Hematopoiesis in the age of the single cell.

Sten Eirik W. Jacobsen[1-5] and Claus Nerlov[1].

[1]MRC Molecular Hematology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DS, UK. [2]Haematopoietic Stem Cell Laboratory, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DS, UK. [3]Department of Cell and Molecular Biology, Wallenberg Institute for Regenerative Medicine, Karolinska Institutet, Stockholm SE-171 77, Sweden. [4]Department of Medicine Huddinge, Center for Hematology and Regenerative Medicine, Karolinska Institutet, Stockholm SE-171 77, Sweden. [5]Karolinska University Hospital, Stockholm SE-171 77, Sweden.

Correspondence:
Professor Claus Nerlov
MRC Molecular Hematology Unit
MRC Weatherall Institute of Molecular Medicine
University of Oxford John Radcliffe Hospital
Oxford OX3 9DS, United Kingdom
Email: claus.nerlov@imm.ox.ac.uk
Phone: +44 1865222324

Professor Sten Eirik W. Jacobsen
Department of Medicine Huddinge
Center for Hematology and Regenerative Medicine
Karolinska Institutet
Karolinska University Hospital
Stockholm SE-141 86, Sweden
Email: sten.eirik.jacobsen@ki.se
Phone: +46 766507816

Abstract.

The molecular and functional characterization of single cells at scale has emerged as a key driver of tissue biology. It is therefore important to understand the properties and limitations of the single cell transcriptomics, molecular barcoding and functional assays used to study cellular properties at the single cell level. We here review recent studies from the hematopoietic system, and discuss how to interpret and integrate data obtained with these different technologies.

Introduction.

The hematopoietic system is one of the most well-studied and well-characterized differentiation hierarchies in the mammalian body, and hematopoiesis has for decades served as a paradigm for how stem cells maintain complex, self-renewing tissues under homeostatic, physiological stress and regenerative conditions. Progress in studying hematopoiesis has been facilitated by the development of robust clonal assays for hematopoietic stem- and progenitor cell function, and by the development of multi-parameter flow cytometry and cell sorting, together allowing the systematic functional characterization of highly purified cell populations. Using these methodologies hematopoietic stem cells (HSCs), defined as cells capable of reconstituting multi-lineage hematopoiesis in lethally irradiated recipients upon transplantation, were first identified by Till and McCulloch[1], isolated by Weissman and colleagues[2], and subsequently purified to increasing degrees of homogeneity[3,4]. Similarly, the discovery of hematopoietic cytokines[5] enabled the development of lineage-specific colony-forming cell (CFC) assays, allowing progenitor cells to be isolated and their lineage potentials individually measured. This powerful combination of technologies allowed hierarchical models of the hematopoietic system to be generated[6-8], underpinning numerous molecular and genetic studies into the development and regulation of the hematopoietic system and the etiology of hematological disorders[9]. In all proposed hierarchies hematopoiesis originates in multi-potent repopulating HSCs, and lineage development occurs via discrete and increasingly lineage-restricted progenitor populations. However, as the underlying studies predominantly involved the analysis of purified cell populations, as opposed to individual cells, functional heterogeneity within the identified stem- and progenitor populations may have been underestimated. In addition, while transplantation of irradiated recipients with purified prospective HSC populations has proven a powerful assay for defining repopulating HSCs it may not be able to de-convolute HSC heterogeneity.

Recently a number of studies have explored technologies that have the potential to overcome these limitations. Single cell transcriptome analysis has allowed molecular heterogeneity to be systematically explored across the hematopoietic system. In addition, lineage tracing and *in vivo* barcoding have emerged as means to study the lineage fates and lineage contributions of individual stem– and progenitor cells within larger populations, including their roles in unperturbed hematopoiesis. Together, these technologies have the potential to considerably advance our understanding of the development and physiology of hematopoiesis, as well as other tissues. However, as is the case with the classical technologies used to investigate hematopoiesis, also these approaches have technical and conceptual limits that need to be borne in mind when interpreting the results they generate. To this end we will here discuss the key concepts used to functionally define hematopoietic stem- and progenitor cell types, and how these can be applied to data generated by both classical and emerging technologies.

Key concepts, current technologies and their limitations.

A number of concepts frequently used to functionally annotate stem- and progenitor cells based on their behavior in the assays described above, are listed in Table 1. An important consideration when interpreting experimental data in light of these concepts is how the properties of the assays may influence the readout, and therefore limit the conclusions that can be reached. For example, while HSC transplantation assays is a powerful tool to assess the ability of putative HSCs to repopulate the hematopoietic system, this assays does not address the extent to which the transplanted cells contributed to

hematopoiesis in their native, steady-state setting. Indeed, highly quiescent HSCs (which by inference generate little or no cellular output) have the highest repopulating capacity[10,11]. Likewise, the *in vivo* lineage fates and the *in vitro* lineage potentials revealed by the same single HSCs may differ greatly[12], as may their lineage fates in different *in vivo* settings.

For progenitors, a key issue is the measurement of lineage potentials and lineage fate, and establishing the relationship between the two. Here, their limited proliferative potential, combined with the distinct, transient kinetics with which different lineages are replenished, makes *in vivo* readout of multiple lineage fates particularly challenging, and *in vitro* CFC and other clonal assays remain the principal tool. However, individual progenitor cells do not necessarily express all their available potentials *in vitro*[13]. Therefore, while positive lineage readout for an individual cell demonstrates the presence of a potential, a negative readout does not necessarily demonstrate that the potential was absent. Consequently, lineage restriction can only be demonstrated through the consistent absence of specific lineage readouts across an entire cell population, requiring its prospective purification. Importantly, similarly to what is observed for lineage biased HSCs[12], progenitor lineage fate can be distinct from lineage potential, as not all fate options may be utilized under a given condition, with the additional challenge that combined potentials of a progenitor need to be established at the single cell level.

Defining the self-renewing component(s) of hematopoiesis at the single cell level.

HSC heterogeneity was initially identified by careful tracking of myeloid and lymphoid lineage output from transplanted single HSC-derived cell populations or single HSCs, observing that HSC clones existed that generated a higher proportion of myeloid (M-biased) or lymphoid cells (L-biased) compared to the HSC population average, a bias stable during up to 3 serial transplantations[14,15], identifying HSC lineage bias as an intrinsically programmed HSC property. However, while transplanting experimental cells with a CD45.2 allotype into congenic CD45.1 recipients allows highly sensitive and specific tracking in peripheral blood of CD45.2 myeloid (M; granulocyte-monocyte), B and T lymphoid cells derived from a transplanted single HSC, mature erythrocytes (E) and platelets (P) lack CD45 expression, and replenishment of these lineages by experimental HSCs could therefore not be tracked in these previous studies. More recent studies using a transgenic reporter to track P-lineage replenishment by transplanted HSCs suggested that a large fraction of LT-HSCs previously classified as M-biased are rather P- or PM-biased[16], and that functional HSC P-bias increases significantly with age[17]. Other studies, using a fluorescent reporter enabling detection of all (i.e. P, E, M, B and T) lineage outputs, identified progenitors capable of extended, lineage-restricted P, PE and PEM repopulation[18]. Most of these progenitors did not sustain stable lineage output in secondary recipients, distinguishing them from long term (LT-) HSCs with extensive self-renewal. Also, although their lineage potentials were not investigated *in vitro*, it was concluded that these were lineage committed oligo-potent and uni-potent progenitors, and therefore that the stem cell properties of self-renewal and multi-potency could be segregated. More recent single HSC transplantation studies revealed, in addition to HSCs repopulating all blood cell lineages, LT-HSCs fully and stably restricted to P, PE, PEM or PEMB fates (but not to any other linage combinations)[12]. In particular, the M-biased HSCs previously identified (based on tracking of myeloid and lymphoid, but not platelet and erythroid output) gave rise to even higher levels of platelet, compared to myeloid, reconstitution. While stably fate-restricted *in vivo,* these HSCs remained in all cases multi-potent, as potentials not expressed upon *in vivo* transplantation could easily be revealed upon *in vitro* culture, and in some instances also upon secondary transplantation. Multi-potency therefore seems to be a general property of LT-HSCs, and HSC *in vivo* fate-restriction appears to be guided by strict rules determining possible and apparently hierarchically related lineage combinations.

Studying HSCs in native hematopoiesis: single cell and population-based strategies.

Single-cell transplantation has thus proven a highly valuable tool in the unraveling of HSC heterogeneity. However, the above studies also show that fully understanding HSC properties will require a combination of assays, including both *in vivo* and *in vitro* single cell-based analysis. In particular, while transplantation can accurately identify repopulating HSCs this assay does not necessarily inform on their fate during native hematopoiesis. In particular, some HSCs may be highly quiescent, and therefore contribute minimally to steady-state hematopoiesis, while demonstrating high repopulating activity upon transplantation[10,11]. To measure steady-state hematopoietic output with single cell resolution cellular clones have been labeled through molecular barcoding. One methodology used is the induction of Sleeping Beauty-mediated transposition, leading to the generation of unique transposon insertions in transposed cells that can be identified through ligation-mediated PCR[19,20]. This is a powerful tool for assigning the clonal relationships of mature hematopoietic cell types, and hematopoietic stem- and progenitor cells. Also this technique identified clones originating in phenotypic HSCs restricted to platelet lineage output, consistent with P-biased HSCs identified by single cell transplantation[12,16] being active during steady-state hematopoiesis. However, also barcoding studies carry intrinsic limitations. Barcoding techniques are not yet fully quantitative, and in steady state the size of many clones is likely to be small and close to the detection level of the method, so negative findings with regard clonal involvement of different lineages must be interpreted with caution. This is particularly relevant when seeking to address whether or not a stable hematopoietic clone is sustained independently of HSCs or not. The clone size at the apex of a barcoded hierarchy would typically consist of very few or even a single cell and may therefore not be consistently detected. This, along with the current inability to reliably define HSCs and multi-potent progenitor cells by virtue of cell surface phenotype alone, makes it challenging to accurately determine how a clone is sustained by barcoding alone.

Pei et al.[21] addressed this by using a LoxP recombination-based molecular barcode through induced by an HSC-selective *Tie2-Cre* driver: here the barcode was generated in the HSC compartment and barcoded output could therefore be definitively traced to single HSCs. Consistent with previous studies[14,15] both myeloid-biased and lymphoid-biased HSC-derived clones were detected. However, lymphoid-biased HSC clones generally fail to serially repopulate[15], and do not sustain HSCs[12]. An open and interesting question in unperturbed hematopoiesis is therefore to which extent barcoded clones with lymphoid bias maintain the originally labeled HSC.

The importance of cautious and alternative interpretation of data generated by single cell technologies is illustrated by the example shown in Figure 1, where the same sustained population of a subset of hematopoietic lineages by a cellular clone in a barcoding experiment (Fig 1a), based on the existing literature could be sustained by either a fate-restricted multi-potent LT-HSC[12] (Fig 1b), a fate-restricted multi-potent progenitor with limited self-renewal, (Fig 1c) or a oligo-potent progenitor with more limited self-renewal[18] (Fig 1d). A more definitive understanding of how native hematopoiesis is sustained will therefore require an integrated approach combining the herein described single cell-based technologies, and in particular combining *in vivo* fate mapping with assays that determine the lineage restriction and self-renewal properties of the cells sustaining clonal output.

The correspondence principle: studying HSC populations.

Single cell-level analysis is critical in order to understand the heterogeneity underlying systems behavior. In that regard, the collective properties of single cells should correspond to those of the entire system. To address this quantitative measurement of the lineage replenishment from the HSC population during steady-state hematopoiesis through lineage tracing has been employed. Here the aim is activation of a Cre recombinase-inducible reporter specifically in HSCs in order to permanently label HSCs as well as their progeny. The critical issue with this approach (and therefore also when interpreting the results) is ensuring that Cre is expressed in repopulating HSCs in a representative manner covering the functional diversity of HSCs[12]. In addition it is important to avoid labeling of down-stream progenitors. If this is not

demonstrated, interpretations of the results will have to take these significant limitations into consideration. Using *Tie2*-driven inducible Cre Busch et al.[22] were able to label a subset (ca. 0.5%) of HSCs in young adult mice, and observed that their replenishment of downstream progenitors and mature cells was very slow, suggesting that steady-state hematopoiesis is sustained largely independently of HSCs. In contrast, Sawai et al. used a *Pdzk1ip1* driven Cre to quantitatively label HSCs, and observed near-complete labeling of all definitive hematopoietic lineages within a few months[23], suggesting that steady-state blood lineage replenishment is highly HSC-dependent. This discrepancy could be explained by the above-described limitations. In the studies of Busch et al.[22] *Tie2* may be expressed in a select subset of HSCs. This possibility is supported by the observation that few Tie2+ HSCs repopulate lymphoid lineages when transplanted at limiting dilution[22] compared to cells from an unselected HSC population[15]. In contrast, Sawai et al. observed balanced repopulation across all definitive hematopoietic lineages, but here the labeling of a subset of down-stream progenitors cannot be excluded[23]. Interestingly, Sawai et al. observed clearly distinct kinetics of lineage repopulation from lineage traced HSCs, with platelets being the most rapidly reconstituted lineage. While this correlates with the observation that HSC barcoding initially labels mostly platelet-lineage progenitors it also raises the possibility that pervasive P-biased HSC readout at early time points after barcoding[20] is due to the more rapid kinetics of replenishment of this lineage, rather than specific readout of P-restricted HSCs. Indeed, most barcoded HSCs with early P-bias displayed multi-lineage repopulation upon transplantation[20].

Overall, HSC barcoding shows similar HSC heterogeneity as that observed in single HSC transplantation, indicating that the different subsets of HSCs identified through single cell transplantation might all be actively employed in native hematopoiesis. This is also consistent with the observation by Yu et al.[24] that repopulation patterns of native HSCs identified by multicolor fluorescent tagging were preserved upon transplantation. However, to fully reconcile the different observations a systematic integrated approach where lineage tracing is combined with single cell transplantation of the labeled HSCs would likely be required.

Single cell transcriptome analysis and barcoding of progenitors: how to decipher molecular and functional heterogeneity.

Defining the cellular pathways and progenitor cell populations through which hematopoietic lineages are generated remains an important task: lineage relationships provide important information about the evolutionary ontogeny of hematopoietic cell types, and knowledge of the branch points where lineages segregate facilitates unraveling of the molecular mechanisms by which cellular fates are specified. Furthermore, progenitor cells are templates for cellular transformation in many hematopoietic malignancies, and knowledge of their molecular characteristics allows accurate comparison of normal and transformed progenitors, facilitating identification of transformation-specific therapeutic targets. Particular challenges exist when studying progenitor behavior: progenitor cells are by their nature short-lived with little or no self-renewal potential, and therefore their ability to repopulate hematopoiesis in a transplant setting limited, in particular for short-lived non-lymphoid cell types. Therefore progenitor populations have generally been characterized using *in vitro* CFC and other clonal assays and transient *in vivo* repopulation.

Using these classical technologies the Weismann laboratory identified multi-potent cell populations that exclusively contained all lymphoid (common lymphoid progenitor (CLP))[7] or all myelo-erythroid lineage potentials (common myeloid progenitor (CMP))[6]. This indicated a hierarchical model where the initial lineage bifurcation downstream of the HSC would strictly separate the lymphoid and myelo-erythroid lineages, with subsequent sequential branching leading to specification of individual lineages via intermediate GMPs (granulocyte-macrophage progenitors) and MEPs (megakaryocyte-erythroid progenitors) (Fig 2a). Subsequent studies identified additional sub-populations with predominant erythroid (CFU-E, preCFU-E) or myeloid (preGM) potentials[8], but also progenitors with combined lymphoid and myeloid potential, but lacking erythroid/megakaryocyte potentials (lymphoid-primed multi-potent

progenitor or LMPP)[25,26], as well as revised definitions of the CMP[27,28], resulting in a revised hierarchical model (Fig 2b). Although similar progenitor populations were identified in human hematopoiesis[29-31], it is important to bear in mind that differences may exist between human and murine hematopoiesis. For example, recent studies have proposed that adult human lineage development occurs mainly through multi- and mono-potent progenitors, with oligo-potent progenitors cells rarely being detected[32,33]. While these studies have the caveat that they are largely based on *in vitro* readouts that may not efficiently detect lineage co-potentials, they emphasize the need for further development of technologies to study human hematopoiesis, with ongoing and expanding gene therapy trials potentially providing an opportunity to obtain information about commitment pathways *in vivo*[34].

While these models proved a highly useful guide for numerous subsequent studies, and allowed flow cytometry-based phenotypic and functional progenitor analysis in genetic and pharmacological perturbation studies, they also highlighted key technical limitations to the methodologies used in the field. In particular, progenitor cell populations, even if defined by specific combinations of cell surface markers, may nevertheless be functionally heterogeneous. In addition, the *in vivo* and *in vitro* differentiation assays used to define the lineage potentials of individual cells do not detect these potentials with complete efficiency. Therefore, even if the overall lineage potentials present (or not present) in a population can be confidently measured *in vitro*, it cannot be reliably determined how many of these reside in each cell, making assessment of population homogeneity uncertain. Finally, while *in vitro* assays inform about lineage potential, they do not necessarily reflect cellular fate *in vivo*.

The latter issue was addressed by *in vivo* molecular barcoding of CMPs[35] and LMPPs[36]. These studies indicated that the vast majority of CMPs assumed either erythroid or myeloid fate *in vivo*, consistent with results obtained *in vitro* on CMPs sub-fractionated by cell sorting[8]. Similarly, most phenotypic LMPPs were observed to have a biased or restricted lineage fate *in vivo,* with production of myeloid, dendritic and B-cells highly correlated between cells derived from the same original LMPP[36]. While these studies helped unravel the functional hetereogeneity within phenotypically defined progenitor populations, they also confirmed the existence of progenitors within the investigated CMP and LMPP populations adopting the predicted combined lineage fates.

To access heterogeneity not resolved by surface markers comprehensive profiling and clustering of both murine and human hematopoietic progenitor cells using single cell RNA-seq has been performed[37-39]. These studies readily identified cellular subsets with lineage-selective gene expression, including putative erythroid, platelet, monocytic and granulocytic lineage-restricted progenitors, but did not reproducibly identify previously characterized multi- and oligo-potent cell types, including human LMPPs/MLPs and murine MPPs, that were shown to harbor distinct subset of lineage potentials[40]. A more sensitive approach to analysis of transcriptional lineage programming at the single cell level single is targeted PCR. Using this approach co-expression of myeloid and lymphoid genes was consistently observed in single LMPPs, whereas Flt3– MPPs and HSCs co-expressed megakaryocyte/erythroid and myeloid programs[26], demonstrating distinct lineage programming of MPP subsets, consistent with their *in vivo* lineage output[40]. More recently Gou et al.[41] used highly multiplexed single cell qRT-PCR to identify surface markers that separate transcriptionally and functionally distinct myelo-erythroid progenitor populations, demonstrating that targeted approaches to resolution of cellular heterogeneity are an important and viable alternative to whole transcriptome analysis.

Transcriptional lineage priming has become a common tool for the assessment of lineage potential, based on the initial observation by Enver and colleagues that multi-potent progenitor cells co-express, at the single cell level, genes characteristic of the lineages that they can differentiate into[42]. The computational assignment of lineage potentials to progenitor cells is therefore based on their expression of lineage-specific genes, building on the assumption that oligo- and multi-potent cell types will co-express the relevant lineage programs. However, it remains unclear to which extent expression of lineage-associated genes reflects cellular potential, and in addition the interplay between lineage programs is complex. In

particular, while the expression of lineage specific genes in multi-potent stem- and progenitor cells can predict corresponding lineage potentials, the opposite seems less clear. For example, while HSCs are myeloid and megakaryocyte lineage primed, they do not display distinct priming for the lymphoid lineage potentials they possess[25,26]. However, upon differentiation into lymphoid-primed multi-potent progenitors (LMPPs) the acquisition of lymphoid lineage priming coincides with loss of megakaryocyte-erythroid lineage potential[25,26], compatible with functional antagonism between these programs contributing to concomitant increased megakaryocyte lineage priming and loss of lymphoid biased HSCs with age[17]. Accordingly, antagonistic lineage programs may be sequentially, rather than simultaneously, expressed during differentiation, disguising latent potentials. It is therefore important when using lineage priming as a proxy for lineage potential that these factors are taken this account, and that predictions based on molecular signatures are functionally validated before interpreting their implications for lineage potentials and fates.

The identification of myeloid progenitor sub-populations with basophil and eosinophil lineage priming[37], distinct from the neutrophil/monocyte priming generally used as a proxy for myeloid lineage potential, raised the issue of how these distinct myeloid lineages segregate. Computational modeling of human single cell RNA-seq data suggested that all myeloid lineages derive from a single GMP, consistent with the predominant models of hematopoiesis[39]. However, combined single cell sequencing and single cell qRT-PCR of murine myeloid progenitors revealed GMP heterogeneity based on *Gata1* expression, a finding used to prospectively separate *Gata1*-negative progenitors with neutrophil/monocyte-restricted potential from *Gata1*-positive eosinophil/mast cell-restricted potential[43]. Furthermore, eosinophil/mast cell potential and megakaryocyte/erythroid potential were observed to co-segregate from monocyte/neutrophil and lymphoid potential, indicating that *Gata1*-expressing myeloid lineages (eosinophil, basophil, mast cell) segregate from *Gata1*-negative myeloid lineages (neutrophils, monocytes) very early during lineage specification (Fig 2c). These observations again underscore the importance of complementing computational analysis with the necessary functional validation studies.

Perspectives.

The application of single cell-based technologies to hematopoiesis clearly has significantly expanded and improved our knowledge of the system. Single cell transcriptome analysis, barcoding, *in vitro* clonal functional assays, and single cell transplantation has all allowed us access to molecular heterogeneity that would have been difficult or impossible to identify using conventional cell population-based technologies. An important challenge is now to use this new knowledge to understand how the hematopoietic system functions on a global scale. Importantly, as they all have intrinsic limitations, no current individual technology is likely to fully resolve the complexity of the hematopoietic system. Rather, it will require effective integration of different single cell molecular, genetic and functional technologies to understand the properties and roles played by distinct stem and progenitor populations in steady state, as well as during tissue stress and cellular transformation.

Acknowledgements.

We thank S. Orkin and S. Morrison for critical reading of the manuscript, and P. Woll for Figure preparation. Research in the Jacobsen and Nerlov laboratories is supported by Bloodwise, MRC and BBSRC.

Author Contributions.

S.E.W.J. and C.N. contributed equally to the writing and editing of the manuscript as well as to figure preparation.

Author Information.

The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence should be addressed to S.E.W.J. (sten.eirik.jacobsen@ki.se) or C.N. (claus.nerlov@imm.ox.ac.uk).

References.

1 McCulloch, E. A. & Till, J. E. The radiation sensitivity of normal mouse bone marrow cells, determined by quantitative marrow transplantation into irradiated mice. *Radiat Res* **13**, 115-125 (1960).

2 Spangrude, G. J., Heimfeld, S. & Weissman, I. L. Purification and Characterization of Mouse Hematopoietic Stem-Cells. *Science* **241**, 58-62, doi:DOI 10.1126/science.2898810 (1988).

3 Osawa, M., Hanada, K., Hamada, H. & Nakauchi, H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* **273**, 242-245 (1996).

4 Kiel, M. J. *et al.* SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109-1121 (2005).

5 Metcalf, D. Hematopoietic cytokines. *Blood* **111**, 485-491, doi:10.1182/blood-2007-03-079681 (2008).

6 Akashi, K., Traver, D., Miyamoto, T. & Weissman, I. L. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**, 193-197, doi:10.1038/35004599 (2000).

7 Kondo, M., Weissman, I. L. & Akashi, K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91**, 661-672 (1997).

8 Pronk, C. J. *et al.* Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. *Cell Stem Cell* **1**, 428-442 (2007).

9 Orkin, S. H. & Zon, L. I. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631-644, doi:10.1016/j.cell.2008.01.025 (2008).

10 Wilson, A. *et al.* Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* **135**, 1118-1129, doi:10.1016/j.cell.2008.10.048 (2008).

11 Bernitz, J. M., Kim, H. S., MacArthur, B., Sieburg, H. & Moore, K. Hematopoietic Stem Cells Count and Remember Self-Renewal Divisions. *Cell* **167**, 1296-1309 e1210, doi:10.1016/j.cell.2016.10.022 (2016).

12 Carrelha, J. *et al.* Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. *Nature* **554**, 106-111, doi:10.1038/nature25455 (2018).

13 Rieger, M. A., Hoppe, P. S., Smejkal, B. M., Eitelhuber, A. C. & Schroeder, T. Hematopoietic cytokines can instruct lineage choice. *Science* **325**, 217-218, doi:10.1126/science.1171461 (2009).

14 Muller-Sieburg, C. E., Cho, R. H., Thoman, M., Adkins, B. & Sieburg, H. B. Deterministic regulation of hematopoietic stem cell self-renewal and differentiation. *Blood* **100**, 1302-1309 (2002).

15 Dykstra, B. *et al.* Long-term propagation of distinct hematopoietic differentiation programs in vivo. *Cell Stem Cell* **1**, 218-229, doi:10.1016/j.stem.2007.05.015 (2007).

16 Sanjuan-Pla, A. *et al.* Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature* **502**, 232-236, doi:10.1038/nature12495 (2013).

17 Grover, A. *et al.* Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells. *Nat Commun* **7**, 11075, doi:10.1038/ncomms11075 (2016).

18 Yamamoto, R. *et al.* Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. *Cell* **154**, 1112-1126, doi:10.1016/j.cell.2013.08.007 (2013).

19 Sun, J. *et al.* Clonal dynamics of native haematopoiesis. *Nature* **514**, 322-327, doi:10.1038/nature13824 (2014).

20 Rodriguez-Fraticelli, A. E. *et al.* Clonal analysis of lineage fate in native haematopoiesis. *Nature* **553**, 212-216, doi:10.1038/nature25168 (2018).

21 Pei, W. *et al.* Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature* **548**, 456-460, doi:10.1038/nature23653 (2017).

22 Busch, K. *et al.* Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature* **518**, 542-546, doi:10.1038/nature14242 (2015).

23 Sawai, C. M. *et al.* Hematopoietic Stem Cells Are the Major Source of Multilineage Hematopoiesis in Adult Animals. *Immunity* **45**, 597-609, doi:10.1016/j.immuni.2016.08.007 (2016).

24 Yu, V. W. *et al.* Epigenetic Memory Underlies Cell-Autonomous Heterogeneous Behavior of Hematopoietic Stem Cells. *Cell* **168**, 944-945, doi:10.1016/j.cell.2017.02.010 (2017).

25 Adolfsson, J. *et al.* Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. *Cell* **121**, 295-306, doi:10.1016/j.cell.2005.02.013 (2005).

26 Mansson, R. *et al.* Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors. *Immunity* **26**, 407-419, doi:10.1016/j.immuni.2007.02.013 (2007).

27 Arinobu, Y. *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* **1**, 416-427, doi:10.1016/j.stem.2007.07.004 (2007).

28 Miyawaki, K. *et al.* CD41 marks the initial myelo-erythroid lineage specification in adult mouse hematopoiesis: redefinition of murine common myeloid progenitor. *Stem Cells* **33**, 976-987, doi:10.1002/stem.1906 (2015).

29 Manz, M. G., Miyamoto, T., Akashi, K. & Weissman, I. L. Prospective isolation of human clonogenic common myeloid progenitors. *Proc Natl Acad Sci U S A* **99**, 11872-11877, doi:10.1073/pnas.172384399 (2002).

30 Doulatov, S. *et al.* Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nat Immunol* **11**, 585-593, doi:10.1038/ni.1889 (2010).

31 Goardon, N. *et al.* Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer Cell* **19**, 138-152, doi:10.1016/j.ccr.2010.12.012 (2011).

32 Notta, F. *et al.* Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* **351**, aab2116, doi:10.1126/science.aab2116 (2016).

33 Karamitros, D. *et al.* Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. *Nat Immunol* **19**, 85-97, doi:10.1038/s41590-017-0001-2 (2018).

34 Biasco, L. *et al.* In Vivo Tracking of Human Hematopoiesis Reveals Patterns of Clonal Dynamics during Early and Steady-State Reconstitution Phases. *Cell Stem Cell* **19**, 107-119, doi:10.1016/j.stem.2016.04.016 (2016).

35 Perie, L., Duffy, K. R., Kok, L., de Boer, R. J. & Schumacher, T. N. The Branching Point in Erythro-Myeloid Differentiation. *Cell* **163**, 1655-1662, doi:10.1016/j.cell.2015.11.059 (2015).

36 Naik, S. H. *et al.* Diverse and heritable lineage imprinting of early haematopoietic progenitors. *Nature* **496**, 229-232, doi:10.1038/nature12013 (2013).

37 Paul, F. *et al.* Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell* **164**, 325, doi:10.1016/j.cell.2015.12.046 (2016).

38 Nestorowa, S. *et al.* A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood* **128**, e20-31, doi:10.1182/blood-2016-05-716480 (2016).

39 Velten, L. *et al.* Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol* **19**, 271-281, doi:10.1038/ncb3493 (2017).

40 Pietras, E. M. *et al.* Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions. *Cell Stem Cell* **17**, 35-46, doi:10.1016/j.stem.2015.05.003 (2015).

41 Guo, G. *et al.* Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. *Cell Stem Cell* **13**, 492-505, doi:10.1016/j.stem.2013.07.017 (2013).

42 Hu, M. *et al.* Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev* **11**, 774-785 (1997).

43 Drissen, R. *et al.* Distinct myeloid progenitor-differentiation pathways identified through single-cell RNA sequencing. *Nat Immunol* **17**, 666-676, doi:10.1038/ni.3412 (2016).

Table 1: Key concepts, brief definitions, assays used to measure them.

Concept	Definition	Assay(s)	Limitations
Repopulating stem cell	Cell that upon transplantation will stably repopulate all or part of the hematopoietic system	Transplantation of prospective HSCs into suitably conditioned recipient	The detection of HSCs in this assay may be limited by HSC engraftment, quiescence and the technical ability of the assay setup to detect all lineage outputs.
Self-renewal	The ability of a cell to perpetuate itself during extended cell divisions	Serial transplantation of purified putative self-renewing cells	Serial transplantation of unpurified HSCs (e.g. total bone marrow) can propagate long-lived lymphoid cells in the absence of detectable HSCs.
Lineage potential	The capacity of a cell to differentiate along (a) certain lineage(s) under appropriate conditions.	*In vitro* clonal or colony forming or *in vivo* transplantation assays that measure the relevant potential(s).	Not all potentials will necessarily be detected, due to distinct differentiation kinetics and/or culture requirements of lineages. A negative finding for a particular cell therefore does not demonstrate that the potential was not present.
Lineage restriction/commitment	A loss of lineage potential that restricts a cell to a subset of its original potentials.	*In vitro* clonal or colony forming or *in vivo* transplantation assays that measure both retained and lost potentials.	Conversely, a potential being undetected at the single cell level does not allow the conclusion that it was absent. Confirmation of loss of a potential requires the prospective purification of a cell population and consistent negative readout.
Lineage fate	The cellular fate(s) that a stem- or progenitor will give rise to.	Lineage fate is specific to a particular condition. It can only be measured under these (typically *in vivo*) conditions, and lineage fate may change if the conditions are altered.	Lineage fate cannot be inferred from assays that remove cells from the relevant environment. As such, fate during unperturbed hematopoiesis cannot be measured *in vitro* or by re-transplantation.
Lineage bias	A term applied to hematopoietic stem cells (HSCs) that, while nominally multi-potent, display a stable restriction or strong bias of their	Lineage bias of multi-potent HSCs has been demonstrated *in vivo* in serial transplantations initiated by a single HSC.	*In vivo* lineage bias cannot be inferred from *in vitro* measurement of lineage potential, as biased/restricted HSCs can retain *in vitro*

	lineage output towards particular cellular fates.		potentials that they do not express *in vivo*. *In vivo* lineage bias may be context dependent
Lineage priming	The expression by a multi- or oligo-potent stem- or progenitor cell of genes associated with the terminal differentiation of (a) specific lineage(s).	Single cell gene expression analysis.	Cells may have potentials that are not measurable by gene expression analysis, e.g. lymphoid potential of HSCs.

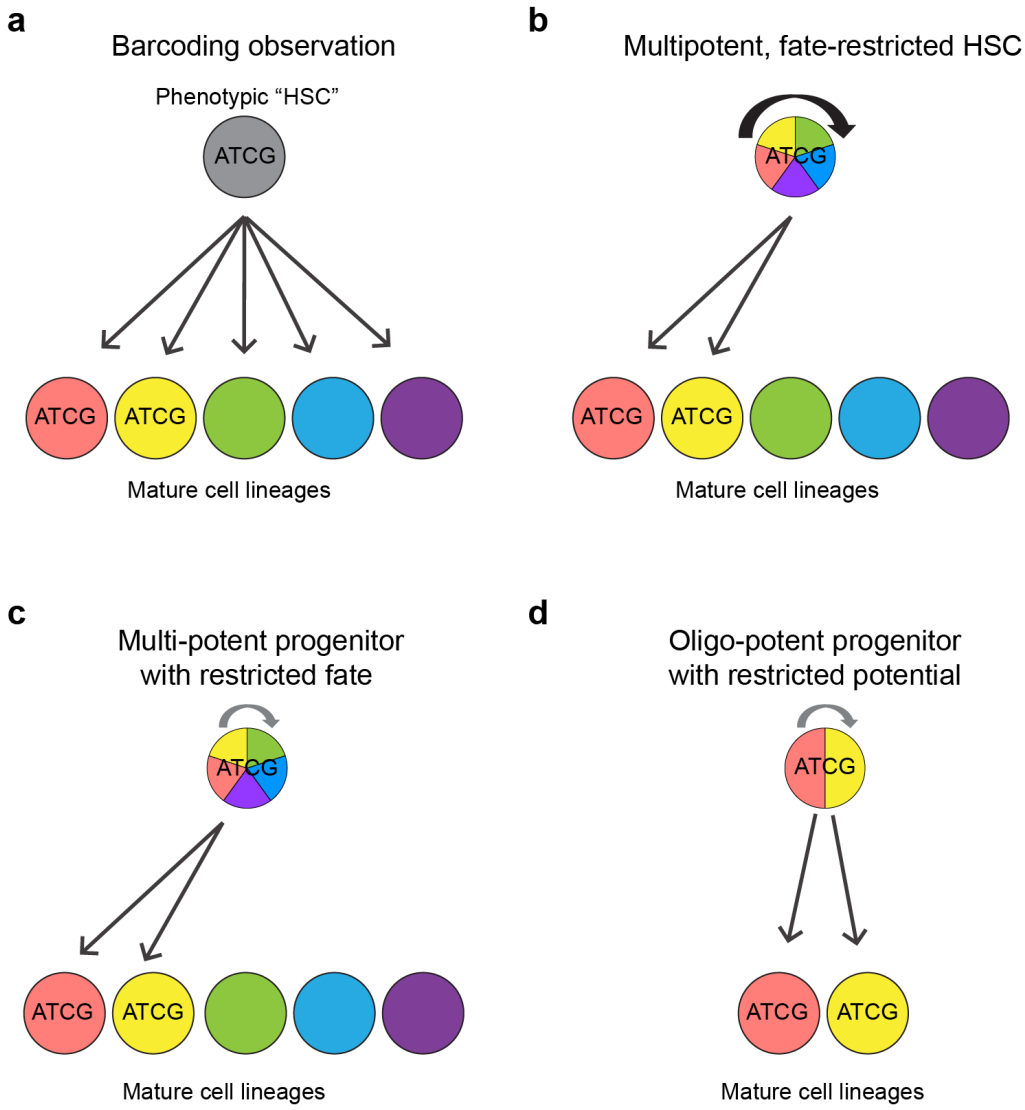

Figure 1: Possible scenarios underlying *in vivo* fate mapping outcomes.

a) An *in vivo* barcoding outcome where clonal replenishment of a restricted set of lineages is observed during steady state hematopoiesis, with the clone originating in a cell with an HSC immuno-phenotype. This outcome could be generated in several distinct scenarios.

b) One option is that labeling occurred in a fate-restricted (i.e. lineage biased) multi-potent repopulating HSC. In this scenario the cell-of-origin would be i) multi-potent when assayed in clonal lineage assays and ii) possess extensive self-renewal potential (large black arrow) and be capable of long-term reconstitution in a serial transplantation assay.

c) A second option is that the labeled clone originated in a multi-potent progenitor with restricted self-renewal potential (small grey arrow) with little or no ability to long-term reconstitute upon transplantation, but revealing multi-potency when assayed in *in vitro* clonal lineage assays.

d) Finally, the cell-of-origin may represent a lineage-restricted progenitor, *in vitro* only possessing the same lineage potentials expressed *in vivo*, and with little or no long-term repopulating activity.

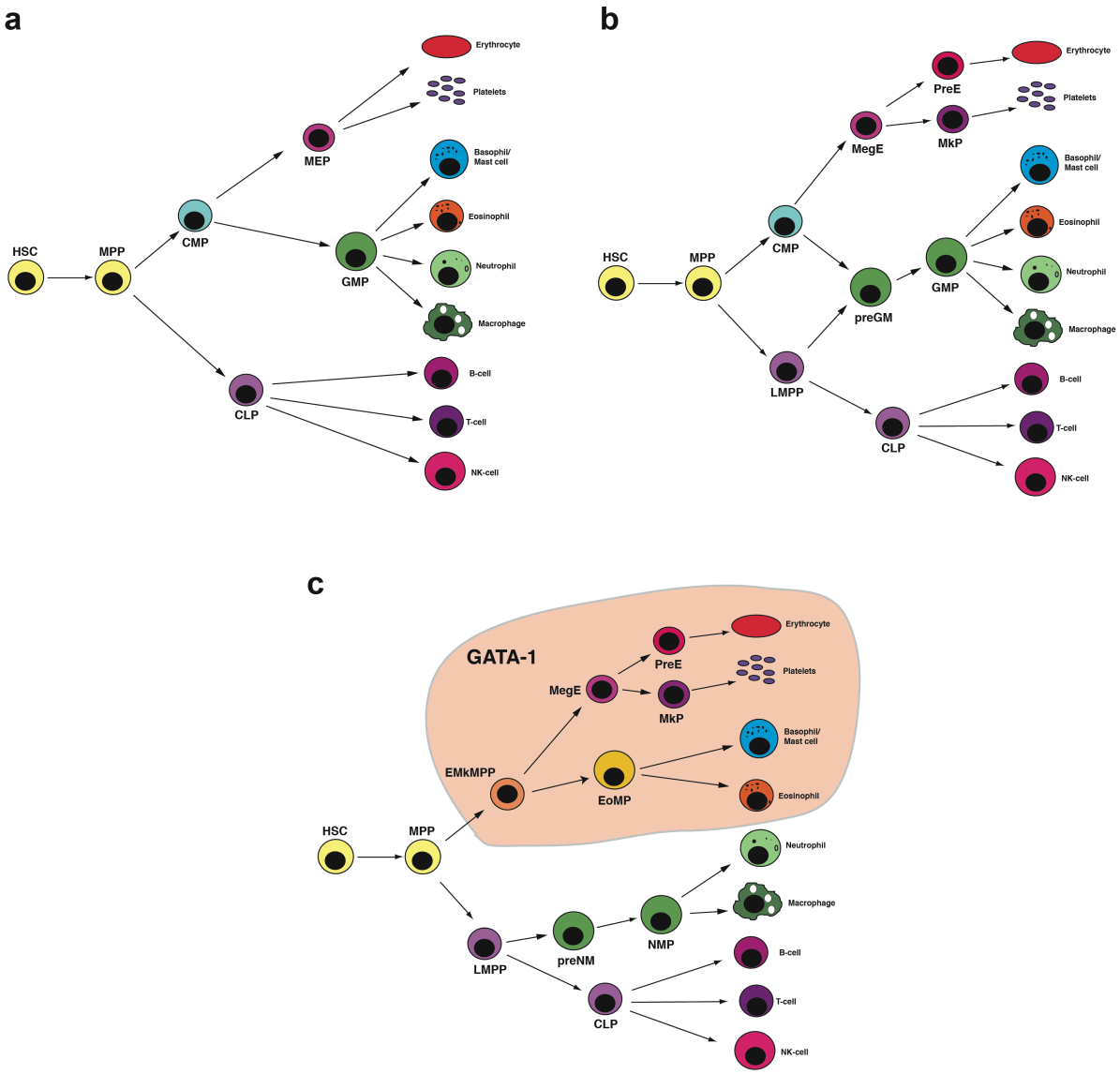

Figure 2: Alternative hierarchical models of hematopoiesis.

a) CMP–CLP model, based on the prospective isolation of cell populations exclusively committed to myeloid/erythroid (CMP) or lymphoid fates (CLP).

b) CMP–LMPP model, incorporating the identification of a myeloid/lymphoid-restricted progenitor population (LMPP), as well as the further fractionation of committed myeloid and megakaryocytic/erythroid progenitors.

c) EMkMPP-LMPP model, where myelopoiesis proceeds via distinct *Gata1*+ (generating eosinophil/mast cells/basophils) and *Gata1*– pathways (generating neutrophils/monocytes). Shaded area indicates the *Gata1* expression domain.