMPs) was constructed based on a comprehensive survey of the literature.⁷ The investigators encoded the combinatorial logic governing the interactions between these 11 factors using a Boolean modeling approach, where each factor is represented by a variable that can be either “on” or “off,” and the activating/inhibitory interactions are encoded using combinations of the logical functions “And,” “Or,” and “Not.”

Execution of this model, where simulation is started with an expression state representative of the CMP, not only captured putative intermediate stages of myeloid differentiation but was also able to reach 4 distinct terminal steady states, which corresponded in their expression pattern to megakaryocytes, erythrocytes, granulocytes, and monocytes. Moreover, in silico simulation of gene knockouts recapitulated known lineage depletion results, and in silico overexpression reproduced known TF-mediated reprogramming data. Importantly, the CMP progenitor state itself was unstable in this model and would always differentiate spontaneously into 1 of the 4 terminal myeloid differentiation states. The model therefore did not capture those aspects of blood stem cell regulatory networks that mediate the relative stability of the blood stem/progenitor state. A second literature-curated model based on cross-regulation of 10 hematopoietic stem cell (HSC) TFs, on the other hand, identified HSC-like gene expression as a stable attractor state.⁸ However, this stable state now required simulation of external triggers to allow exit toward more differentiated states. Of note, this second model focused on cross-regulation between HSC TFs that resulted in numerous positive feedback loops and therefore presumably stabilized the HSC expression state. Importantly, HSCs have the capacity to balance self-renewal and differentiation, which suggests that neither of the 2 modeling approaches outlined above is yet able to capture the complexities inherent in HSC function. In addition to capturing both the self-renewal and differentiation function of HSCs, future modeling efforts should also attempt to connect HSC regulatory network models with literature curated and validated network models for more differentiated cell lineages such as the myeloid and T-lymphoid models generated by the Singh and Rothenberg groups, respectively.^{9,10}

Inferring transcriptional regulatory networks from microarray gene expression profiles

Because the TF proteins that control gene expression are themselves gene products, the levels of TF protein and target gene mRNA abundance will correlate unless post-translational regulation overrides the link between mRNA levels and active protein concentration. Statistical associations between mRNA levels of candidate upstream regulators and downstream target genes can therefore provide clues about possible gene regulatory hierarchies. Following the introduction of microarray technology for genome-wide measurements of mRNA abundance, researchers quickly realized that such datasets could be used to construct more global “gene network” models that encapsulate the regulatory wiring underlying the gene expression identity of a given cell type. Importantly, simple clustering of genes with similar expression across a set of microarray expression profiles cannot distinguish between direct and indirect regulation as the likely cause of coexpression. More sophisticated tools have therefore been developed to reverse engineer network models from expression profiling data, such as graphical modeling,¹¹ integrative methods that use, for example, protein-protein interaction data,¹² and statistical/information theoretical methods.¹³

Particularly promising results for higher eukaryote systems were obtained following the introduction of the Aracne tool,¹⁴ which reverse engineers transcriptional networks from microarray data by identifying irreducible statistical dependencies, which cannot be explained by any other statistical dependencies and are therefore much more likely to represent direct regulatory interactions in the network. Aracne has been used to reverse engineer regulatory network models for normal B cells,¹⁵ as well as B- and T-cell leukemias.¹⁶⁻¹⁸ Of note, Aracne requires hundreds of individual microarray samples to achieve statistical power and therefore has thus far not been used for the construction of HSPC regulatory networks because the required expression datasets do not exist. Moreover, despite their computational accessibility, predictive network models based on correlations between expressed genes are biased toward detecting positive regulatory connections and find it much harder to distinguish between negative regulation and a lack of coincidence for some other reason. However, there is strong evidence from most developmental systems that have been dissected in depth that negative regulatory interactions are key determinants of the decision making processes when multipotent progenitors choose between alternative cell fates. Computational tools designed to specifically uncover inhibitory interactions from gene expression profiling data have been applied to expression datasets from lower eukaryote model systems¹⁹ but are not yet widely adopted by the hematopoiesis research community.

An alternative approach to dissect the architecture of gene expression programs is the Module Networks algorithm,²⁰ a probabilistic method that identifies so-called regulatory modules from collections of gene expression profiling data. The identified modules not only report sets of coregulated genes, but also their candidate upstream regulators and the cell types where this regulation occurs. As such, this approach is particularly suited to analyze gene expression profiling data that cover the hematopoietic differentiation hierarchy. Its application by Novershtern et al to expression data for 38 distinct human blood progenitor and mature cell types resulted in the identification of 80 expression modules, several of which were associated with specific expression in HSPCs.²¹ In parallel, the authors generated network models based on a combination of TF expression and sequence motif content in promoter regions. Of note, the overlap of predicted regulatory interactions generated by these 2 approaches was minimal, because 70% of TFs covered in both the Module Networks and Promoter Motif Analysis showed no significant expression correlation with the module to which they had been assigned by promoter motif scanning.²¹ One obvious limitation of the promoter-motif based approach is illustrated by previous observations that much of the TF-binding conferring cell type-specific expression occurs at distal sites away from the promoter regions.²² More fundamentally, however, the Novershtern study exemplifies that microarray expression profiling data may yield testable hypotheses about regulatory interactions, but on their own are unlikely to be sufficient for the generation of reliable network models.

Impact of genome-wide TF binding maps for network reconstruction

The apparent problems of reconstructing regulatory networks from expression data alone are possibly most easily explained by the fact that information on actual regulatory elements is missing in this approach. As outlined above, regulatory elements constitute major

building blocks of transcriptional regulatory networks, yet traditional approaches to define their location and function have been slow. The recent advent of high-throughput sequencing technology has greatly accelerated the generation of genome-wide TF binding maps, by coupling chromatin immunoprecipitation to high-throughput sequencing (ChIP-Seq).^{23,24} Due to the rapid decline in sequencing costs, ChIP-Seq technology is now readily available to individual research laboratories, and data integration across laboratories has generated a publicly accessible resource of several hundred TF ChIP-Seq studies across a wide range of hematopoietic cell types.²⁵

However, although ChIP-Seq experiments are increasingly used to complement gene expression profiling,²⁶⁻³⁴ no major progress has yet been reported toward the generation of validated HSPC regulatory network models. This is at least partly due to the observation that TFs are commonly bound to thousands if not tens of thousands of sites in the genome, often near genes that would not be affected by ablation of the factor in question. This has raised the issue of a need to distinguish functionally significant from nonsignificant binding events.³⁵ However, there are no consensus definitions for these terms, thus leaving the community with no reliable strategy to predict whether or not a given TF binding event is functionally relevant. Robust conclusions from ChIP-Seq studies thus far therefore are largely based on statistically significant associations across a range of binding events, where it is clear that as a group, a specific characteristic is enriched, even if no definitive statement can be made about the nature of any given individual binding event. Of relevance to HSPC regulatory networks, statistical analysis has identified previously unknown combinatorial TF interactions between known HSC regulatory TFs³⁶ and global reorganization of HSPC TF assemblies during the specification of blood progenitors from differentiating embryonic stem cells.³⁷ Moreover, a recent comparative analysis of 10 TFs in HSPCs and mast cells showed that HSPC TFs can actively participate in transcriptional programs of HSPCs, as well as mature hematopoietic lineages, largely by binding to distinct sets of target regions.³⁸

To move beyond generalized statistical associations, ChIP-Seq studies need to be complemented with functional analysis of specific binding events, so that putative regulatory relationships inferred from TF binding are functionally validated. TF knockdown followed by gene expression profiling can identify those TF-bound gene loci where reducing levels of the TF causes either down- or upregulation of the putative target gene. However, this approach is not straightforward because (1) compensatory processes may maintain expression levels of true target genes, (2) expression measurements need to be performed shortly after TF knockdown, because otherwise many expression changes will be due to indirect regulation, and (3) sequence motifs with the highest affinity may retain bound TF the longest in an acute knockdown experiment so that the most important target genes may actually change expression last. High-throughput functional assays of promoter and enhancer regions provide a complementary approach that can assign regulatory function to TF-bound regions from ChIP-Seq studies. Of particular interest are recently reported embryonic stem cell differentiation-based assays,^{39,40} as they do not rely on immortalized cell line models but instead allow the reporting of regulatory element functionality in a range of primary cell lineages thus approximating transgenic mouse analysis. Although ChIP-Seq technology on its own did not fulfill its original expectation as the silver bullet for comprehensive identification of transcriptional hierarchies, combination with expression profiling and emerging high-throughput enhancer analysis undoubtedly has the potential to make a real impact in the future to accelerate the construction of comprehensive gene regulatory network models.

Reconstructing regulatory hierarchies from single cell analysis

Blood stem cell research has long made use of single cell assays to address fundamental questions such as the self-renewal and differentiation potential of individual cells within heterogeneous cell populations.⁴¹ As outlined above, gene expression is controlled by complex gene regulatory networks, and gene expression measurements can be used to infer the potential nature of those underlying networks. However, although gene expression measurements in single mammalian cells were pioneered in the hematopoietic system,⁴² these early studies relied on Southern blot hybridization of amplified cDNA 3' ends⁴³ or standard polymerase chain reaction methodology⁴⁴ and therefore were unable to generate datasets of the scale required for network inference. Recent technological advances in microfluidic technology now readily permit quantification of tens to hundreds of genes in hundreds and even thousands of single cells. In addition to providing insights into population heterogeneity and putative transition stages during blood cell differentiation,⁴⁵⁻⁴⁷ the scale of these datasets provides a substrate for robust statistical analysis of gene expression correlation, as each single cell analyzed represents an independent biological measurement.⁴⁸ Moignard et al surveyed the expression of 18 transcription factor genes in 690 single blood stem and progenitor cells,⁴⁷ and went on to exploit pairwise correlation analysis to identify putative regulatory relationships between individual TFs. This analysis highlighted a previously unrecognized triad composed of Gata2, Gfi1, and Gfi1b, which might play an important role in regulating early fate choice decisions of differentiating blood stem cells. Importantly, the authors went on to validate direct regulatory interactions within this triad using a variety of functional assays, thus moving the analysis beyond simple statistical correlation.

A subsequent study profiled >80 TFs in >250 blood stem and progenitor cells⁴⁹ and again used correlation analysis to predict putative regulatory connections. Integration with previously published ChIP-Seq data for 10 key HSPC TFs³⁶ was used to highlight potentially direct regulatory links, and a central role for the transcription factor Gata2 was further supported by analysis of Gata2^{+/-} HSPCs, which showed altered expression of some Gata2 targets that were consistent with the inferred network model. These early successes in regulatory network inference using single cell expression profiling suggest that this approach may be widely embraced by the scientific community. Future efforts should pay attention to the principles of partial correlation analysis discussed above in the context of the Aracne tool for network reconstruction from microarray expression profiles, because partial correlation analysis measures the degree of association between 2 genes with the potential effects of all other genes removed and therefore focuses analysis on associations that are more likely to be the result of direct regulatory relationships.

High-throughput single cell analysis also offers new opportunities to reconstruct the molecular nature of differentiation trajectories, based on the principle that within a population of differentiating yet unsynchronized cells, each single cell represents a snapshot that corresponds to a particular expression state on a given differentiation pathway. Computer tools have recently been reported that attempt to arrange such single cell measurements into putative differentiation time courses, thus introducing the concept of differentiation pseudotime.⁵⁰⁻⁵² From a network reconstruction perspective, this will offer exciting new opportunities that go beyond simple correlation analysis, because a temporal sequence of expression changes offers

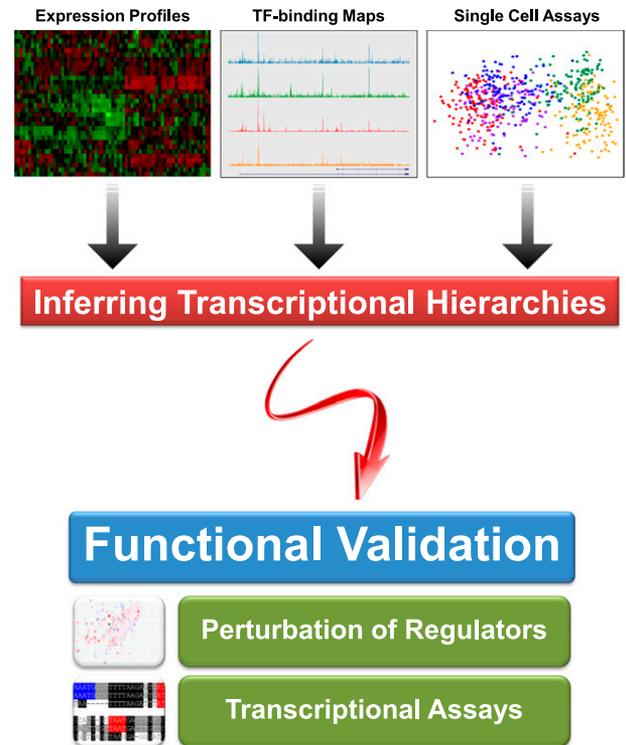


Figure 2. High-throughput technologies can be used to infer transcriptional hierarchies, but reconstruction of regulatory network models requires experimental validation of inferred regulatory connections. Shown at the top are representative examples of microarray expression profiles,⁷³ ChIP-Seq TF binding maps,⁷⁴ and single cell expression analysis,⁴⁷ which can all be used to infer regulatory hierarchies. The bottom part of the figure illustrates the importance of follow-up functional experiments such as perturbation and high-throughput transcriptional assays^{39,40} to validate inferred regulatory connections.

potential insights into cause-effect relationships and transcriptional hierarchies. Successful reconstruction of high-quality regulatory network models will require large-scale datasets, probably consisting of thousands of single cells analyzed along a differentiation time course. Further development of algorithms will also be required, for example, to permit accurate reconstruction of pseudotime for branched differentiation trajectories. Moreover, single cell RNA-Seq is rapidly overtaking microfluidics-based quantitative reverse transcriptase-polymerase chain reaction due to falling costs and increased robustness of experimental protocols.⁵³ Significant challenges remain, however, not just in terms of distinguishing experimental noise from true biological heterogeneity,⁵⁴ but also in the context of network reconstruction where information on thousands of genes is likely to require even larger numbers of single cells to achieve robust inference of regulatory connections.

Outlook

The reconstruction of ever more accurate regulatory network models for HSPCs will require integration of data from all the experimental platforms discussed above (Figure 2). Moreover, availability of additional data types will be essential, particularly with regard to accurate measurements of TF protein abundance and activity. Given the known phenotypic heterogeneity and rarity of HSCs, assay miniaturization to the single cell level will be critical just as it has been for mRNA expression measurements. Mass cytometry,⁵⁵ as

well as recently reported single cell western blot technology,⁵⁶ offer exciting perspectives, which will be enhanced further once protocols are developed that permit quantification of TF protein and target gene mRNA levels in the same single cell.

Owing to the central position of gene regulatory elements as building blocks of transcriptional regulatory networks, future progress in network reconstruction is also likely to come from emerging genome-scale “epigenetic” data, because TFs both influence, and are influenced by, the chromatin template at their target regulatory elements. Several recent studies have provided genome-wide DNA methylation maps alongside transcriptome analysis of hematopoietic differentiation and HSC aging,⁵⁷⁻⁶¹ with the latter study by Goodell and colleagues further interrogating histone modifications, in addition to DNA methylome and transcriptome analysis. Such integrated approaches provide very welcome information on the likely activity and potential dynamic status of individual gene regulatory elements, and as such can be used to examine to what extent emerging network models are consistent with the likely activity of specific gene regulatory elements. However, epigenomic analysis is no substitute for functional dissection of regulatory mechanisms, which remains essential for the generation of validated network models. Of note, recent developments in genome editing (eg, CRISPR/Cas9) provide powerful new means for saturation mutagenesis⁶² that might become useful for interrogating regulatory elements controlling HSC fate and thereby accelerate our ability to functionally validate regulatory network components.

TF input is particularly important when cells change state from for example a progenitor type to a descendant cell type. However, most of the multigene models and methods discussed in this review identify distinct stable states and not precursor-product relationships or developmental kinetics. Attempts to explain hematopoietic differentiation dynamics are only seen in the small-scale models based on 2 or 3 well-studied genes.^{5,6} Moreover, all current models are limited by their closed-system scope, and the challenge for whole-genome approaches is to make such explanations fully definitive and inclusive. Large-scale experimental perturbation tests (Figure 2) will be vital, not only to convert the evidence for different stable states into deeper insights into precursor-product relationships and developmental kinetics, but also to make use of new genome-scale datasets, to ultimately generate network models that actually explain the regulation of blood development.

Both experimental approaches and network modeling algorithms will need to take into account the importance of cell-cell interactions. It was shown several years ago that megakaryocyte-derived stimulatory growth factors can promote self-renewal of human HSPCs.⁶³ Using combined genomic and phenotype data, this work has recently been expanded to generate a directional cell-cell communication network between 12 human hematopoietic cell types isolated from umbilical cord blood.⁶⁴ Microfluidics technology already offers opportunities to monitor single cells responding to an external signal, as well as to quantify signaling molecules secreted by individual HSPCs.⁶⁵ A blueprint for the integration of inter-cellular communication into regulatory network models has been provided by Eric Davidson’s pioneering work on sea urchin development,¹ which is likely to provide an important starting point for the future generation of similarly comprehensive network models for blood cell development.

Genome-scale expression datasets have also been generated for more specialized (but perhaps equally clinically important) fates of HSCs such as mobilization, homing, and the changes that accompany physiological processes such as aging.⁶⁶⁻⁶⁸ More recently, genome-wide epigenomic maps have also been reported comparing young and

aged HSCs.⁶¹ It will be important to assess the extent to which regulatory network models for the more classical HSC fate choices (eg, lineage choice and self-renewal) can accommodate these additional functions of HSCs. Importantly, emerging genome-scale data coupled with targeted functional experiments will open the door to widen network analysis to encompass the whole spectrum of HSC physiology.

The generation of comprehensive regulatory network models for normal hematopoiesis will have major implications for our understanding of hematological pathologies. Many of the most commonly mutated genes in hematological malignancies encode transcriptional and epigenetic regulators. The functional consequences of such mutations can only be truly appreciated by taking into account the connectivity of a given gene within wider regulatory networks, because this critically determines how a given mutation will impact on the activity of other key regulators and therefore ultimately the behavior of the entire cell. Future therapeutic interventions are likely to increasingly depend on combination therapies for the concurrent targeting of multiple proteins, yet the number of possible permutations increases exponentially with the number of proteins under consideration. Computational modeling of leukemogenic mutations as network perturbations offers exciting opportunities to use computer simulations as a rapid means to identify those combinations of genes and/or regulatory interactions, where therapeutic intervention may have the potential to restore normal network behavior and therefore counteract the detrimental effect of the original mutation.

Finally, much effort in the field is directed at deriving transplantable HSCs from other cells types, via differentiation of induced pluripotent stem cells,⁶⁹ transdifferentiation of somatic cells such as fibroblasts⁷⁰ and endothelial cells,⁷¹ and also reprogramming/respecification of committed blood cells.⁷² Ultimately, the success of such efforts is contingent on instating (in the case of directed differentiation or transdifferentiation) or reinstating (in the case of reprogramming from differentiated blood cells) the regulatory networks governing HSC potential onto these other cell types. A more thorough understanding of HSC regulatory networks undoubtedly has the potential to greatly accelerate such research efforts.

Acknowledgments

Research in the author’s laboratory is supported by the Biotechnology and Biological Sciences Research Council, the Medical Research Council (MRC), Leukaemia and Lymphoma Research, Cancer Research UK, the Leukemia and Lymphoma Society, Microsoft Research, and core support grants by the Wellcome Trust to the Cambridge Institute for Medical Research and Wellcome Trust–MRC Cambridge Stem Cell Institute.

Authorship

Contribution: B.G. wrote the paper.

Conflict-of-interest disclosure: The author declares no competing financial interests.

Correspondence: Berthold Göttgens, Cambridge Institute for Medical Research, Hills Rd, Cambridge CB2 0XY, United Kingdom; e-mail: bg200@cam.ac.uk.

References

- Davidson EH. *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution*. San Diego, CA: Academic Press; 2006.
- Hamlett I, Draper J, Strouboulis J, Iborra F, Porcher C, Vyas P. Characterization of megakaryocyte GATA1-interacting proteins: the corepressor ETO2 and GATA1 interact to regulate terminal megakaryocyte maturation. *Blood*. 2008; 112(7):2738-2749.
- Sherwood RI, Hashimoto T, O'Donnell CW, et al. Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat Biotechnol*. 2014;32(2):171-178.
- Roeder RG. Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Lett*. 2005;579(4):909-915.
- Chickarmame V, Enver T, Peterson C. Computational modeling of the hematopoietic erythroid-myeloid switch reveals insights into cooperativity, priming, and irreversibility. *PLoS Comput Biol*. 2009;5(1):e1000268.
- Narula J, Smith AM, Gottgens B, Igoshin OA. Modeling reveals bistability and low-pass filtering in the network module determining blood stem cell fate. *PLoS Comput Biol*. 2010;6(5):e1000771.
- Krumsiek J, Marr C, Schroeder T, Theis FJ. Hierarchical differentiation of myeloid progenitors is encoded in the transcription factor network. *PLoS ONE*. 2011;6(8):e22649.
- Bonzanni N, Garg A, Feenstra KA, et al. Hard-wired heterogeneity in blood stem cells revealed using a dynamic regulatory network model. *Bioinformatics*. 2013;29(13):i80-i88.
- Kueh HY, Rothenberg EV. Regulatory gene network circuits underlying T cell development from multipotent progenitors. *Wiley Interdiscip Rev Syst Biol Med*. 2012;4(1):79-102.
- Singh H, Khan AA, Dinner AR. Gene regulatory networks in the immune system. *Trends Immunol*. 2014;35(5):211-218.
- Friedman N. Inferring cellular networks using probabilistic graphical models. *Science*. 2004; 303(5659):799-805.
- Ideker T, Thorsson V, Ranish JA, et al. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*. 2001;292(5518):929-934.
- Butte AJ, Kohane IS. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pac Symp Biocomput*. 2000;4:18-429.
- Margolin AA, Nemenman I, Basso K, et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*. 2006;7(Suppl 1):S7.
- Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A. Reverse engineering of regulatory networks in human B cells. *Nat Genet*. 2005;37(4):382-390.
- Palomero T, Lim WK, Odum DT, et al. NOTCH1 directly regulates c-MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. *Proc Natl Acad Sci USA*. 2006;103(48):18261-18266.
- Della Gatta G, Palomero T, Perez-Garcia A, et al. Reverse engineering of TLX oncogenic transcriptional networks identifies RUNX1 as tumor suppressor in T-ALL. *Nat Med*. 2012;18(3): 436-440.
- Mani KM, Lefebvre C, Wang K, et al. A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas. *Mol Syst Biol*. 2008;4:169.
- Zeng T, Li J. Maximization of negative correlations in time-course gene expression data for enhancing understanding of molecular pathways. *Nucleic Acids Res*. 2010;38(1):e1.
- Segal E, Shapira M, Regev A, et al. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet*. 2003;34(2):166-176.
- Novershtern N, Subramanian A, Lawton LN, et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell*. 2011;144(2):296-309.
- Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010; 38(4):576-589.
- Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. *Science*. 2007;316(5830): 1497-1502.
- Robertson G, Hirst M, Bainbridge M, et al. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods*. 2007;4(8):651-657.
- Ruau D, Ng FS, Wilson NK, et al. Building an ENCODE-style data compendium on a shoestring. *Nat Methods*. 2013;10(10):926.
- Beck D, Thoms JA, Perera D, et al. Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes. *Blood*. 2013; 122(14):e12-e22.
- Goldberg L, Tijssen MR, Birger Y, et al. Genome-scale expression and transcription factor binding profiles reveal therapeutic targets in transgenic ERG myeloid leukemia. *Blood*. 2013;122(15): 2694-2703.
- Mylona A, Andrieu-Soler C, Thongjuea S, et al. Genome-wide analysis shows that Ldb1 controls essential hematopoietic genes/pathways in mouse early development and reveals novel players in hematopoiesis. *Blood*. 2013;121(15): 2902-2913.
- Tanaka Y, Joshi A, Wilson NK, Kinston S, Nishikawa S, Göttgens B. The transcriptional programme controlled by Runx1 during early embryonic blood development. *Dev Biol*. 2012; 366(2):404-419.
- Xu J, Shao Z, Glass K, et al. Combinatorial assembly of developmental stage-specific enhancers controls gene expression programs during human erythropoiesis. *Dev Cell*. 2012; 23(4):796-811.
- Zhang JA, Mortazavi A, Williams BA, Wold BJ, Rothenberg EV. Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. *Cell*. 2012;149(2):467-482.
- Cheng Y, Wu W, Kumar SA, et al. Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. *Genome Res*. 2009;19(12):2172-2184.
- Dawson MA, Foster SD, Bannister AJ, et al. Three distinct patterns of histone H3Y41 phosphorylation mark active genes. *Cell Reports*. 2012;2(3):470-477.
- Yu S, Cui K, Jothi R, et al. GABP controls a critical transcription regulatory module that is essential for maintenance and differentiation of hematopoietic stem/progenitor cells. *Blood*. 2011; 117(7):2166-2178.
- DeVilbiss AW, Sanalkumar R, Johnson KD, Keles S, Bresnick EH. Hematopoietic transcriptional mechanisms: from locus-specific to genome-wide vantage points. *Exp Hematol*. 2014;42(8): 618-629.
- Wilson NK, Foster SD, Wang X, et al. Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell*. 2010;7(4):532-544.
- Lichtinger M, Ingram R, Hannah R, et al. RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. *EMBO J*. 2012;31(22): 4318-4333.
- Calero-Nieto FJ, Ng FS, Wilson NK, et al. Key regulators control distinct transcriptional programmes in blood progenitor and mast cells. *EMBO J*. 2014;33(11):1212-1226.
- Dickel DE, Zhu Y, Nord AS, et al. Function-based identification of mammalian enhancers using site-specific integration. *Nat Methods*. 2014;11(5): 566-571.
- Wilkinson AC, Goode DK, Cheng YH, et al. Single site-specific integration targeting coupled with embryonic stem cell differentiation provides a high-throughput alternative to in vivo enhancer analyses. *Biol Open*. 2013;2(11):1229-1238.
- Ema H, Morita Y, Suda T. Heterogeneity and hierarchy of hematopoietic stem cells. *Exp Hematol*. 2014;42(2):74-82.e72.
- Brady G, Barbara M, Iscove N. Representative in vitro cDNA amplification from individual hemopoietic cells and colonies. *Methods Mol Cell Biol*. 1990;2:17-25.
- Hoang T, Paradis E, Brady G, et al. Opposing effects of the basic helix-loop-helix transcription factor SCL on erythroid and monocytic differentiation. *Blood*. 1996;87(1):102-111.
- Hu M, Krause D, Greaves M, et al. Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev*. 1997;11(6): 774-785.
- Pina C, Fugazza C, Tipping AJ, et al. Inferring rules of lineage commitment in haematopoiesis. *Nat Cell Biol*. 2012;14(3):287-294.
- Glotzbach JP, Januszky M, Vial IN, et al. An information theoretic, microfluidic-based single cell analysis permits identification of subpopulations among putatively homogeneous stem cells. *PLoS ONE*. 2011;6(6):e21211.
- Moignard V, Macaulay IC, Swiers G, et al. Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. *Nat Cell Biol*. 2013;15(4):363-372.
- Moignard V, Göttgens B. Transcriptional mechanisms of cell fate decisions revealed by single cell expression profiling. *BioEssays*. 2014; 36(4):419-426.
- Guo G, Luc S, Marco E, et al. Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. *Cell Stem Cell*. 2013;13(4):492-505.
- Durruthy-Durruthy R, Gottlieb A, Hartman BH, et al. Reconstruction of the mouse otocyst and early neuroblast lineage at single-cell resolution. *Cell*. 2014;157(4):964-978.
- Bendall SC, Davis KL, Amir AD, et al. Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell*. 2014;157(3):714-725.
- Trapnell C, Cacchiarelli D, Grimsby J, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol*. 2014;32(4):381-386.
- Mahata B, Zhang X, Kolodziejczyk AA, et al. Single-cell RNA sequencing reveals T helper cells synthesizing steroids de novo to contribute to

- immune homeostasis. *Cell Reports*. 2014;7(4):1130-1142.
54. Brennecke P, Anders S, Kim JK, et al. Accounting for technical noise in single-cell RNA-seq experiments. *Nat Methods*. 2013;10(11):1093-1095.
 55. Bandura DR, Baranov VI, Ornatsky OI, et al. Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal Chem*. 2009;81(16):6813-6822.
 56. Hughes AJ, Spelke DP, Xu Z, Kang CC, Schaffer DV, Herr AE. Single-cell western blotting. *Nat Methods*. 2014;11(7):749-755.
 57. Bock C, Beerman I, Lien WH, et al. DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. *Mol Cell*. 2012;47(4):633-647.
 58. Ji H, Ehrlich LI, Seita J, et al. Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature*. 2010;467(7313):338-342.
 59. Attema JL, Papathanasiou P, Forsberg EC, Xu J, Smale ST, Weissman IL. Epigenetic characterization of hematopoietic stem cell differentiation using miniChIP and bisulfite sequencing analysis. *Proc Natl Acad Sci USA*. 2007;104(30):12371-12376.
 60. Beerman I, Bock C, Garrison BS, et al. Proliferation-dependent alterations of the DNA methylation landscape underlie hematopoietic stem cell aging. *Cell Stem Cell*. 2013;12(4):413-425.
 61. Sun D, Luo M, Jeong M, et al. Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal. *Cell Stem Cell*. 2014;14(5):673-688.
 62. Findlay GM, Boyle EA, Hause RJ, Klein JC, Shendure J. Saturation editing of genomic regions by multiplex homology-directed repair. *Nature*. 2014;513(7516):120-123.
 63. Kirouac DC, Ito C, Cszaszar E, et al. Dynamic interaction networks in a hierarchically organized tissue. *Mol Syst Biol*. 2010;6:417.
 64. Qiao W, Wang W, Laurenti E, et al. Intercellular network structure and regulatory motifs in the human hematopoietic system. *Mol Syst Biol*. 2014;10(7):741.
 65. Zhao JL, Ma C, O'Connell RM, et al. Conversion of danger signals into cytokine signals by hematopoietic stem and progenitor cells for regulation of stress-induced hematopoiesis. *Cell Stem Cell*. 2014;14(4):445-459.
 66. Forsberg EC, Passegué E, Prohaska SS, et al. Molecular signatures of quiescent, mobilized and leukemia-initiating hematopoietic stem cells. *PLoS ONE*. 2010;5(1):e8785.
 67. Rossi DJ, Bryder D, Zahn JM, et al. Cell intrinsic alterations underlie hematopoietic stem cell aging. *Proc Natl Acad Sci USA*. 2005;102(26):9194-9199.
 68. Chambers SM, Shaw CA, Gatz C, Fisk CJ, Donehower LA, Goodell MA. Aging hematopoietic stem cells decline in function and exhibit epigenetic dysregulation. *PLoS Biol*. 2007;5(8):e201.
 69. Doulatov S, Vo LT, Chou SS, et al. Induction of multipotential hematopoietic progenitors from human pluripotent stem cells via respecification of lineage-restricted precursors. *Cell Stem Cell*. 2013;13(4):459-470.
 70. Pereira CF, Chang B, Qiu J, et al. Induction of a hemogenic program in mouse fibroblasts. *Cell Stem Cell*. 2013;13(2):205-218.
 71. Sandler VM, Lis R, Liu Y, et al. Reprogramming human endothelial cells to haematopoietic cells requires vascular induction. *Nature*. 2014;511(7509):312-318.
 72. Riddell J, Gazit R, Garrison BS, et al. Reprogramming committed murine blood cells to induced hematopoietic stem cells with defined factors. *Cell*. 2014;157(3):549-564.
 73. Hebestreit K, Gröttrup S, Emden D, et al. Leukemia gene atlas—a public platform for integrative exploration of genome-wide molecular data. *PLoS ONE*. 2012;7(6):e39148.
 74. Hannah R, Joshi A, Wilson NK, Kinston S, Göttgens B. A compendium of genome-wide hematopoietic transcription factor maps supports the identification of gene regulatory control mechanisms. *Exp Hematol*. 2011;39(5):531-541.



blood[®]

2015 125: 2614-2620

doi:10.1182/blood-2014-08-570226 originally published
online March 11, 2015

Regulatory network control of blood stem cells

Berthold Göttgens

Updated information and services can be found at:

<http://www.bloodjournal.org/content/125/17/2614.full.html>

Articles on similar topics can be found in the following Blood collections

[Hematopoiesis and Stem Cells](#) (3363 articles)

[Review Articles](#) (651 articles)

[Review Series](#) (109 articles)

Information about reproducing this article in parts or in its entirety may be found online at:

http://www.bloodjournal.org/site/misc/rights.xhtml#repub_requests

Information about ordering reprints may be found online at:

<http://www.bloodjournal.org/site/misc/rights.xhtml#reprints>

Information about subscriptions and ASH membership may be found online at:

<http://www.bloodjournal.org/site/subscriptions/index.xhtml>