Mechanisms of fate decision and lineage commitment during haematopoiesis

Ana Cvejic[1, 2, 3]

[1]Department of Haematology, University of Cambridge, CB2 0PT, UK

[2]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1HH, UK

[3]Wellcome Trust – Medical Research Council Cambridge Stem Cell Institute, Cambridge, CB2 1QR, UK

Contact Information:

Ana Cvejic

E-mail: as889@cam.ac.uk

Abstract

Blood stem cells need to both perpetuate themselves (self-renew) and differentiate into all mature blood cells to maintain blood formation throughout life. However, it is unclear how the underlying gene regulatory network maintains this population of self-renewing and differentiating stem cells, and how it accommodates the transition from a stem cell to a mature blood cell. Our current knowledge of transcriptomes of various blood cell types has mainly been advanced by population-level analysis. However, a population of seemingly homogenous blood cells may include many distinct cell types with substantially different transcriptomes and abilities to make diverse fate decisions. Therefore, understanding the cell-intrinsic differences between individual cells is necessary for a deeper understanding of the molecular basis of their behavior. Here we review recent single cell studies in the haematopoietic system and their contribution to our understanding of the mechanisms governing cell fate choices and lineage commitment.

Introduction

A single cell type, the haematopoietic stem cell (HSC), is responsible for generating all blood cells throughout the lifetime of an organism [1]. The HSC is a rare cell that resides primarily in the bone marrow of adult mammals. It has the ability to either self-renew, and generate more stem cells or differentiate and generate over 10 different blood cell types. These different blood cells provide functions such as protection against infections, oxygen transport and maintaining haemostasis. Thus, over time each HSC makes essential fate decisions by integrating a wide array of signals from the microenvironment and completing complex changes in the regulation of gene expression. Clarifying how HSCs differentiate into diverse cell types is important for understanding how they attain their various functions and offers the potential for therapeutic manipulation.

Traditionally, different blood cells are distinguished from each other based on the expression of a handful of cell-specific surface markers [2, 3]. Our knowledge of the different haematopoietic cell types is a direct result of the development of reagents to distinguish these various cells by their cell surface markers, followed by functional, transplant-based tracking of their activities [4]. An inherent problem with this approach is that the presence of specific cell surface markers doesn't directly reflect the transcriptional state of a cell. In addition, the variable loss or gain of marker expression occurs according to the activation/proliferative state of the cell [5]. Since the typical mammalian cell expresses around 10^4 genes [6], a more comprehensive and objective strategy is needed to define cell types. Although the transcriptomes of populations of HSC and progenitor cells have previously been assessed on microarrays [7] and more recently using RNA-Sequencing of bulk cells [8], individual cells can exhibit substantial heterogeneity in gene expression (Figure 1) with important functional consequences [9]. Understanding the cell-intrinsic differences between individual cells is necessary for a deeper understanding of the molecular basis of their behavior.

A population of seemingly homogenous haematopoietic cells captured at any one time, using well defined combinations of cell surface markers, may include many distinct or intermediate cellular states [9-11]. This has been nicely illustrated by recent studies showing that early diversification into cells with distinct lineage bias within the HSC compartment may exist and that individual HSCs lead to different reconstitution patterns. The balanced production of myeloid and lymphoid cells or deficiency in lymphoid potential as well as long term self-renewal potential are shown to be intrinsic HSCs properties that are stably inherited by their HSC offspring [10, 12, 13]. Considering only average properties, for example by bulk transcriptomics analysis, masks subpopulations of cells (Figure 1) [14]. The unique epigenetic signature and transcription factor networks in individual progenitor cells can influence the fate-choice decisions and the response of these cells to extrinsic signals. Therefore, a study of blood development with single-cell resolution is required to define cellular level heterogeneities that presage distinct differentiation decisions.

Recent technological advances now provide us with tools that allow single-cell molecular profiling and acquisition of transcriptomic data from hundreds to thousands of individual cells. These methods range from fluorescent in situ hybridization (FISH) as a probe-dependent method, RT-qPCR where cDNA obtained from a single cell can be used to quantify the expression of up to 100 transcripts to single-cell RNA sequencing which provides a highly resolved picture of whole-genome gene expression patterns of a single cell (reviewed in [15-18]). The development of automated massively parallel RNA single cell protocols that allow molecule counting of the transcripts from thousands of cells (such as MARS-Seq) enabled dissection of complex tissues into distinct cell types [19]. The wider application of these methods will permit agnostic interrogation of haematopoiesis and transform our view of how cellular decisions are made. An integrative strategy, combining genetic perturbation with computational sequence and network analysis methods, will further reveal regulatory networks that maintain the dynamic balance between different blood cell types.

The traditional model of haematopoiesis assumes a stepwise set of binary choices across the full haematopoietic spectrum

Haematopoiesis is often depicted by a hierarchical differentiation tree, with HSCs at the root and the mature blood cells as the branches. Self-renewing HSCs give rise to intermediate, transiently amplifying multipotent progenitor cells (MPPs) that subsequently, through lineage commitment and differentiation, generate all cells of the haematopoietic system. The multipotent state of stem and progenitor cells is characterized by the co-expression of numerous lineage-affiliated genes. This "multilineage priming" is thought to be functionally related to the cell's ability to "chose" the cell fate prior to unilineage commitment and differentiation i.e. irreversibly losing the capacity to differentiate to any other cell type [20-23].

The conventional model of haematopoiesis assumes a stepwise set of binary choices across the full haematopoietic spectrum [24]. In the first step, MPPs produce common lymphoid (CLP) or common myeloid (CMP) progenitor cells. CMP cells are restricted in their development to cell types of the megakaryocyte-erythroid and myeloid (i.e. granulocyte-monocyte) lineages whereas CLP cells show restricted development into T-cells, B-cells, and natural killer cells [25-28]. After the early and irreversible segregation of lymphoid and myeloid differentiation pathways, progenitor cells usually differentiate into one of two cell fates. For example, megakaryocyte-erythroid progenitors can differentiate to either megakaryocytes or erythroid cells and granulocyte-monocyte progenitors give rise to neutrophils and monocytes. This descriptive and widely accepted model of haematopoiesis is, however, probably too simplistic. The identification of early progenitors with lymphoid and myeloid potential (EPLMs) [29] and lymphoid-primed multipotent progenitors (LMPP) [30, 31] that have granulocytic, monocytic and lymphoid potential but low potential to form megakaryocyte and erythroid lineages suggests that myeloid and lymphoid potential remain associated downstream of the HSCs. This has prompted development of new, alternative models of haematopoiesis [32-35].

Single cell analysis supports an alternative model of the cellular organization of the haematopoietic lineage tree

Analysing transcriptomes of cells is a standard molecular approach to better understanding how cells function and to uncover the state of a cell in a specific environment. Single cell analysis proved to be very valuable in resolving some of the outstanding issues in the cellular organisation of the haematopoietic lineage tree. Single cell gene expression analysis of 280 cell surface markers and 180 intracellular regulators, including important transcription factors (TFs), revealed considerable heterogeneity within each of, over ten, examined haematopoietic populations [9]. The transcriptional differences were present in even closely related cells emphasizing that no two cells are truly identical in terms of gene expression. However, it also suggested that populations of progenitor cells, sorted based on well-defined cell surface markers and previously considered to be functionally homogenous, are actually a mixture of cells with distinct differentiation fates. The hierarchical clustering of most variable genes within the CMP population revealed that CMPs are composed of at least two distinct populations, which could be separated based on CD55 expression. When transplanted in irradiated mouse recipients, CD55+ CMPs produced predominantly platelets, whereas CD55- CMPs gave rise mainly to myeloid cells [9]. Therefore, transcriptional differences between CD55+ and CD55- populations were mirrored in their functional properties. To recover the cellular hierarchy from these single cell expression data, spanning-tree progression analysis of density-normalized events (SPADE) was used. The generated model implied that the megakaryocyte-erythroid lineage ($CD55^+$ compartment) is closely linked to long-term repopulating HSCs and separated early from the lympho-myeloid lineage ($CD55^-$ compartment). Single cell *in vitro* assays further confirmed that the megakaryocyte colonies are the first to emerge from HSCs. In line with these findings, no single cells have been identified so far, in the HSPC pool, that co-express lymphoid and megakaryocyte-erythroid lineage affiliated genes; thus, the megakaryocyte-erythroid lineage separates early from myeloid and lymphoid fates. This early separation of the megakaryocyte-erythroid lineage from HSCs has been supported by other, alternative

models of haematopoiesis, which assume that myeloid, and megakaryocyte-erythroid priming of HSPCs precedes lymphoid lineage priming.

Interestingly, recent studies suggested that platelet-primed HSCs are at the top of haematopoietic cellular hierarchy. Platelet-primed stem cells were identified within vWf (von Willebrand factor) expressing long term HSCs (LT-HSC). Single transplanted vWf+ HSCs gave rise predominantly to platelets and myeloid cells, and were able to self-renew and give rise to lymphoid-biased HSCs [36]. Thus, commitment to the megakaryocyte lineage starts in the most primitive stem cell compartment [36].

Principles of lineage commitment: insights from single cell gene expression analysis

The competence of multipotent cells to differentiate into diverse blood lineages is followed by their irreversible commitment towards a specific lineage. Understanding of the molecular basis of lineage commitment relies on the ability to asses transcriptional changes in individual cells at either side of the "commitment boundary". While this is experimentally challenging *in vivo*, *in vitro* single cell gene expression studies offered some interesting insights. By investigating Sca1^{lo} CD34+ and Sca1^{lo} CD34- cells from a multipotent haematopoietic cell line EML (Erythroid-Myeloid-Lymphoid cells) [37] the experimental system has been established in which it is possible to examine transcriptional changes that underlie transition from self-renewal to lineage commitment. Sca1^{lo} CD34+ cells reconstituted in culture and are multipotent with erythroid and neutrophil differentiation potential whereas Sca1^{lo} CD34- are erythroid-committed cells with no culture reconstitution potential. On a population level, the two compartments were transcriptionally different confirming that self-renewal and lineage commitment are two independent and separable cell states. Examination of these two states at the single cell level revealed that erythroid cell commitment could occur in the absence of a complete erythroid transcriptome and some of the key lineage regulators (e.g. *Gata1*) but also prior to silencing components of alternative lineage programs. Self-renewing cells were more similar to each other, with no clear

subpopulations that would reflect differences in the functional culture-reconstitution potential of these cells. In contrast, a significant cell to cell variation was present in early-committed erythroid cells further suggesting that commitment can occur through multiple entry points. In other words, commitment occurs due to stochastic independent expression of key regulatory factors rather than coordinated expression of lineage programs. Computational modeling of this data set further revealed that an increase in *Gata2* and a decrease in *Mpo*~~,~~ were the best predictors of erythroid commitment, whereas *Gata1* was important to a lesser extent [38]. However, no significant level of gene expression coordination was detectable in the commitment transition. Due to the lack of correlation between the expression of *Gata2* and other genes, it remained unclear what the specific molecular mechanisms by which *Gata2* can drive cells into commitment are.

Similar principals of lineage restriction in lymphoid cells have been reported during B-cell lineage commitment [39]. Lymphoid lineage priming starts in LMPPs, which express the repertoire of lymphoid associated genes but do not express genes specific for B- or T-lineage cells; this lineage commitment starts in a subpopulation of CLPs. Analysis of the expression of key B lineage genes in individual CLP cells revealed that B lineage commitment occurs in a stepwise manner, prior to expression of B lineage specific surface markers [40]. CLPs initially lose NK-cell potential to become B/T cell–restricted progenitors, and then mature into B cell–restricted progenitors. Therefore, the CLP compartment, as defined based on the classical combination of cell surface markers, is composed of a mixture of cells with relatively restricted lineage potentials. On a molecular level, *Ebf1* (early B cell factor 1) has been recognized as essential for expression of B lineage genes and B cell specification. However, physiological levels of *Ebf1* are not sufficient to cause lineage restriction. Instead, the stable commitment of B cells is dependent on the action of *Pax5* (paired box 5) and *Ikaros* (IKAROS Family Zinc Finger 1). *Pax5* deficiency results in accumulation of cells specified for development toward the B lineage but these cells display defective lineage commitment [41]. *Pax5*–deficient progenitors maintained the ability to differentiate to T cells and to a lesser extent to myeloid cells, suggesting a minimal role of

Pax5 in lympho-myeloid lineage restriction. Therefore, distinct regulatory networks control a coordinated activation of B lineage associated genes and those linked to specification of B cells as opposed to genes responsible for B lineage commitment, which restricts the cell to a single, B lineage fate.

Cellular decisions: discrete steps or gradual process?

One of the drawbacks of traditional methods of purifying haematopoietic cells to relative homogeneity is the limited number of cell surface markers that is used simultaneously to define the blood cell type. Therefore, the sorted populations are transcriptionally and functionally heterogeneous [9, 40]. In addition, only a subpopulation of the overall cellular pool is examined which restricts the ability to characterize transitional populations and the relationships between them. Due to these technical limitations, the prevalent model for haematopoiesis assumes a stepwise set of obligatory steps through which progeny of haematopoietic stem cells passes during lineage development. The transcriptional programs that govern this process have mainly been investigated by population-level analysis of these discrete steps, which cannot reveal the continuous nature of the differentiation process. It is now appreciated that differentiation of haematopoietic cells is governed by gradual changes in the expression of the array of key TFs [42]. Since these changes in expression of TFs occur at different rates, progenitor cells with varying degrees of lineage potential are generated; therefore accounting for progenitors with varying combinations of myeloid and lymphoid potential. Indeed, rather than a process of immediate transition between strictly defined stages, haematopoiesis is more likely a progressive and continual sequence of fate decisions that the cell makes (Figure 2) in response to intrinsic factors and environmental signals.

One of the early computational methods designed to capture trends in high-dimensional data, including ordering of cells based on their developmental progression, was SPADE [43]. For example, SPADE was used to computationally reconstruct the haematopoietic lineage

progression, in an unbiased way, from HSCs using 40 distinct gene expression clusters [9]. However, while SPADE allowed easier visualisation and interpretation of the data, its main application was not to infer a trajectory and the precise order of events nor their progression. The minimum spanning tree (MST) that underlies SPADE could provide only an estimation of the developmental ordering. Therefore, single cell resolution was lost in SPADE and it had limited success in depicting the continuous nature of transcriptional changes occurring during haematopoietic cell differentiation.

Of the more recently developed algorithms, amongst others, Monocle [44], Wanderlust [45] and Waterfall [46] were designed to order single cells along a one-dimensional trajectory and track their developmental chronology. Instead of clustering blood cells in similar groups, Wanderlust was designed to position individual cells in a graph structure. Antibodies coupled to different, stable, transition element isotopes were used to bind target epitopes within and on the surface of the cells. By using information from single-cell mass cytometry data and simultaneous measurement of different surface markers, as well as internal functional proteins, Wanderlust generated trajectories that enabled ordering of human cells from HSCs through to naive B cells [45]. The unbiased ordering of cells identified a population of very early B cell progenitors that developmentally precede cells that express canonical B cell surface markers, CD10 and CD19. CD10 and CD19 have been considered to be the earliest markers traditionally used in the identification of human B cells. The newly identified, early population of human B cells in the marrow was marked by expression of CD24 on their surface. Evidence from single cell expression analysis of murine CLPs also suggests that even in the absence of detectable CD19 surface expression, a significant proportion of progenitor cells are B lineage committed [40]. Therefore, single cell analysis allows identification of previously unrecognized subsets of progenitor cells.

Gene regulatory networks: blueprint of the functional cooperativity among genes

It is well accepted that the spatio-temporal control of transcription of thousands of genes

during fate determination is not regulated by a single lineage-specific TF. Instead, fate choice is enforced by combinatorial diversity generated by the binding of multiple TFs at key regulatory sites [47, 48]. The regulatory regions serve to combine the input signals from the regulators (e.g. TFs) and thereby control the level of gene expression during transcription. The regulators and genes, together with the regulatory connections and interactions between them, form the gene regulatory networks (GRNs). Therefore, GRNs can be regarded as a blueprint for understanding the functional cooperativity among genes and can offer elucidation of developmental processes and cell differentiation on a systems level.

The total amount of mRNA produced during transcription is a measure of how functional or active a gene is. Although gene expression analysis from populations of cells can be used to infer gene networks, population level analysis can mask important gene correlations. For example, while a group of genes can be regulated independently by two TFs, the GRN generated from the bulk RNA-sequencing data might predict that these two TFs are co-expressed or that one regulates the other one. Single cell RNA-sequencing can reveal that the examined population of cells is heterogeneous and that actually these two TFs are exclusively expressed in any given cell. GRNs constructed from single cells would reveal that the group of genes is regulated by two different TFs. Therefore, identifying GRNs that regulate the behavior of individual haematopoietic cells is essential for understanding the underlying transcriptional programs that drive lineage specification and commitment.

Single cell analysis, in combination with functional studies, discovered a GRN that is associated with early segregation of the megakaryocyte-erythroid lineage from the lympho-myeloid lineage [9]. The correlated gene expression in individual cells from five haematopoietic populations (HSC, MPP, CMP, MEP, GMP, CLP) combined with an existing ChIP-seq binding dataset for ten major TFs [47] revealed *Gata2* as a central regulator in the megakaryocyte-erythroid versus lympho-myeloid fate choice. The functional relevance of these findings was confirmed by single cell analysis of *Gata2* haploinsufficient HSPCs, which

showed down-regulation of megakaryocyte-erythroid and up-regulation of lympho-myeloid markers [9]. Similarly, single cell gene expression analysis of 18 TFs in hundreds of primary blood stem and progenitor cells also identified *Gata2* as an important regulator of commitment of HSCs toward the megakaryocyte-erythroid lineage [11]. Examination of the existing ChIP-Seq data combined with transgenic and transcriptional assays revealed a regulatory triad that consisted of *Gata2*, *Gfi1* and *Gfi1b*. *Gata2* has been proposed to function in a regulatory loop to modulate *Gfi1*/*Gfi1b* cross-antagonism and thus control specification of HSCs into megakaryocyte-erythroid and lympho-myeloid lineages [11]. All together, these findings supported the idea that the commitment to the megakaryocyte-erythroid lineage might occur, at the top of the haematopoietic lineage tree, in HSCs.

Gene regulatory networks can also be used to delineate lineage progression during early blood development [49]. Murine blood and vessel development initiates during gastrulation from mesodermal multipotent cells that express *Flk1* (Fetal Liver Kinase 1). These cells initially have the potential to form blood, endothelium and smooth muscle cells. Gene expression analysis of over 40 TF and marker genes involved in endothelial and hematopoietic development of post-implantation mouse embryos, at four sequential stages of development, identified key events necessary for the development of mesoderm toward blood and endothelial-like fates and specification of primitive erythroid cells. The analysis revealed asynchronous maturation of individual cells while they underwent commitment to haemogenic endothelium. The analysis of the underlying regulatory network predicted a key role of *Sox7* in the development of erythroid fate, which was subsequently experimentally validated using a transgenic mouse assay. The generated network coupled with experimental validation further identified transcriptional regulator *Erg* as a downstream target of *Sox* and *Hox* factors during early blood specification. Therefore, single-cell expression profiling, combined with novel computational approaches for network synthesis, can provide insight into the molecular basis of early blood development.

Concluding remarks

Traditionally HSCs are defined as cells with the long-term ability to self-sustain and to persistently generate both lymphoid and myeloid cells. However, intrinsic heterogeneity of classically defined HSCs in their differentiation output has challenged the concept that all HSCs have equal multi-lineage differentiation potential. HSCs isolated using any of the known combinations of cell surface markers, are transcriptionally heterogeneous, underscoring the importance of analyses at the single-cell level.

Single cell gene expression analysis has revealed that populations of progenitor cells, sorted based on the well-defined sets of specific cell surface markers, such as CMPs, CLPs and possibly other progenitors, are also transcriptionally and functionally heterogeneous. An additional level of complexity has been introduced, into the traditional hierarchical model of haematopoiesis, with the identification of progenitor cells with mixed lineage potential; these populations cannot be integrated into the current model of haematopoiesis that posits early segregation of myeloid and lymphoid potential. Therefore, it is necessary to revise the current view of the lineage development in haematopoiesis.

Future studies will need to address the exact composition of the stem and progenitor's transcriptional states *in vivo,* as well as the relationships between them. It is now more appreciated that HSPCs' differentiation is a gradual, continuous process both on a transcriptional and functional level rather than a set of obligatory, discrete steps. However, what these transcriptional states are *in vivo* and what the key genes and networks that regulate the transition of cells from one state to another are remains unclear. Single cell transcriptome analysis might provide answers to these questions.

Although current methodology for sorting haematopoietic stem and progenitor cells, based on well-defined cell surface markers, have long provided means for classifying haematopoietic cells, the identification of the whole spectra of blood transcriptional states will

require alternative methodology for isolating cells. The marker-free isolation of heterogeneous populations of blood cells, followed by single cell transcriptome analysis could provide means for data-driven classification of cell types. The high-resolution transcriptional maps of both main and intermediate cellular states during differentiation will further help in (re)defining the haematopoietic lineage branching map. Importantly, the relationships in the GRNs that regulate blood cell state transitions will need to be assessed for functional relevance by perturbation experiments. Single cell RNA-sequencing has been used, for example, to identify network controlling differential responses of dendritic cell subtypes, in the spleen and bone marrow, to different stimuli [19, 50]. Application and development of appropriate computational and statistical methods will be essential to fully exploit the wealth of generated data [51].

Gene expression analysis of multiple genes in a limited number of individual cells has identified that stochastic, independent activation of regulatory factors leads to lineage commitment prior to establishment of a coordinated lineage program. Single cell RNA sequencing will provide information on the full spectrum of molecular changes that are necessary for the transition from multipotent to committed state to occur. It should be noted however, that fluctuations in gene expression at the RNA level do not necessarily reflect the level of protein present in individual cells. The simultaneously measured absolute mRNA and protein abundance in both bacteria and eukaryotes, suggests that the cellular concentrations of mRNAs and corresponding proteins do correlate. It has been estimated that between 40% and 80% of the variation in protein concentration can be explained by variation in mRNA abundances [52-56]. Single-cell mass cytometry theoretically allows simultaneous detection of over 100 different proteins in thousands of cells. Development of methodologies for detection of protein and mRNA level in the same cell will be necessary to relate these parameters with cell behavior and fate choices.

By sequencing thousands of single cells, future studies are poised to go beyond traditional approaches in examining the complex relationships between the continuous spectra of blood

cells, and will provide unprecedented insight into the regulation of blood cell formation. They will illuminate novel disease-causing genes and inform approaches aimed at manipulating the expansion, directed-differentiation or reprogramming of stem cell fate for therapeutic advantage.

Conflict of interest disclosure

The author reports no potential conflicts of financial interests.

References:

1. Orkin SH, Zon LI. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* 2008; **132**: 631-644.

2. Muller-Sieburg CE, Whitlock CA, Weissman IL. Isolation of two early B lymphocyte progenitors from mouse marrow: a committed pre-pre-B cell and a clonogenic thy-1 lo hematopoietic stem cell. *Cell* 1986; **44**:653-662.

3. Spangrude GJ, Heimfeld S, Weissman IL. Purification and characterization of mouse hematopoietic stem cells. *Science* 1988; **241**: 58-62.

4. Osawa M, Hanada K-i, Hamada H, Nakauchi H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* 1996; **273**: 242-245

5. Schroeder T. Hematopoietic stem cell heterogeneity: subtypes, not unpredictable behavior. *Cell Stem Cell* 2010; **6**: 203-207.

6. Hastie ND, Bishop JO. The expression of three abundance classes of messenger RNA in mouse tissues. *Cell* 1976; **9**:761-774.

7. Watkins NA, Gusnanto A, de Bono B, De S, Miranda-Saavedra D, Hardie DL, et al. A HaemAtlas: characterizing gene expression in differentiated human blood cells. *Blood* 2009; **113**: e1-e9.

8. Chen L, Kostadima M, Martens JHA, Canu G, Garcia SP, Turro E, et al. Transcriptional diversity during lineage commitment of human blood progenitors. *Science* 2014; **345**: 1251033

9. Guo G, Luc S, Marco E, Lin T-W, Peng C, Kerenyi MA, et al. Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. *Cell Stem Cell* 2013; **13**: 492-505.

10. Muller-Sieburg CE, Sieburg HB, Bernitz JM, Cattarossi G. Stem cell heterogeneity: implications for aging and regenerative medicine. *Blood* 2012; **119**: 3900-3907.

11. Moignard V, Macaulay IC, Swiers G, Buettner F, Schütte J, Calero-Nieto FJ, 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**: 363-372.

12. Sieburg HB, Cho RH, Dykstra B, Uchida N, Eaves CJ, Muller-Sieburg CE. The hematopoietic stem compartment consists of a limited number of discrete stem cell subsets. *Blood* 2006; **107**:2311-2316.

13. Dykstra B, Kent D, Bowie M, McCaffrey L, Hamilton M, Lyons K, et al. Long-term propagation of distinct hematopoietic differentiation programs *in vivo*. *Cell Stem Cell* 2007; **1**: 218-229.

14. Sandberg R. Entering the era of single-cell transcriptomics in biology and medicine. *Nat Methods* 2014; **11**: 22-24.

15. Kalisky T, Blainey P, Quake SR. Genomic analysis at the single-cell level. *Annu Rev Genet* 2011; **45**: 431-445.

16. Tang F, Lao K, Surani MA. Development and applications of single-cell transcriptome analysis. *Nat Methods* 2011; 8: 6-11.

17. Saliba A-E, Westermann AJ, Gorski SA, Vogel J. Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res* 2014; **42**: 8845-8860

18. Etzrodt M, Endele M, Schroeder T. Quantitative single-cell approaches to stem cell research. *Cell Stem Cell* 2014; **15**: 546-558.

19. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, et al. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* 2014; **343**: 776-779.

20. Hu M, Krause D, Greaves M, Sharkis S, Dexter M, Heyworth C, et al. Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev* 1997; **11**: 774-785.

21. Brady G, Billia F, Knox J, Hoang T, Kirsch IR, Voura EB, et al. Analysis of gene expression in a complex differentiation hierarchy by global amplification of cDNA from single cells. *Curr Biol* 1995; **5**: 909-922.

22. Miyamoto T, Iwasaki H, Reizis B, Ye M, Graf T, Weissman IL, et al. Myeloid or lymphoid promiscuity as a critical step in hematopoietic lineage commitment. *Dev Cell* 2002; **3**: 137-147.

23. Akashi K, He X, Chen J, Iwasaki H, Niu C, Steenhard B, et al. Transcriptional accessibility for genes of multiple tissues and hematopoietic lineages is hierarchically controlled during early hematopoiesis. *Blood* 2003; **101**: 383-389.

24. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. *Nature* 2001; **414**: 105-111.

25. Akashi K, Traver D, Miyamoto T, Weissman IL. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature. 2000;404(6774):193-197.

26. Kondo M, Weissman IL, Akashi K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* 1997; **91**: 661-672.

27. Karsunky H, Inlay MA, Serwold T, Bhattacharya D, Weissman IL. Flk2+ common lymphoid progenitors possess equivalent differentiation potential for the B and T lineages. *Blood*. 2008; **111**: 5562-5570.

28. Serwold T, Ehrlich LIR, Weissman IL. Reductive isolation from bone marrow and blood implicates common lymphoid progenitors as the major source of thymopoiesis. *Blood* 2009; **113**: 807-815.

29. Balciunaite G, Ceredig R, Massa S, Rolink AG. A B220+ CD117+ CD19±hematopoietic progenitor with potent lymphoid and myeloid developmental potential. *Eur J Immunol* 2005; **35**: 2019-2030.

30. Adolfsson J, Månsson R, Buza-Vidas N, Hultquist A, Liuba K, Jensen CT, et al. Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential: a revised road map for adult blood lineage commitment. *Cell* 2005; **121**: 295-306.

31. Arinobu Y, Mizuno S-i, Chong Y, Shigematsu H, Iino T, Iwasaki H, et al. Reciprocal activation of GATA-1 and PU. 1 marks initial specification of hematopoietic stem cells into myeloerythroid and myelolymphoid lineages. *Cell stem cell* 2007; **1**: 416-427.

32. Katsura Y. Redefinition of lymphoid progenitors. *Nat Rev Immunol* 2002; **2**: 127-132.

33. Lai AY, Kondo M. Asymmetrical lymphoid and myeloid lineage commitment in multipotent hematopoietic progenitors. *J Exp Med* 2006; **203**: 1867-1873.

34. Ye M, Graf T. Early decisions in lymphoid development. *Curr Opin Immunol* 2007; **19**: 123-128.

35. Adolfsson J, Borge OJ, Bryder D, Theilgaard-Mönch K, Åstrand-Grundström I, Sitnicka E, et al. Upregulation of Flt3 expression within the bone marrow Lin− Sca1+ c-kit+ stem cell compartment is accompanied by loss of self-renewal capacity. *Immunity* 2001; **15**: 659-669.

36. Sanjuan-Pla A, Macaulay IC, Jensen CT, Woll PS, Luis TC, Mead A, et al. Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature* 2013; **502**: 232-236.

37. Pina C, Fugazza C, Tipping AJ, Brown J, Soneji S, Teles J, et al. Inferring rules of lineage commitment in haematopoiesis. *Nat Cell Biol* 2012; 14: 287-294.

38. Teles J, Pina C, Edén P, Ohlsson M, Enver T, Peterson C. Transcriptional regulation of lineage commitment–a stochastic model of cell fate decisions. *PLoS Comput Biol* 2013; **9**: e1003197.

39. Månsson R, Hultquist A, Luc S, Yang L, Anderson K, Kharazi S, et al. Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors. *Immunity* 2007; 26: 407-419.

40. Mansson R, Zandi S, Welinder E, Tsapogas P, Sakaguchi N, Bryder D, et al. Single-cell analysis of the common lymphoid progenitor compartment reveals functional and molecular heterogeneity. *Blood* 2010; **115**: 2601-2609.

41. Zandi S, Åhsberg J, Tsapogas P, Stjernberg J, Qian H, Sigvardsson M. Single-cell analysis of early B-lymphocyte development suggests independent regulation of lineage specification and commitment *in vivo*. *PNAS* 2012; **109**: 15871-15876.

42. Mingueneau M, Kreslavsky T, Gray D, Heng T, Cruse R, Ericson J, et al. The transcriptional landscape of αβ T cell differentiation. *Nat Immunol* 2013; **14**: 619-632.

43. Qiu P, Simonds EF, Bendall SC, Gibbs Jr KD, Bruggner RV, Linderman MD, et al. Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat Biotechnol* 2011; **29**: 886-891.

44. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. Pseudo-temporal ordering of individual cells reveals dynamics and regulators of cell fate decisions. *Nat Biotechnol* 2014; **32**: 381-386.

45. Bendall SC, Davis KL, Amir E-aD, Tadmor MD, Simonds EF, Chen TJ, et al. Single-Cell Trajectory Detection Uncovers Progression and Regulatory Coordination in Human B *Cell Dev Cell* 2014; **157**: 714-725.

46. Shin J, Berg DA, Zhu Y, Shin JY, Song J, Bonaguidi MA, et al. Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades underlying Adult Neurogenesis. *Cell Stem Cell* 2015; **17**: 360-372.

47. Wilson NK, Foster SD, Wang X, Knezevic K, Schütte J, Kaimakis P, et al. Combinatorial Transcriptional Control In Blood Stem/Progenitor Cells: Genome-wide Analysis of Ten Major Transcriptional Regulators. *Cell Stem Cell* 2010; **7**: 532-544.

48. Tijssen Marloes R, Cvejic A, Joshi A, Hannah Rebecca L, Ferreira R, Forrai A, et al. Genome-wide Analysis of Simultaneous GATA1/2, RUNX1, FLI1, and SCL Binding in Megakaryocytes Identifies Hematopoietic Regulators. *Dev Cell* 2011; **20**: 597-609.

49. Moignard V, Woodhouse S, Haghverdi L, Lilly AJ, Tanaka Y, Wilkinson AC, et al. Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat Biotechnol* 2015: **33**: 269-276.

50. Shalek AK, Satija R, Shuga J, Trombetta JJ, Gennert D, Lu D, et al. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* 2014; **510**: 363-369.

51. Stegle O, Teichmann SA, Marioni JC. Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet* 2015; **16**: 133-145.

52. de Sousa Abreu R, Penalva LO, Marcotte EM, Vogel C. Global signatures of protein and mRNA expression levels. *Mol BioSyst* 2009; **5**: 1512-1526.

53. Maier T, Güell M, Serrano L. Correlation of mRNA and protein in complex biological samples. *FEBS letters* 2009; **583**: 3966-3973.

54. Lundberg E, Fagerberg L, Klevebring D, Matic I, Geiger T, Cox J, et al. Defining the transcriptome and proteome in three functionally different human cell lines. *Mol Syst Biol* 2010; **6**: 450.

55. Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J,Wolf J, Chen W, Selbach M. Global quantification of mammalian gene expression control. *Nature* 2011; **473**: 337–342.

56. Li JJ, Bickel PJ, Biggin MD. System wide analyses have underestimated protein abundances and the importance of transcription in mammals. *PeerJ* 2014; **2**: e270

Figure legends:

Figure 1: Bulk transcriptomics masks heterogeneity at the single cell level.

Our knowledge of transcriptomes of various haematopoietic lineages has been advanced mainly by population-level analysis. However, a population of seemingly homogenous haematopoietic cells may include many distinct cell types. Thus, the transcriptome of an individual cell may differ substantially from the averaged transcriptome of the population of cells. These transcriptional differences between individual cells likely have important functional consequences by altering the ability of cells to make diverse differentiation decisions.

Figure 2: Gradual transition of cells between different cell states.

Rather than a process of abrupt transition between strictly defined stages, haematopoiesis is more likely a continuous process both on a transcriptional and functional level. Ultimately, at the end of this process, functional, fully differentiated, lineage restricted cells are formed.