Understanding hematopoiesis from a single-cell standpoint

Konstantinos D. Kokkaliaris[1#], Daniel Lucas[2#], Isabel Beerman[3#], David G. Kent[4*], and Leila Perié[5*]

[1] Cell Systems Dynamics Research Group, ETH Zürich, Department of Biosystems Science and Engineering, Basel, Switzerland
[2] Department of Cell and Developmental Biology, University of Michigan Medical School, MI, USA
[3] Boston Children's Hospital, Boston MA, USA
[4] Department of Haematology, Wellcome Trust and MRC Cambridge Stem Cell Institute, Cambridge University, Cambridge, UK
[5] Institut Curie, PSL Research University, Paris, France

[#] Co-first authors
[*] Co-senior authors

Corresponding author: Konstantinos D. Kokkaliaris,
konstantinos.kokkaliaris@bsse.ethz.ch, Cell Systems Dynamics Research Group, ETH Zürich, Department of Biosystems Science and Engineering, Basel, Switzerland

Abstract

The cellular diversity of the hematopoietic system has been extensively studied and a plethora of cell-surface markers have been used to discriminate and prospectively purify different blood cell types. However, even within phenotypically-identical fractions of hematopoietic stem and progenitor cells (HSPCs) or lineage-restricted progenitors, significant functional heterogeneity is observed when single cells are analyzed. To address these challenges, researchers are now utilizing techniques to follow single cells and their progeny in order to improve our understanding for the underlying functional heterogeneity. On November 19th 2015 Drs. David Kent and Leïla Perié, two emerging young group leaders, presented their recent efforts to dissect the functional properties of individual cells in a webinar series organized by the International Society for Experimental Hematology (ISEH). Here, we provide a summary of the presented

methods for cell labeling and clonal tracking and discuss how these different techniques have been employed to study hematopoiesis.

Introduction

Cellular heterogeneity within defined populations is becoming increasingly evident, and examination of cellular cohorts at the population level may thus obscure unique properties of individual cells. For example, hematopoietic stem and progenitor cells (HSPCs) are defined as the multipotent cells able to give rise to all hematopoietic (myeloid, lymphoid and thrombo-erythroid) lineages. However, there is growing evidence that subpopulations with inherent lineage bias exist. In addition, it has been postulated that committed progenitor populations may also be inherently heterogeneous. Given the heterogeneity of those cellular compartments, single-cell analysis is essential to define their functional potential.

Single-cell sorting has been employed by the stem-cell field to address function of individual cells through either in vivo transplantation or in vitro culture experiments. With advances in sequencing technology, single cells can be assayed for their entire DNA sequence (genome) [1], RNA expression (transcriptome) [2], DNA methylation, chromatin structure (epigenomes) [3], and most recently the combination of both epigenome and transcriptome [4,5]. Evaluation of genome-wide information at the single-cell level provides unique insights into the potential of individual cells, but requires the destruction of the starting cell, and thus functional output cannot be performed in tandem [6–8]. However, several tools have been developed to address this problem. First, flow cytometric index-sorting allows for retrospective analysis by collecting and comparing parameters (light scattering properties, cell-surface marker expression levels) from each of the individual sorted cells from the same experiment. Second, viral barcoding provides a powerful way to assay multiple single cells in the same assay, but is limited by the genetic manipulation of starting cells. In tandem, such powerful methods can provide novel insights into the cellular heterogeneity of

defined hematopoietic cell types. On November 19th 2015 Drs. David Kent and Leïla Perié highlighted techniques employed by their groups to study the functional properties of individual cells in a webinar series organized by the International Society for Experimental Hematology (ISEH) [9,10] and moderated by Dr. Claudia Waskow. Here, we present an overview of this webinar together with advantages and limitations of the main techniques used to identify functional differences between hematopoietic populations; index sorting and viral barcoding (Figure 1).

Linking genome-wide expression data with functional properties in single cells- David Kent

One long-standing challenge in stem cell biology is the identification of distinct molecular markers that would allow isolation of pure, functional HSCs. Over the last decades, a number of laboratories have developed different cell-surface marker combinations or used reporter gene constructs to prospectively isolate HSCs with achieved purities ranging from 20 to 50% [11–15]. While some transplantation failures may be partially attributed to the technical challenges of single-cell transplants, it appears that a sizeable fraction of analyzed cells do not have stem-cell properties. These "contaminating cells" within the isolated HSC population therefore obscure subsequent functional or gene expression analyses. As mentioned above, a variety of functional assays have revealed vast heterogeneity within the HSC pool, since single stem cells show differences on lineage output [16–18], repopulation kinetics [19,20], and response to extrinsic factors [21].

To address these challenges, Dr Kent presented his recent work in the first part of the webinar. In collaboration with Bertie Gottgens' laboratory, Dr. Kent hypothesized that comparing gene expression profiles of HSCs isolated with different strategies would reveal a conserved/overlapping molecular profile between HSCs that would not be shared by various contaminating cell fractions. Excluding contaminating cells based on the expected purity of each HSC population sorted, would thus reveal the molecular signature of "true" stem cells and lead to identification of markers enabling HSC isolation with higher purity.

To test this hypothesis, they combined single-cell gene expression techniques with single-cell in vivo assays and bioinformatic analysis. Initially, they isolated HSCs through four different immunophenotypic strategies (CD34$^-$Flt3$^-$CD48$^-$CD150$^+$KSL, CD45$^+$EPCR$^+$CD48$^-$CD150$^+$, CD34$^-$Flt3$^-$KSL and SP CD150$^+$KSL) as well as five types of progenitor cells. The expression of 43 genes was compared across those 9 cell populations (1800 single cells total) by single-cell qRT-PCR [22]. Using multidimensional mathematical analysis (t-distributed stochastic neighbor embedding analysis, t-SNE), they presented data in single-cell plots confirming that most cells of the same population clustered together. As initially hypothesized, they indeed identified a region where differently sorted HSC populations overlapped, thus sharing a common gene expression profile (termed molecular overlapping HSCs - MolO) compared to those outside that region (no molecular overlap - NoMo).

Taking advantage of flow cytometric index-sorting, a technique which quantifies the intensity for all parameters used for the isolation of single cells, they were able to retrospectively link the cell-surface marker expression of sorted cells with their outcome in downstream assays [23,24]. This permitted Dr Kent and colleagues to associate gene expression data of MolO HSCs with the expression levels of all fluorescent markers used for their isolation. By doing so, they revealed that 28 of 43 genes were differentially expressed between HSCs located in the overlapping or non-overlapping region. Interestingly, MolO HSCs exhibited higher expression for Sca1 and CD150 and lower for CD48 than NoMo. To functionally test these results, CD48$^-$CD150$^+$Sca1$^+$ HSCs were divided to Sca1 high (SLAM Sca1^{hi}) or low expressing cells (SLAM Sca1^{lo}) and their HSC potential was assessed in both in vitro and in vivo assays [22]. Monitoring the cell-cycle profile, colony size and immunophenotype after in vitro culture revealed that SLAM Sca1^{hi} cells were enriched for behaviors typically associated with stem cells (slow division kinetics, small colony size, retaining cell surface marker expression in culture). SLAM Sca1^{hi} HSCs led to higher donor chimerism while producing all hematopoietic lineages compared to the myeloid-deficient SLAM Sca1^{lo} cells in bulk transplantation experiments. In addition, single-cell transplants illustrated that

this 3-marker based strategy (CD48, CD150 and Sca1) yields HSCs with at least 50% purity, comparable to previous schemes utilizing three markers (CD150, CD48, CD41) [12].

To further investigate the underlying HSC heterogeneity at the transcriptome level, single-cell RNA sequencing was performed and identified differential expression of 4533 genes between single $CD34^-Flt3^-CD48^-CD150^+KSL$ cells. Going one step further, Dr Kent presented plots which could link these single-cell RNA sequencing with single-cell transplantation data, since populations used in both assays were index-sorted for the exact same flow cytometry parameters. Bioinformatic analysis of those data identified EPCR as a marker positively correlating with retention of functional stem-cell properties, while negatively correlating with differentiation. Indeed, isolating SLAM Sca1^{hi} EPCR^{hi} cells improved HSC purity to almost 70% as shown by single-cell transplantations [22], providing a novel strategy for functional HSC isolation.

Cellular Barcoding Following Multiple Single Cell Lineages in vivo - Leila Perié

In her seminar Dr. Perié presented the methods utilized by her laboratory to perform barcoding and lineage tracing of hematopoietic progenitors. Her data demonstrate that lymphoid-primed multipotent progenitors and common myeloid progenitor populations are highly heterogeneous containing lineage restricted cells of different commitment potential.

The Perié laboratory utilizes a library of small, non-coding, DNA sequences as barcodes. These are cloned into lentiviral vectors that also express a fluorescent reporter for easy isolation of transduced cells. For lineage tracing cultured hematopoietic progenitors are transduced with the lentiviral library. Those sequences will integrate into the genome of transduced cells allowing identification of their progeny by presence of unique DNA barcodes. The transduced progenitors are then injected into myeloablated recipients and the lineage contribution of each barcode is assessed by purifying specific hematopoietic populations at different

time points after transplantation and performing nested PCR amplification and next-generation sequencing.

During her seminar, Dr Perié emphasized on some important aspects of cellular barcoding. For this method to be successful, validating the ability of utilized viral vectors to transduce the cells of interest while keeping transduction efficiency low (between 5-10%) is essential to ensure a single DNA barcode per progenitor. The size of the library is also important: the number of cells to be transduced should be several orders of magnitude smaller than the diversity of the library (to ensure that each progenitor has a unique barcode). Also, the length of the barcodes will affect sequencing costs. Dr Perié recommended to sequence the full library before any experiment, in order to create a reference library and facilitate bioinformatics analysis of generated data. Another important consideration is controling the number of different progenitors transduced with the same barcode. To do so, Dr Perié suggested transplanting the pool of transduced cells into at least two separate recipients and check whether the same barcode appears in both mice.

It is also important to be aware of the limitations of cellular barcoding. The first limitation is that it provides no information about the exact time of commitment; if a progenitor gives rise to two different cell types it is impossible to determine whether this occur early or late during cell maturation. It also provides no information whether the transduced progenitor underwent trans-differentiation or de-differentiation instead of commitment to one or more lineages. An important technical limitation is that the technique requires in vitro culture, use of lentivirus and lengthy transplantation of the transduced progenitors into myeloablated recipients. All those steps can affect lineage commitment decisions and not reflect actual lineage differentiation during homeostasis.

Utilizing the methods described above, Dr. Perié presented data showing that early murine hematopoietic progenitors are highly heterogeneous and, in most cases, already committed to specific lineages. Transplanting lentivirally barcoded $CD16/32^-CD127^-CD117^{hi}Sca1^+CD135^{hi}$ lymphoid-primed multipotent progenitors (LMPP) into myeloablated recipients revealed that LMPPs are extremely

heterogeneous in their lineage output. Most LMPP were already committed to dendritic, myeloid or B-cell lineage and only a small fraction was capable of multilineage reconstitution [25]. There results demonstrated that LMPP could generate dendritic cells directly without passing through a common myeloid progenitor (CMP) or common lymphoid progenitor (CLP) stage [25]. These, together with mathematical modeling, suggested that the classical hematopoietic differentiation tree should be revised to include new LMPP sub-types [26]. Using the same technique, Dr. Perié examined lineage commitment to erythroid and myeloid lineages from the CMP stage. They found that the CMP population was also heterogeneous with most CMP already committed to either myeloid or erythroid lineages, whereas only 5% of CMP are bipotent [27]. These studies revealed the heterogeneity of hematopoietic progenitors defined by cell-surface markers and highlighted the power of cellular barcoding to investigate progenitors' potential with single-cell resolution.

Together, these single-cell methodologies and recent studies utilizing these powerful methods have provided insights into the heterogeneity of the primitive hematopoietic compartment including committed progenitor cells and not just the early stem cell compartment, rigorously defined a cell-surface marker combination for HSCs. Also, the power of single-cell studies was highlighted, which will soon become the norm for evaluating cell function and potential, as further optimization and enhancements of current methods to analyze genome-wide information generated from single cells is ongoing.

The webinar can be viewed at the ISEH website at: http://iseh.site-ym.com/?ISEHWebinars

References

1.   Gawad C, Koh W, Quake SR. Single-cell genome sequencing: current state of the science. *Nat Rev Genet.* 2016;
2.   Wu AR, Neff NF, Kalisky T, et al. Quantitative assessment of single-cell RNA-sequencing methods. *Nat Methods.* 2014;11(1):41–46.

3. Schwartzman O, Tanay A. Single-cell epigenomics: techniques and emerging applications. *Nat Rev Genet.* 2015;16(12):716–726.

4. Angermueller C, Clark SJ, Lee HJ, et al. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat. Methods.* 2016;

5. Cabezas-Wallscheid N, Klimmeck D, Hansson J, et al. Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. *Cell Stem Cell.* 2014;15(4):507–522.

6. Schroeder T. Long-term single-cell imaging of mammalian stem cells. *Nat Methods.* 2011;8(4):30–35.

7. Kokkaliaris KD, Loeffler D, Schroeder T. Advances in tracking hematopoiesis at the single-cell level. *Curr Opin Hematol.* 2012;19(4):243–9.

8. Etzrodt M, Endele M, Schroeder T. Quantitative single-cell approaches to stem cell research. *Cell Stem Cell.* 2014;15(5):546–558.

9. Singbrant S, van Galen P, Lucas D, et al. Two new routes to make blood: Hematopoietic specification from pluripotent cell lines versus reprogramming of somatic cells. *Exp Hematol.* 2015;43(9):756–759.

10. Sykes SM, Kokkaliaris KD, Milsom MD, Levine RL, Majeti R. Clonal evolution of pre-leukemic hematopoietic stem cells in acute myeloid leukemia. *Exp Hematol.* 2015;43(12):989–992.

11. Osawa M, Hanada K, Hamada H, Nakauchi H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science.* 1996;273(5272):242–5.

12. Kiel MJ, Yilmaz OH, Iwashita T, et al. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell.* 2005;121(7):1109–21.

13. Kent DG, Copley MR, Benz C, et al. Prospective isolation and molecular characterization of hematopoietic stem cells with durable self-renewal potential. *Blood.* 2009;113(25):6342–50.

14. Gazit R, Mandal PK, Ebina W, et al. Fgd5 identifies hematopoietic stem cells in the murine bone marrow. *J. Exp. Med.* 2014;211(7):1315–31.

15. Acar M, Kocherlakota KS, Murphy MM, et al. Deep imaging of bone marrow shows non-dividing stem cells are mainly perisinusoidal. *Nature.* 2015;526(7571):126–30.

16. Muller-Sieburg CE, Cho RH, Karlsson L, Huang J-F, Sieburg HB. Myeloid-biased hematopoietic stem cells have extensive self-renewal capacity but generate diminished lymphoid progeny with impaired IL-7 responsiveness. *Blood.* 2004;103(11):4111–8.

17. Dykstra B, Kent D, Bowie M, et al. Long-term propagation of distinct hematopoietic differentiation programs in vivo. *Cell Stem Cell.* 2007;1(2):218–29.

18. Beerman I, Bhattacharya D, Zandi S, et al. Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. *Proc Natl Acad Sci U S A.* 2010;107(12):5465–70.

19. Sieburg HB, Cho RH, Dykstra B, et al. The hematopoietic stem compartment consists of a limited number of discrete stem cell subsets. *Blood*. 2006;107(6):2311–6.

20. Morita Y, Ema H, Nakauchi H. Heterogeneity and hierarchy within the most primitive hematopoietic stem cell compartment. *J Exp Med*. 2010;207(6):1173–1182.

21. Challen G a, Boles NC, Chambers SM, Goodell M a. Distinct hematopoietic stem cell subtypes are differentially regulated by TGF-beta1. *Cell Stem Cell*. 2010;6(3):265–78.

22. Wilson NK, Kent DG, Buettner F, et al. Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. *Cell Stem Cell*. 2015;16(6):712–724.

23. Osborne GW. Recent advances in flow cytometric cell sorting. *Methods Cell Biol*. 2011;102:533–556.

24. Schulte R, Wilson NK, Prick JCM, et al. Index sorting resolves heterogeneous murine hematopoietic stem cell populations. *Exp Hematol*. 2015;43(9):803–811.

25. Naik SH, Perie L, Swart E, et al. Diverse and heritable lineage imprinting of early haematopoietic progenitors. *Nature*. 2013;496(7444):229–232.

26. Perié L, Hodgkin PD, Naik SH, et al. Determining Lineage Pathways from Cellular Barcoding Experiments. *Cell Rep*. 2014;6(4):617–624.

27. Perié L, Duffy KR, Kok L, et al. The Branching Point in Erythro-Myeloid Differentiation. *Cell*. 2015;163(7):1655–1662.

Figure legend

Figure 1: Single-cell methods to define properties of individual cells that are masked in population-based experimental paradigms. Index sorting allows for the retrospective analysis of FACs data post experiment (ie after RNA-Sequensing, single-cell transplant, clonal culture assays). Lentiviral barcoding allows for tagging a plethora of single cells (after purification or enrichment of a population) that can then be used to track individual cells' potential. There are benefits and drawbacks to each method, but both have been utilized to establish more in-depth appreciation of the heterogeneity in primitive hematopoietic cell potential.

FACS Sort

Assay
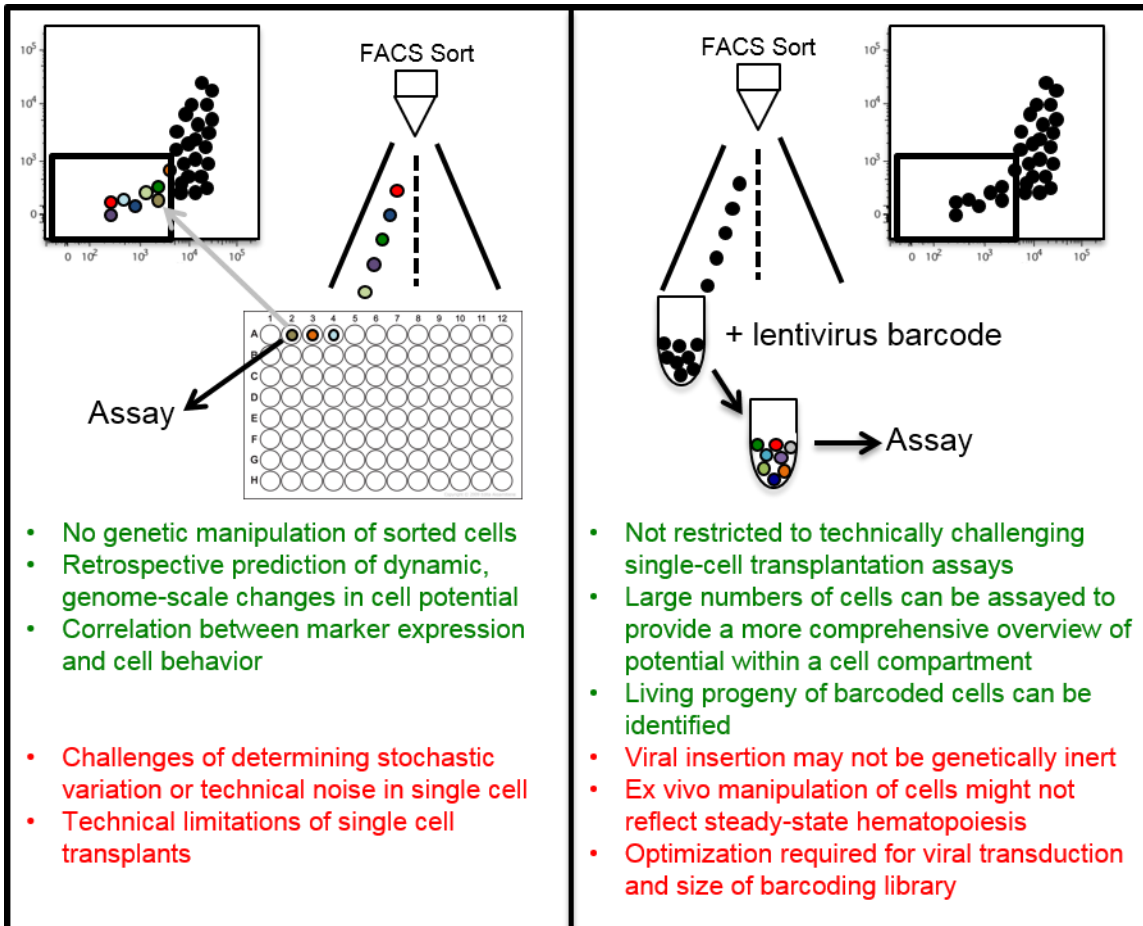
- No genetic manipulation of sorted cells
- Retrospective prediction of dynamic, genome-scale changes in cell potential
- Correlation between marker expression and cell behavior

- Challenges of determining stochastic variation or technical noise in single cell
- Technical limitations of single cell transplants

FACS Sort

+ lentivirus barcode

Assay

- Not restricted to technically challenging single-cell transplantation assays
- Large numbers of cells can be assayed to provide a more comprehensive overview of potential within a cell compartment
- Living progeny of barcoded cells can be identified
- Viral insertion may not be genetically inert
- Ex vivo manipulation of cells might not reflect steady-state hematopoiesis
- Optimization required for viral transduction and size of barcoding library