Title: Demystifying blood stem cell fates

Authors:
Fiona K. Hamey and Berthold Göttgens

Affiliations:
Department of Haematology, Wellcome Trust–Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, Cambridge Institute for Medical Research, University of Cambridge, Cambridge CB2 0XY, United Kingdom.

Determining the differentiation potential of stem and progenitor cells is essential for understanding their function, yet our ability to do so is limited by the restrictions of experimental assays. Based on single-cell functional and molecular profiling experiments, a new computational approach shows how lineage commitment may occur in human haematopoiesis.

Diagrams depicting haematopoiesis as a branching process, with stem cells at the top of a differentiation hierarchy, are a familiar sight in any textbook referring to the haematopoietic system[1]. Developed and refined over the past few decades, this classical model proposes that haematopoietic stem cells (HSCs) undergo stepwise differentiation into discrete progenitor populations on their journey to mature blood cells. Yet, in recent years, advances in single-cell technologies have provided new insights challenging this classical view of haematopoiesis, which instead argue for a more continuous process and against the existence of some of the previously defined progenitor populations. In this issue, Velten and colleagues[2] investigate the fate decisions made during human adult haematopoiesis by presenting combined analysis of single-cell transcriptional and functional data. Similar to a recent report[3], they find that whilst multipotent cells exist in the adult stem cell compartment, the majority of blood progenitor cells are in fact unipotent. This observation is in marked contrast to the traditional branching model with its stepwise progression from stem cells via multipotent to bipotent and finally unipotent progenitors.

To define the gene expression states associated with cell fate decision making in adult human haematopoiesis, the authors generated single-cell RNA-sequencing (scRNA-Seq) profiles by sampling the stem and progenitor compartment from adult human bone marrow. Importantly, the authors used index flow sorting[4] so that each single-cell molecular profile could be assigned to a classically-defined cell type based on combinatorial surface marker expression. This approach revealed that cells within the more mature progenitor compartment (defined based on surface marker expression) could be separated into discrete cell states by approaches including clustering algorithms, with different states showing gene expression that suggested clear lineage priming. By contrast, less mature cells did not resolve into similarly clear populations when analysed with the same bioinformatic tools. The less mature cell compartment included cells with surface marker phenotypes historically associated with cells such as multipotent progenitors (MPPs) and multilymphoid progenitors (MLPs), which retain the multipotent state of HSCs, yet have lost the potential for long term self-renewal. These findings therefore hinted at an incompatibility with the classical branching model, as there was little evidence for cell states that might occupy intermediate bi- or tri-potent states on the journey from multipotent to unipotent cells.

It is worth highlighting the highly interactive website that accompanies the paper, which allows researchers to perform basic analyses on data from the study without the need for any specialist bioinformatics expertise. To process the raw data from scratch would require investing considerable time and effort, hence making the data available for online exploration represents a highly valuable resource. A particular asset for the haematopoiesis community (which we hope will continue to be available online) is that the flow cytometry index data are linked with the scRNA-Seq molecular profiles, so that any user can query gene expression and/or molecular heterogeneity for the exact populations of human blood stem and progenitor cells that for over a decade scientists in the field have worked hard to define.

As cells differentiate, they undergo noticeable changes in gene expression, which are reflected in the molecular profiles measured by scRNA-Seq. This raises the exciting possibility of utilising single-cell snapshot data to either validate or invalidate the proposed structure of the haematopoietic hierarchy. As individual profiles represent expression states of real cells progressing through differentiation, software can be developed with the specific aim of exploiting the strength of similarities between single-cell profiles to construct putative differentiation journeys. Existing algorithms, such as diffusion pseudotime (DPT)[5] and Wishbone[6], aim to organize differentiating cells along branching trajectories. For this study on human haematopoiesis the authors developed a new approach specifically suited to a system with extensive pre-existing knowledge about the start and end points of individual trajectories, which does not assume the constraint of a branching structure. These requirements led to the innovation of the STEMNET algorithm, which uses genes specifically expressed in different blood lineages to score each cellular profile based on its lineage bias and strength of commitment. Assessment of these scores across the whole data set enables identification of cells that possess the potential for multiple fates, and also of the precise combinations of multi- or bipotent fates that exist.

Importantly, scRNA-Seq profiles report expression states in unperturbed haematopoiesis, for which specific fate tracking of individual stem or progenitor cells has been impossible to date. Classical functional assays involve a perturbation of the system, such as placing cells into *in vitro* culture or transplanting them into an irradiated and therefore abnormal bone marrow environment. For this reason, functional assays report the potential of a cell in those specific assay conditions, rather than its fate in unperturbed haematopoiesis. Bearing in mind these caveats, it is nevertheless critical that computational predictions should be accompanied by functional validation. A particular challenge of high dimensional transcriptomics data is that profiling gene expression of a cell via RNA-Seq necessitates its destruction, thereby ruling out the possibility of testing its potential in any way. To circumvent this problem, the authors used the same cell surface markers, measured by index sorting, for the functional assays and scRNA-Seq, an approach that has

previously been described for studies in mouse[4,7]. The combined results of *in vitro* culture and xenotransplantation assays with the STEMNET predictions challenged the classical step-by-step branching model of haematopoiesis. Strikingly, the authors showed a lack of well-defined hierarchically organized discrete progenitor populations, with the majority of cells in fact appearing to be either multipotent, or committed to just a single fate. The uncommitted cells were described as CLOUD-HSPCs (a Continuum of LOw-primed UnDifferentiated haematopoietic stem and progenitor cells) as they could not be separated based on lineage commitment, but instead seemed to reside in a more continuous compartment (Figure 1). However, it remains to be established to what extent a continuous change in gene expression indicates a continuous change in potential, as variations in other factors that may not be directly reflected in the transcriptome of a cell are likely to play an important role in its fate restriction. Another possibility is that transcriptional noise in biologically irrelevant genes may obscure expression changes in a small number of key regulators from current bioinformatic analysis tools. Such changes in key regulators may in turn define discrete, rather than continuous cellular states within the immature HSPC compartment.

The results of single-cell culture assays described in this paper echo those published last year by Dick and colleagues[3], who contrasted the output of culturing single human HSPCs in colony assays at different developmental time points. Both studies found that in human adult bone marrow the majority of bi- or multipotent cells (based on colony output) were restricted to the stem and early progenitor compartment, with only small numbers of later progenitor cells giving rise to multiple lineages. Although the two studies observed that differentiation potential was clearly reduced as cells matured from the earlier to the later compartment, it is worth noting that both still found a small proportion of bi- and multipotent cells present in the progenitor compartment. The detection of even small numbers of multipotent progenitors raises the question of what makes these cells different from the majority of unipotent cells? Is this small percentage a true reflection of the *in vivo* potential or is it linked to limitations of culture assays? Finding an answer to these questions will pose a challenge especially in the human system, for which it is currently

impossible to assay the 'true' *in vivo* potential of a cell, as all conclusions are derived from *in vitro* and xenotransplantation experiments. Additionally, assays are limited by their ability to only resolve broad classes of cell types (erythroid, megakaryocytic and myeloid), without the full range of more specific fates that can be detected based on gene expression states.

Applied to single-cell molecular profiles, methods such as STEMNET promise the exciting ability to look into the future of an individual cell, predicting all its possible fates. Intriguingly, methods are also emerging that will allow us to look back into the history of a cell. Transgenic CrispR scarring approaches for example can generate barcode sequences that are maintained over subsequent generations and can therefore be used to retrospectively assign lineage relationships[8]. While transgenic methods are not applicable in human, their use in model systems such as the mouse may still provide insights of relevance to human biology. Rapid developments in 'multi-omics' single-cell techniques[9,10,11,12] will allow the determination of multiple properties of individual cells (for instance, through parallel sequencing of a single cell's transcriptome and genome). When combined with appropriate computational methods, these approaches open up promising new opportunities to capture the past as well as predict the future of a single cell from static snapshot data alone, and thus substantially advance our ability to gain new insights into how cell fate decisions are regulated across multiple biological systems.

References:

1. Orkin, S. H. & Zon, L. I. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631-644 (2008).

2. Velten, L., Haas, S. F., Raffel, S., Blaszkiewicz, S., Islam, S., Hennig, B. P., Hirche, C., Lutz, C., Buss, E. C., Nowak, D., Boch, T., Hofmann, W.-K., Ho, A. D., Huber, W., Trumpp, A., Essers, M. A. G. & Steinmetz, L. M. Human haematopoietic stem cell lineage commitment is a continuous process *Nat. Cell Biol.* **XXX**, XXX-XXX (2017).

3. Notta, F., Zandi, S., Takayama, N., Dobson, S., Gan, O. I., Wilson, G., Kaufmann, K. B., McLeod, J., Laurenti, E., Dunant, C. F., McPherson, J. D., Stein, L. D., Dror, Y. & Dick, J. E. Distinct routes of lineage development

reshape the human blood hierarchy across ontogeny. *Science* **351**, aab2116 (2016).
4. Wilson, N. K., Kent, D. G., Buettner, F., Shehata, M., Macaulay, I. C., Calero-Nieto, F. J., Sanchez Castillo, M., Oedekoven, C. A., Diamanti, E., Schulte, R., Ponting, C. P., Voet, T., Caldas, C., Stingl, J., Green, A. R., Theis, F. J. & Göttgens, B. Combined single-cell functional and gene expression analysis resolves heterogeneity within stem cell populations. *Cell Stem Cell* **16**, 712-724 (2015).
5. Haghverdi, L., Buttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845-848 (2016).
6. Setty, M., Tadmor, M. D., Reich-Zeliger, S., Angel, O., Salame, T. M., Kathail, P., Choi, K., Bendall, S., Friedman, N., & Pe'er, D. Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nat. Biotechnol.* **34**, 637-645 (2016).
7. Nestorowa, S., Hamey, F. K., Pijuan Sala, B., Diamanti, E., Shepherd, M., Laurenti, E., Wilson, N. K., Kent, D. G. & Göttgens, B. *Blood* **128**, e20-31 (2016).
8. McKenna, A., Findlay, G. M., Gagnon, J. A., Horwitz, M. S., Schier, A. F., & Shendure, J. *Science* **353**, aaf7907 (2016).
9. Macaulay, I. C., Teng, M. J., Haerty, W., Kumar, P., Ponting, C. P. & Voet T. Separation and parallel sequencing of the genomes and transcriptomes of single cells using G&T-seq. *Nat. Methods* **12**, 519-522 (2015).
10. Dey, S. S., Kester, L., Spanjaard, B., Bienko, M. & van Oudenaarden, A. Integrated genome and transcriptome sequencing of the same cell. *Nat. Biotechnol.* **33**, 285-289 (2015).
11. Angermueller, C., Clark, S. J., Lee, H. J., Macaulay, I. C., Teng, M. J., Hu, T. X., Krueger, F., Smallwood, S. A., Ponting, C. P., Voet, T., Kelsey, G., Stegle, O. & Reik, W. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat Methods* **13**, 229-232 (2016).
12. Hou, Y., Guo, H., Cao, C., Li, X., Hu, B., Zhu, P., Wu, X., Wen, L., Tang, F., Huang, Y. & Peng, J. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Research*. **26**, 304-319 (2016).

Figure legend:
Left hand side: In the classical model of haematopoiesis, haematopoietic stem cells (HSCs) undergo stepwise differentiation, first losing their self-renewal potential and differentiating into multipotent progenitors (MPPs) before further commitment into pools of progenitors committed to different lineages. Right hand side: Alternatively, Velten and colleagues[2] propose a model of a Continuum of LOw-primed UnDifferentiated ("CLOUD")-HSPCs, in which differentiation does not occur as a branching process. The continuum consists of un-committed cells in transitory states, which gradually acquire transcriptomic lineage priming, ultimately resulting in the emergence of differentiated uni-lineage progenitors.