Dissecting Stem Cell Differentiation using Single Cell Expression Profiling

Victoria Moignard[1] and Berthold Göttgens[1]*

Department of Haematology and Wellcome Trust - Medical Research Council Cambridge Stem Cell Institute, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK.

*Corresponding author: bg200@cam.ac.uk

Abstract

Many assumptions about the way cells behave are based on analyses of populations. However, it is now widely recognized that even apparently pure populations can display a remarkable level of heterogeneity. This is particularly true in stem cell biology where it hinders our understanding of normal development and the development of strategies for regenerative medicine. Over the past decade technologies facilitating gene expression analysis at the single cell level have become widespread, providing access to rare cell populations and insights into population structure and function. Here we review the contributions of single cell biology to understanding stem cell differentiation so far, both as a new methodology for defining cell types and a tool for understanding the complexities of cellular decision-making.

Introduction

Embryonic development and multilineage differentiation require that diversity be generated from individual cells, whether the zygote or adult stem cells. While the cell populations produced by these processes show stereotypical behaviours with regards to stability and potential that are vital to normal development and homeostasis, there is now recognized to be huge variation in populations at the cellular level [1,2]. For example, embryonic stem cells (ESCs) are heterogeneous and prone to differentiation in conventional serum and Lif culture conditions, but are transcriptionally and phenotypically more homogeneous in the 'ground state' 2i and Lif conditions and stably self renew [3,4]. With the current drive to understand and mimic cell fate decisions in culture for regenerative medicine, it is vital to understand how diversity arises, what causal role or effect heterogeneity has in differentiation and whether it can be modulated to produce phenotypically pure populations.

Heterogeneity can be due to the presence of multiple cell subpopulations, asynchrony in cell cycle progression [5], or stochasticity in molecular processes including transcriptional bursting [6]. This has great implications for the extrapolation of population studies to individual cells as changes in population gene expression can correspond to changes in individuals or to a change in the cellular composition of a tissue (Figure 1). Therefore strategies to analyse individual cells have great promise in increasing our understanding of stem cell biology. Furthermore, stem cells are often rare populations and not amenable to conventional studies requiring millions of cells. Single cell studies therefore facilitate molecular analyses of previously intractable cells.

Historically, attempts to study single cells have been based on imaging or flow cytometry, limiting the number of parameters that can be investigated. Two key developments over the last decade have opened up the era of single cell biology, most notably in the case of transcriptomics: the introduction of many –omics technologies and their reduction to the single cell level [7], and the use of microfluidics to miniaturise and parallelise procedures [8–10]. It is now possible to assay the entire transcriptome of individual cells, and although there are still technical challenges - the low efficiency (5-25%) of reverse transcription means that lowly expressed genes may not be captured [11] - it is usually possible to obtain biologically meaningful information for several thousand genes per cell depending on cell type and sequencing depth [11–13].

Here we will explore the lessons learned from single cell transcriptomics regarding the nature of cellular decision-making and the function of heterogeneity, and how single cell transcriptomics is redefining lineages. We will not cover the technical aspects of experimental design and analysis, which are reviewed elsewhere [13–18].

Towards an atlas of cell fate

Homing in on HSCs

Since the discovery of the haematopoietic stem cell [19], decades of work have gone into identifying this rare population and the hierarchy through which it produces the diverse mature cell types of the blood system [20,21]. Key to this process has been prospective isolation of cell types using antibody staining and FACS, and their characterization with functional assays. Not surprisingly given the promiscuity of many surface markers, most populations remain impure with at best 50% of cells immunophenotypically defined as HSCs exhibiting true HSC activity in transplantation assays [22,23].

Trying to distinguish between transcriptional noise and HSC subpopulations, Glotzbach et al., [24] identified nine genes that differed between $CD34^{lo}$ cells (enriched for HSC activity) and $CD34^{high}$ progenitor cells. They identified a cluster of $CD34^{lo}$ cells that was also found in HSCs sorted using another strategy, but was underrepresented in $CD34^{high}$ cells, potentially representing part of the transcriptional program of true HSCs. However, analyzing so few genes provides little insight into the regulation of HSCs, and this study offered no means for isolating these cells.

To enrich for HSCs, Wilson et al., [12] collected single cells using four common immunophenotypes for mouse HSCs (Figure 2A). They identified the molecular overlap (termed 'MolO' cells) between the four populations using 48 genes analysed by single cell quantitative real-time PCR (sc-qRT-PCR), specifically identifying a region in the 48-dimensional space where the frequency of cells of each sorting strategy matched the percentage that read out as HSCs in transplantation assays. The authors used index sorting – where the fluorescence data relating to each cell is retained – to associate the MolO gene expression profile with a refined $CD150^{+}CD48^{-}Sca1^{high}EPCR^{high}$ surface profile that further enriched HSCs to 67% of the sorted population. Additional whole transcriptome analysis using single cell RNA sequencing (scRNA-seq) identified a cell cycle signature consistent with quiescent HSCs [12]. While clearly of specific interest to the haematopoietic community, this

strategy can be applied to other systems to identify potential new stem cell markers for functional testing.

Re-routing myelopoiesis

The origins of each haematopoietic lineage are also a source of continued debate. The traditional hierarchical model of binary fate choices between alternative lineages (Figure 1C) [25,26] has been called into question with reports that the common myeloid progenitor (CMP) is not a bipotent progenitor but a heterogeneous mixture of different lineage progenitors [27], and that the megakaryocyte-erythroid lineage differentiates directly from the HSC rather than through the CMP [28].

A large-scale single cell qRT-PCR study of multiple adult haematopoietic populations showed that the CMP could be subdivided by expression of *CD55* [29]. Prospective isolation and transplantation of $CD55^{+}$ and $CD55^{-}$ populations identified megakaryocyte-erythroid and myeloid biases in lineage output respectively, in line with the transcriptomic data. Further upstream, high levels of CD150 in HSCs, already shown to enrich for long-term HSC activity [23], correlated with a megakaryocyte-erythroid gene signature [29], in line with the early emergence of platelet-primed progenitors [28].

scRNA-seq has been used to dissect myelopoiesis further. Massively parallel single cell RNA-seq (MARS-seq) was developed to sequence tens of thousands of cells at low coverage, with only 200-1,500 mRNA molecules detected per cell, but provided enough information to define dendritic cell types [30]. The same method has since been applied to the whole haematopoietic stem/progenitor compartment [31]. Expression of CD34 and FcgR – used to separate the CMP from its supposed progeny, the megakaryocyte-erythroid (MEP) and granulocyte-monocyte (GMP) progenitors – was also recorded by index sorting. Analysis of nearly 3,000 cells identified 19 myeloid clusters which could be associated with particular lineages and differentiation stages based on the expression of key markers, but which did not cleanly segregate into the CMP, MEP and GMP gates conventionally used for sorting (Figure 2B).

The myeloid compartment has already been further subdivided by flow cytometry [32], so it would be interesting to see how the clusters identified by Paul et al., fit within this more refined view of myelopoiesis. Nevertheless, the hierarchy proposed between clusters was consistent with early lineage commitment rather than the

existence of multipotent progenitors (Figure 2C), which agrees with recent barcoding-based lineage tracing experiments in native haematopoiesis [33,34].

Lineage decomposition in solid tissues

Despite these ongoing debates, haematopoiesis remains the best-characterized stem cell system. Analysis of lineages and cellular potential has been harder in non-haematopoietic tissues where cells are more difficult to access and assays less well developed. Here, the value of scRNA-seq for non-biased lineage decomposition and marker identification is clear.

A similar approach to myelopoiesis was taken in the intestine, where single cells from intestinal organoids – in vitro 3D cultures originating from single stem/progenitor cells that recapitulate normal intestinal structures – were sequenced to investigate the cellular composition of the tissue [35]. Although the cell numbers were relatively small compared to the haematopoietic studies, the authors developed a new computational tool, RaceID, to identify rare cell types, even where they make up as few as 1 cell in the population, by identifying cells that express 'outlier' genes above levels expected based on population noise. Using this tool, *Reg4* was identified and validated as a new marker to enrich for enteroendocrine cells, and sequencing of $Reg4^+$ cells identified several cellular subtypes not previously known to exist in the small intestine, with implications for understanding the endocrine control of digestion [35]. The authors also used RaceID on sequenced $Lgr5^+$ cells from organoids and primary mouse samples to contribute to the ongoing debate as to whether the $Lgr5^+$ stem cell pool is heterogeneous, but could not detect subpopulations [35]. Given the current noisy nature of scRNA-seq, RaceID is unlikely to be accurate when faced with lowly-expressed genes which are more prone to dropout. Indeed, the screen of whole organoids failed to identify a stem cell cluster, likely because of the low expression of *Lgr5*, but this should improve with sequencing quality.

It is often impossible to know a priori how many cells will be required to identify particular populations, so being able to capture all cells of a tissue rather than a subset would be beneficial. However, with conventional methods this either requires a sacrifice in sequencing depth for each cell to enable more to be analysed, or makes experiments prohibitively expensive for many researchers. Similar to MARS-seq, another study sought to investigate the level of sequencing depth required to accurately classify cells [36]. Down-sampling indicated that 50,000 sequencing reads per cell are required to distinguish disparate cell types including blood, neural and

epidermal cells [36], compared with the 20,000 indicated by MARS-seq [30]. The authors went on to examine primary neural cells, as understanding development and neurodevelopmental disorders is hindered by the variety of cell types present in the developing brain. Down-sampling to as few as 5,000 reads per cell was still sufficient to coarsely cluster cells, although using a greater fraction of the data allowed for the identification of many putative markers and indicated that cells could be classified on the basis of cell cycle and the activity of signaling pathways, as well as by cell type [36].

Two recent technologies, Drop-seq [37] and inDrop [38], provide increased throughput at a lower cost by capturing thousands of individual cells in parallel in nanolitre-volume droplets. Each droplet functions as a microscopic reaction chamber for library preparation, with barcoded cell libraries later highly multiplexed for sequencing. Drop-seq was used to interrogate nearly 45,000 mouse retinal cells, identifying 39 transcriptionally-distinct clusters down to 50 cells in size, including known and new populations [37].

scRNA-seq has recently been used to identify lineages and lineage relationships in several other tissues, including the lung [39], otocyst [40] and during cardiogenesis [41]. A lot may therefore be learned by stepping away from conventional surface marker-based assays and transplantation models towards single cell profiling of whole native tissues, even where we consider a lineage to be well-defined. Droplet-based technologies will greatly facilitate this process [37,38] and commercial platforms are beginning to emerge. However, prospective isolation will still be required to demonstrate the functionality of each cell type [12,31]. It will also be important to move beyond providing an atlas of cell types to understanding how the differences between them arise. Comparing the results of mutations or disease models to wild type cells is already providing insights into the roles of individual genes [31,42] and comparing populations such as induced HSCs [29,43] and ESCs grown in different conditions [44] to their in vivo counterparts has helped improve experimental strategies.

Lessons in lineage segregation

While morphological and immunophenotypic differences allow us to isolate populations, the events that segregate them may happen many generations previously without any morphological indications, making it difficult to define and capture decision points. One of the earliest single cell transcriptomic studies

analysed individual cells from the mouse zygote through to the 64 cell blastocyst by sc-qRT-PCR [45]. This identified inverse correlations between *Sox2* and *Id2*, and *Fgf4* and *Fgfr2*, indicative of the lineage decisions between trophectoderm and inner cell mass, and epiblast and primitive endoderm (PE), respectively, earlier than they can be morphologically distinguished. These findings are consistent with the blockade of Fgf signaling employed by 2i and Lif ESC culture conditions [3,46], and blocking Fgf signaling in morulae resulted in a down-regulation of PE and up-regulation of epiblast markers [45]. A subsequent study of *Fgf4*$^{-/-}$ embryos indicated that the Fgf pathway functions to reinforce expression patterns resulting from earlier heterogeneity, leading to lineage segregation [47], although the cause of such heterogeneity and the early differences identified by Guo et al., [45] is unresolved. Although mammalian embryos are remarkably plastic and can survive the removal of cells at a very early stage, for example for preimplantation genetic diagnosis, several recent studies have reported differences in gene expression of sister blastomeres in 2- and 4-cell stage mouse embryos that can be related to lineage choice rather than noise [48–50]. A bias in the contribution of blastomeres at the 4 cell stage to different lineages has also been observed by lineage tracing [51], so understanding how these early differences arise will be a key issue.

Recent studies using scRNA-seq have indicated that there is reasonable conservation between gene modules expressed in very early mouse and human embryos (1-8 cells) [52,53]. There are some differences in timing and specificity, including for major pathways such as Tgfβ, which was shown to play a key role in maintaining the pluripotent epiblast in humans only [54]. Studies have also highlighted significant differences between hESCs and human epiblast cells [52,54], although this is reduced in more recent attempts to achieve ‘ground state’ pluripotency in hESCs [54]. These studies suggest that the differences found could help to identify pathways, particularly Wnt and Fgf, that could be modulated to obtained cell lines more representative of the in vivo pluripotent state.

Hierarchical and stochastic stages of commitment and reprogramming

One of the earliest attempts at single cell gene expression analysis suggested that HSCs promiscuously express lineage-affiliated genes, termed ‘lineage priming’, prior to differentiation [55]. A number of studies have subsequently investigated lineage commitment and suggest that the early stages are stochastic, with the heterogeneous expression of lineage-affiliated genes eventually swaying the balance

from self-renewal to differentiation before a lineage programme becomes irreversibly activated.

In the haematopoietic EML cell line, levels of the surface marker Sca-1 were shown to correlate with lineage potential, and cells could reversibly move between subpopulations [56]. However, the potential of individual cells across the distribution of Sca-1 expression was never formally tested. In a follow up study using sc-qRT-PCR [57], the erythroid-biased Sca-1^{lo} population was further divided on the basis of expression of key erythroid regulator Gata1. This correlated with differentiation capacity but not self renewal [57], arguing that the two programs are separate and that self-renewing cells cannot significantly sample lineage programs without committing. However, the newly committed Gata1^{+} cells were transcriptionally more similar to the uncommitted progenitors than mature erythroid cells, with substantially more heterogeneity in expression. Coupled with further computational modeling, this work suggests that while cells ultimately activate the same lineage-specific transcriptional programme, the early stages are stochastic and offer multiple routes into differentiation [57,58].

Heterogeneity of Nanog expression in ESCs [59], among other factors [60,61], has also been linked to differentiation bias, while ESCs can be maintained without Nanog but are more prone to differentiation [59]. Following transient *Nanog* depletion, the pluripotency network is stable enough to be rescued by *Nanog* re-expression for 3 days, before irreversibly breaking down during differentiation [62]. Sc-qRT-PCR analysis of a number of genes indicated that the early changes after *Nanog* depletion are stochastic, with no subpopulations of cells identified. As Nanog is involved in many feedback loops in ESCs, the authors proposed that Nanog fluctuations cause transitions between a feedback-rich pluripotent state, and states with less feedback that are prone to differentiation [62]. The use of InDrops to study ESC heterogeneity and the first stages of differentiation additionally suggested that fluctuations in the expression of pluripotency regulators are weakly coupled within cells, but a strong differentiation stimulus such as loss of LIF results in a more coherent lineage programme [38], as in erythroid cells [57]. This early-stochastic and late-hierarchical pattern also holds true in reprogramming, where single cell analysis indicated that a coherent transcriptional programme only develops after activation of *Sox2*, and that the early stochasticity accounts for low reprogramming efficiency [63].

Several studies have attempted to formalize gene expression changes between cells to generate gene regulatory networks that explain self-renewal and differentiation. Network inference from population studies has been hindered by having few samples relative to the number of genes studied, and due to the asynchrony of cells within those samples. Methods are now improving thanks to the thousands of cells that can be analysed using single cell methods. Using correlation between genes in 600 cells across multiple haematopoietic stem/progenitor populations, we identified a triad of transcription factors (TFs), including Gata2, that seems to act in regulating exit from the stem cell compartment, with the connectivity validated by ChIP-seq and transcriptional assays [64]. Other single cell studies have also highlighted the importance of Gata2 in early stages of HSC differentiation [29,57,58]. Using a more sophisticated synthesis approach we built a boolean network for early haematopoietic development in the embryo that provided a number of hypotheses about gene regulation and allowed us to test the function of each gene in the network [65].

Recreating developmental trajectories using pseudotemporal ordering

A great drawback of current scRNA-seq technologies is the loss of spatiotemporal information associated with cells, an important consideration given the highly ordered structure of tissues and the information passed between cells both through physical contact and paracrine signaling. Several unsupervised approaches have recently been suggested for reconstructing cellular ‘pseudotime’ by ordering cells according to similarities in their transcriptomes and finding the longest continuous path through the data, with no prior knowledge required of expected gene expression patterns.

Monocle was introduced for the analysis of skeletal myoblasts in culture, where it identified clusters of genes with different kinetics [66], and has subsequently been used to study olfactory receptor development [16] and to compare neuronal development in human neocortex and cerebral organoids [67]. Wanderlust was developed for single cell mass cytometry analysis of protein expression in differentiating B cells [68], but is also applicable to scRNA-seq. We used diffusion maps to order 4,000 cells from the earliest stages of blood development in the gastrulating mouse embryo and were able to capture the bifurcation of blood and endothelium from mesoderm [65,69]. Where real-time information is lacking, the pseudotemporal ordering of otherwise asynchronous snapshots of cells therefore allows for greater resolution in understanding the order of gene expression and therefore the regulatory relationships between genes.

Conclusions

The era of single cell biology is upon us, requiring new methods and interdisciplinary collaborations, and a new way of thinking about biological problems [70,71]. In particular, the use of single cell technology as a new method to explore lineage structure seems set to continue with the potential to catalogue the entire human body. The continued decrease in sequencing cost and increase in multiplexing, particularly with droplet-based technologies, will also make the technology available to a wider circle of researchers and topics.

While the promise is great, there are many challenges still to be faced, not least in improving RNA capture and processing and distinguishing biological variation from technical noise. Methods such as SmartSeq2 span whole transcripts and can therefore be used to discover novel isoforms and splicing, but provide relative transcript abundances [72]. Conversely, MARS-seq [30] and droplet technologies [37,38] incorporate unique molecular identifiers into each transcript to enable transcript counting and reduce technical noise, but as a result only provide information about the ends of transcripts while the parallelization prevents phenotypic information such as surface marker expression from being recorded for each cell. Further, the lower depth of sequencing typical with these techniques sacrifices information about lowly expressed genes, which can include key regulatory factors such as TFs. There is necessarily a trade off between the number of cells that can be analysed and the sequencing depth per cell, and researchers must carefully evaluate their needs when designing experiments to select the most appropriate sequencing method for the information required [14,13].

A deeper biological understanding will also require analysis of other aspects of gene regulation and function. Accordingly, methods are emerging to study other molecules and modifications at single cell resolution, from proteins [73] and DNA methylation [74], to protein-DNA interactions [75,76] and chromosome looping [77]. It's also possible to study DNA and RNA [78] or methylation and transcription [79] within the same cell, so single cell systems biology is on the horizon. Spatial context cannot be ignored, and methods are appearing for sequencing in situ [80] and for building tissue maps from in situ hybridization images [81,82] and by sequencing small populations of cells from multiple tissue sites [83], against which single cell transcriptomes can be mapped. Likewise, live imaging provides a temporal context completely unavailable to transcriptomics technologies [84], but is limited in the

number of genes or proteins that can be studied simultaneously. Mapping transcriptomic data against live imaging information could, like arrangement in pseudotime, help to bring order to snapshot data to discover the underlying patterns. These ideas are methodologically challenging, both for experimentalists and computational biologists. A strong, continued relationship between the two is therefore fundamental for the continued success of single cell biology.

Acknowledgements

Research in the authors' laboratories is supported by the Medical Research Council, Biotechnology and Biological Sciences Research Council, Bloodwise, the Leukemia and Lymphoma Society, a Wellcome Trust Strategic Award (Tracing Early Mammalian Lineage Decisions by Single Cell Genomics) and core support grants by the Wellcome Trust to the Cambridge Institute for Medical Research and Wellcome Trust - MRC Cambridge Stem Cell Institute.

Figure legends:

Figure 1: More than the sum of its parts. A) While a population-level approach may indicate that all cells express the same level of three genes (left), single cell analysis shows a wide range of expression patterns. This has great implications for understanding behavior and regulatory interactions between genes. Modified from [85]. **B)** A change in expression of a gene in response to a stimulus can occur due to a change in expression within individual cells (due to a change in regulation; top right) or due to a change in the composition of a population (bottom right), for example with the selective proliferation of one cell subset at the expense of another. Modified from [86].

Figure 2: Redefining haematopoiesis. A) Wilson et al., [12] performed qRT-PCR for a set of known stem cell regulators in single cells of four HSC populations prospectively isolated based on published immunophenotypes. A region was identified in which all four populations overlapped and was associated with a particular immunophenotype (red box on FACS plot) within the HSC gate (blue box on FACS plot). Subsequent transplantation of these 'MolO' cells into mice indicated that they are enriched for HSC activity compared with previous protocols. **B)** Paul et al., [31] used MARS-seq to profile the transcriptomes of ~2,700 cells of the mouse progenitor compartment. They identified 19 cell clusters showing varying degrees of overlap with the MEP, GMP and CMP compartments previously defined. Each cluster was associated with a particular lineage based on the expression of key markers and TFs. Heatmap from [31]. **C)** The conventional haematopoietic hierarchy (left) in which HSCs give rise to mature cell types through a hierarchy of multipotent progenitors. A novel hierarchy (right) has been proposed as a result of work by Paul et al., [31] and others in which only the earliest cells are multipotent, with lineage restriction occurring earlier during differentiation.

References:

1. Graf T, Stadtfeld M: **Heterogeneity of embryonic and adult stem cells. [Internet]**. *Cell Stem Cell* 2008, **3**:480–3.
2. Copley MR, Beer PA, Eaves CJ: **Hematopoietic stem cell heterogeneity takes center stage. [Internet]**. *Cell Stem Cell* 2012, **10**:690–7.
3. Ying Q-L, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, Cohen P, Smith A: **The ground state of embryonic stem cell self-renewal. [Internet]**. *Nature* 2008, **453**:519–23.
4. Hackett JA, Surani MA: **Regulatory principles of pluripotency: from the ground state up. [Internet]**. *Cell Stem Cell* 2014, **15**:416–30.
5. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ, Teichmann SA, Marioni JC, Stegle O: **Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells [Internet]**. *Nat. Biotechnol.* 2015, **33**:155–60.
6. Raj A, van Oudenaarden A: **Nature, nurture, or chance: stochastic gene expression and its consequences. [Internet]**. *Cell* 2008, **135**:216–26.
7. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, et al.: **mRNA-Seq whole-transcriptome analysis of a single cell. [Internet]**. *Nat. Methods* 2009, **6**:377–82.
8. Lecault V, White AK, Singhal A, Hansen CL: **Microfluidic single cell analysis: from promise to practice. [Internet]**. *Curr. Opin. Chem. Biol.* 2012, **16**:381–90.
9. White AK, VanInsberghe M, Petriv OI, Hamidi M, Sikorski D, Marra MA, Piret J, Aparicio S, Hansen CL: **High-throughput microfluidic single-cell RT-qPCR. [Internet]**. *Proc. Natl. Acad. Sci. U. S. A.* 2011, **108**:13999–4004.
10. Citri A, Pang ZP, Südhof TC, Wernig M, Malenka RC: **Comprehensive qPCR profiling of gene expression in single neuronal cells. [Internet]**. *Nat. Protoc.* 2012, **7**:118–27.
11. Islam S, Kjällquist U, Moliner A, Zajac P, Fan J-B, Lönnerberg P, Linnarsson S: **Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. [Internet]**. *Nat. Protoc.* 2012, **7**:813–28.
12. Wilson NK, Kent DG, Buettner F, Shehata M, Macaulay IC, Calero-Nieto FJ, Sánchez Castillo M, Oedekoven CA, Diamanti E, Schulte R, et al.: **Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations [Internet]**. *Cell Stem Cell* 2015, **16**:712–24.

** This study analysed single cells sorted with 4 commonly used mouse HSC

immunophenotypes in order to identify true HSCs within the population. In combination with index sorting and transplantation, they identified a more refined surface profile enriching cells to 67% of the sorted population, compared with 50% in previous protocols.

13. Grün D, van Oudenaarden A: **Design and Analysis of Single-Cell Sequencing Experiments [Internet]**. *Cell* 2015, **163**:799–810.
14. Stegle O, Teichmann SA, Marioni JC: **Computational and analytical challenges in single-cell transcriptomics [Internet]**. *Nat. Rev. Genet.* 2015, **16**:133–45.
15. Woodhouse S, Moignard V, Göttgens B, Fisher J: **Processing, visualising and reconstructing network models from single-cell data. [Internet]**. *Immunol. Cell Biol.* 2015, doi:10.1038/icb.2015.102.
16. Hanchate NK, Kondoh K, Lu Z, Kuang D, Ye X, Qiu X, Pachter L, Trapnell C, Buck LB: **Single-cell transcriptomics reveals receptor transformations during olfactory neurogenesis [Internet]**. *Science (80-.).* 2015, doi:10.1126/science.aad2456.
17. Semrau S, van Oudenaarden A: **Studying Lineage Decision-Making In Vitro: Emerging Concepts and Novel Tools. [Internet]**. *Annu. Rev. Cell Dev. Biol.* 2015, **31**:317–45.
18. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA: **The Technology and Biology of Single-Cell RNA Sequencing [Internet]**. *Mol. Cell* 2015, **58**:610–620.
19. Till JE, McCulloch EA: **A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. [Internet]**. *Radiat. Res.* 1961, **14**:213–22.
20. Bryder D, Rossi DJ, Weissman IL: **Hematopoietic stem cells: the paradigmatic tissue-specific stem cell. [Internet]**. *Am. J. Pathol.* 2006, **169**:338–46.
21. Orkin SH, Zon LI: **Hematopoiesis: an evolving paradigm for stem cell biology. [Internet]**. *Cell* 2008, **132**:631–44.
22. Kiel MJ, Yilmaz OH, Iwashita T, Yilmaz OH, Terhorst C, Morrison SJ: **SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. [Internet]**. *Cell* 2005, **121**:1109–21.
23. Kent DG, Copley MR, Benz C, Wöhrer S, Dykstra BJ, Ma E, Cheyne J, Zhao Y, Bowie MB, Zhao Y, et al.: **Prospective isolation and molecular characterization of hematopoietic stem cells with durable self-renewal**

potential. [Internet]. *Blood* 2009, **113**:6342–50.

24. Glotzbach JP, Januszyk M, Vial IN, Wong VW, Gelbard A, Kalisky T, Thangarajah H, Longaker MT, Quake SR, Chu G, et al.: **An information theoretic, microfluidic-based single cell analysis permits identification of subpopulations among putatively homogeneous stem cells. [Internet]**. *PLoS One* 2011, **6**:e21211.
25. Akashi K, Traver D, Miyamoto T, Weissman IL: **A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. [Internet]**. *Nature* 2000, **404**:193–7.
26. Kondo M, Weissman IL, Akashi K: **Identification of clonogenic common lymphoid progenitors in mouse bone marrow. [Internet]**. *Cell* 1997, **91**:661–72.
27. Arinobu Y, Mizuno S, Chong Y, Shigematsu H, Iino T, Iwasaki H, Graf T, Mayfield R, Chan S, Kastner P, et al.: **Reciprocal activation of GATA-1 and PU.1 marks initial specification of hematopoietic stem cells into myeloerythroid and myelolymphoid lineages. [Internet]**. *Cell Stem Cell* 2007, **1**:416–27.
28. Sanjuan-Pla A, Macaulay IC, Jensen CT, Woll PS, Luis TC, Mead A, Moore S, Carella C, Matsuoka S, Bouriez Jones T, et al.: **Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. [Internet]**. *Nature* 2013, **502**:232–6.
29. Guo G, Luc S, Marco E, Lin T-W, Peng C, Kerenyi MA, Beyaz S, Kim W, Xu J, Das PP, et al.: **Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. [Internet]**. *Cell Stem Cell* 2013, **13**:492–505.
30. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, Mildner A, Cohen N, Jung S, Tanay A, et al.: **Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. [Internet]**. *Science* 2014, **343**:776–9.
31. Paul F, Arkin Y, Giladi A, Jaitin DA, Kenigsberg E, Keren-Shaul H, Winter D, Lara-Astiaso D, Gury M, Weiner A, et al.: **Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors [Internet]**. *Cell* 2015, **163**:1663–1677.

**Using the MARS-seq method previously developed by the same group, this study sequenced thousands of myeloid progenitor cells at a shallow depth. This provided information about up to ~10,000 transcripts per cell and was sufficient to partition cells into 19 clusters representing different stages of myelopoiesis. This indicated that the current immunophenotypic descriptions of myeloid cells do not adequately

describe the continuum of cell types. Additionally, this study found no evidence of co-expression of markers of different lineages, arguing against the conventional hierarchy of multipotent progenitors and for a model in which lineage commitment occurs early and differentiation more gradually.

32. Pronk CJH, Rossi DJ, Månsson R, Attema JL, Norddahl GL, Chan CKF, Sigvardsson M, Weissman IL, Bryder D: **Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. [Internet]**. *Cell Stem Cell* 2007, **1**:428–42.
33. Sun J, Ramos A, Chapman B, Johnnidis JB, Le L, Ho Y-J, Klein A, Hofmann O, Camargo FD: **Clonal dynamics of native haematopoiesis [Internet]**. *Nature* 2014, **514**:322–7.
34. Busch K, Klapproth K, Barile M, Flossdorf M, Holland-Letz T, Schlenner SM, Reth M, Höfer T, Rodewald H-R: **Fundamental properties of unperturbed haematopoiesis from stem cells in vivo [Internet]**. *Nature* 2015, **518**:542–6.
35. Grün D, Lyubimova A, Kester L, Wiebrands K, Basak O, Sasaki N, Clevers H, van Oudenaarden A: **Single-cell messenger RNA sequencing reveals rare intestinal cell types [Internet]**. *Nature* 2015, **525**:251–5.

*This study mapped the different populations of the intestinal epithlium and developed a new computational method, RaceID, for identifying very rare cell populations. This was used to identify a population of enteroendocrine cells that were validated experimentally.

36. Pollen AA, Nowakowski TJ, Shuga J, Wang X, Leyrat AA, Lui JH, Li N, Szpankowski L, Fowler B, Chen P, et al.: **Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex [Internet]**. *Nat. Biotechnol.* 2014, **32**:1053–8.

*This study investigated the depth of sequencing required to accurately classify different cell types, by sequencing at high depth and down-sampling the number of reads used for analysis. A relatively shallow sequencing depth of 50,000 reads per cell was sufficient to classify cells and identify putative biomarkers.

37. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, et al.: **Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets [Internet]**. *Cell* 2015, **161**:1202–1214.

**This study introduced DropSeq, a droplet-based sequencing technology. Each cell is captured in a nanolitre volume droplet which acts as a reaction chamber for

sequencing library generation. Droplets also contain a bead on which barcoded primers are immobilised. After reverse transcription and cDNA amplification are complete, the droplets are broken and cellular libraries are mixed for sequencing. This reduces the cost of scRNA-seq to a few cents, rather than a few dollars per cell.

38. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW: **Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells [Internet]**. *Cell* 2015, **161**:1187–1201.

**Like DropSeq, the InDrop method introduced in this study allows for droplet based generation of barcoded single cell transcriptome libraries.

39. Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, Espinoza FH, Desai TJ, Krasnow MA, Quake SR: **Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. [Internet]**. *Nature* 2014, **509**:371–5.

40. Durruthy-Durruthy R, Gottlieb A, Hartman BH, Waldhaus J, Laske RD, Altman R, Heller S: **Reconstruction of the mouse otocyst and early neuroblast lineage at single-cell resolution. [Internet]**. *Cell* 2014, **157**:964–78.

41. Noseda M, Harada M, McSweeney S, Leja T, Belian E, Stuckey DJ, Abreu Paiva MS, Habib J, Macaulay I, de Smith AJ, et al.: **PDGFRα demarcates the cardiogenic clonogenic Sca1+ stem/progenitor cell in adult murine myocardium. [Internet]**. *Nat. Commun.* 2015, **6**:6930.

42. Tetteh PW, Basak O, Farin HF, Wiebrands K, Kretzschmar K, Begthel H, van den Born M, Korving J, de Sauvage F, van Es JH, et al.: **Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their Enterocyte-Lineage Daughters [Internet]**. *Cell Stem Cell* 2016, **18**:203–13.

43. Riddell J, Gazit R, Garrison BS, Guo G, Saadatpour A, Mandal PK, Ebina W, Volchkov P, Yuan G-C, Orkin SH, et al.: **Reprogramming committed murine blood cells to induced hematopoietic stem cells with defined factors. [Internet]**. *Cell* 2014, **157**:549–64.

44. Kolodziejczyk AA, Kim JK, Tsang JCH, Ilicic T, Henriksson J, Natarajan KN, Tuck AC, Gao X, Bühler M, Liu P, et al.: **Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation [Internet]**. *Cell Stem Cell* 2015, **17**:471–485.

45. Guo G, Huss M, Tong GQ, Wang C, Li Sun L, Clarke ND, Robson P: **Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. [Internet]**. *Dev. Cell* 2010, **18**:675–85.

46. Nichols J, Smith A: **Naive and primed pluripotent states. [Internet]**. *Cell*

Stem Cell 2009, **4**:487–92.

47. Ohnishi Y, Huber W, Tsumura A, Kang M, Xenopoulos P, Kurimoto K, Oleś AK, Araúzo-Bravo MJ, Saitou M, Hadjantonakis A-K, et al.: **Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. [Internet]**. *Nat. Cell Biol.* 2014, **16**:27–37.
48. Biase F, Cao X, Zhong S: **Cell fate inclination within 2-cell and 4-cell mouse embryos revealed by single-cell RNA sequencing. [Internet]**. *Genome Res.* 2014, doi:10.1101/gr.177725.114.
49. Goolam M, Scialdone A, Graham SJL, Macaulay IC, Jedrusik A, Hupalowska A, Voet T, Marioni JC, Zernicka-Goetz M: **Heterogeneity in Oct4 and Sox2 Targets Biases Cell Fate in 4-Cell Mouse Embryos [Internet]**. *Cell* 2016, **165**:61–74.

*Sequencing of mouse blastomeres identified targets of Sox2 and Oct4, in particular Sox21, as heterogeneously expressed at the 4-cell stage. Sox21 expression was correlated with the first lineage decision between extra-embryonic and pluripotent tissues. This finding is an important step forward in understanding early symmetry-breaking events.

50. Burton A, Muller J, Tu S, Padilla-Longoria P, Guccione E, Torres-Padilla M-E: **Single-cell profiling of epigenetic modifiers identifies PRDM14 as an inducer of cell fate in the mammalian embryo. [Internet]**. *Cell Rep.* 2013, **5**:687–701.
51. Tabansky I, Lenarcic A, Draft RW, Loulier K, Keskin DB, Rosains J, Rivera-Feliciano J, Lichtman JW, Livet J, Stern JNH, et al.: **Developmental bias in cleavage-stage mouse blastomeres. [Internet]**. *Curr. Biol.* 2013, **23**:21–31.
52. Yan L, Yang M, Guo H, Yang L, Wu J, Li R, Liu P, Lian Y, Zheng X, Yan J, et al.: **Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. [Internet]**. *Nat. Struct. Mol. Biol.* 2013, **20**:1131–9.
53. Xue Z, Huang K, Cai C, Cai L, Jiang C, Feng Y, Liu Z, Zeng Q, Cheng L, Sun YE, et al.: **Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. [Internet]**. *Nature* 2013, **500**:593–7.
54. Blakeley P, Fogarty NME, Del Valle I, Wamaitha SE, Hu TX, Elder K, Snell P, Christie L, Robson P, Niakan KK: **Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. [Internet]**. *Development* 2015, doi:10.1242/dev.123547.

*Sequencing of human and mouse preimplantation embryos identified conserved and species-specific transcriptional programs. Comparison of hESCs with human

embryos identified pathways that could be modulated to produce hESCs that are transcriptionally more similar to human epiblast.

55. Hu M, Krause D, Greaves M, Sharkis S, Dexter M, Heyworth C, Enver T: **Multilineage gene expression precedes commitment in the hemopoietic system. [Internet]**. *Genes Dev.* 1997, **11**:774–785.
56. Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S: **Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. [Internet]**. *Nature* 2008, **453**:544–7.
57. Pina C, Fugazza C, Tipping AJ, Brown J, Soneji S, Teles J, Peterson C, Enver T: **Inferring rules of lineage commitment in haematopoiesis. [Internet]**. *Nat. Cell Biol.* 2012, **14**:287–94.
58. Teles J, Pina C, Edén P, Ohlsson M, Enver T, Peterson C: **Transcriptional regulation of lineage commitment--a stochastic model of cell fate decisions. [Internet]**. *PLoS Comput. Biol.* 2013, **9**:e1003197.
59. Chambers I, Silva J, Colby D, Nichols J, Nijmeijer B, Robertson M, Vrana J, Jones K, Grotewold L, Smith A: **Nanog safeguards pluripotency and mediates germline development. [Internet]**. *Nature* 2007, **450**:1230–4.
60. Toyooka Y, Shimosato D, Murakami K, Takahashi K, Niwa H: **Identification and characterization of subpopulations in undifferentiated ES cell culture. [Internet]**. *Development* 2008, **135**:909–18.
61. Hayashi K, Lopes SMC de S, Tang F, Surani MA: **Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. [Internet]**. *Cell Stem Cell* 2008, **3**:391–401.
62. MacArthur BD, Sevilla A, Lenz M, Müller F-J, Schuldt BM, Schuppert AA, Ridden SJ, Stumpf PS, Fidalgo M, Ma'ayan A, et al.: **Nanog-dependent feedback loops regulate murine embryonic stem cell heterogeneity. [Internet]**. *Nat. Cell Biol.* 2012, **14**:1139–47.
63. Buganim Y, Faddah DA, Cheng AW, Itskovich E, Markoulaki S, Ganz K, Klemm SL, van Oudenaarden A, Jaenisch R: **Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. [Internet]**. *Cell* 2012, **150**:1209–22.
64. Moignard V, Macaulay IC, Swiers G, Buettner F, Schütte J, Calero-Nieto FJ, Kinston S, Joshi A, Hannah R, Theis FJ, et al.: **Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. [Internet]**. *Nat. Cell Biol.* 2013, **15**:363–72.
65. Moignard V, Woodhouse S, Haghverdi L, Lilly AJ, Tanaka Y, Wilkinson AC,

Buettner F, Macaulay IC, Jawaid W, Diamanti E, et al.: **Decoding the regulatory network of early blood development from single-cell gene expression measurements [Internet]**. *Nat. Biotechnol.* 2015, doi:10.1038/nbt.3154.

66. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL: **The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells [Internet]**. *Nat. Biotechnol.* 2014, **advance on**.

**Introduced the concept of cellular pseudotime, in which cells are ordered based on the similarity of their transcriptomes in lieu of real temporal information. This paper described the popular 'Monocle' algorithm for ordering single cell transcriptomes in pseudotime.

67. Camp JG, Badsha F, Florio M, Kanton S, Gerber T, Wilsch-Bräuninger M, Lewitus E, Sykes A, Hevers W, Lancaster M, et al.: **Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. [Internet]**. *Proc. Natl. Acad. Sci. U. S. A.* 2015, **112**:15672–15677.
68. Bendall SC, Davis KL, Amir E-AD, Tadmor MD, Simonds EF, Chen TJ, Shenfeld DK, Nolan GP, Pe'er D: **Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. [Internet]**. *Cell* 2014, **157**:714–25.
69. Haghverdi L, Buettner F, Theis FJ: **Diffusion maps for high-dimensional single-cell analysis of differentiation data [Internet]**. *Bioinformatics* 2015, **31**:2989–98.
70. MacArthur BD, Lemischka IR: **Statistical mechanics of pluripotency. [Internet]**. *Cell* 2013, **154**:484–9.
71. Eberwine J, Kim J: **Cellular Deconstruction: Finding Meaning in Individual Cell Variation. [Internet]**. *Trends Cell Biol.* 2015, **25**:569–78.
72. Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R: **Smart-seq2 for sensitive full-length transcriptome profiling in single cells. [Internet]**. *Nat. Methods* 2013, **10**:1096–8.
73. Bendall SC, Simonds EF, Qiu P, Amir ED, Krutzik PO, Finck R, Bruggner R V, Melamed R, Trejo A, Ornatsky OI, et al.: **Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. [Internet]**. *Science* 2011, **332**:687–96.
74. Guo H, Zhu P, Wu X, Li X, Wen L, Tang F: **Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed**

using reduced representation bisulfite sequencing. [Internet]. *Genome Res.* 2013, **23**:2126–35.

75. Cusanovich DA, Daza R, Adey A, Pliner H, Christiansen L, Gunderson KL, Steemers FJ, Trapnell C, Shendure J: **Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing [Internet]**. *Science (80-.).* 2015, **348**:910–4.
76. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY, Greenleaf WJ: **Single-cell chromatin accessibility reveals principles of regulatory variation. [Internet]**. *Nature* 2015, **523**:486–490.
77. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, Laue ED, Tanay A, Fraser P: **Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. [Internet]**. *Nature* 2013, **502**:59–64.
78. Macaulay IC, Haerty W, Kumar P, Li YI, Hu TX, Teng MJ, Goolam M, Saurat N, Coupland P, Shirley LM, et al.: **G&T-seq: parallel sequencing of single-cell genomes and transcriptomes [Internet]**. *Nat. Methods* 2015, **12**:519–22.
79. Angermueller C, Clark SJ, Lee HJ, Macaulay IC, Teng MJ, Hu TX, Krueger F, Smallwood SA, Ponting CP, Voet T, et al.: **Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity [Internet]**. *Nat. Methods* 2016, **13**:229–32.
80. Lovatt D, Ruble BK, Lee J, Dueck H, Kim TK, Fisher S, Francis C, Spaethling JM, Wolf JA, Grady MS, et al.: **Transcriptome in vivo analysis (TIVA) of spatially defined single cells in live tissue. [Internet]**. *Nat. Methods* 2014, **11**:190–6.
81. Achim K, Pettit J-B, Saraiva LR, Gavriouchkina D, Larsson T, Arendt D, Marioni JC: **High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin [Internet]**. *Nat. Biotechnol.* 2015, **33**:503–9.
82. Satija R, Farrell JA, Gennert D, Schier AF, Regev A: **Spatial reconstruction of single-cell gene expression data [Internet]**. *Nat. Biotechnol.* 2015, **33**:495–502.
83. Peng G, Suo S, Chen J, Chen W, Liu C, Yu F, Wang R, Chen S, Sun N, Cui G, et al.: **Spatial Transcriptome for the Molecular Annotation of Lineage Fates and Cell Identity in Mid-gastrula Mouse Embryo [Internet]**. *Dev. Cell* 2016, **36**:681–697.

*In this study, a spatial map of gene expression in the mouse epiblast was built by sequencing small groups of cells from multiple regions and sections of mid-gastrulation embryos. Single cell transcriptomes could then be mapped back against

this reference to provide a spatial dimension to sc-RNA-seq data.

84. Hoppe PS, Coutu DL, Schroeder T: **Single-cell technologies sharpen up mammalian stem cell research. [Internet]**. *Nat. Cell Biol.* 2014, **16**:919–27.

85. Macaulay IC, Voet T: **Single cell genomics: advances and future perspectives. [Internet]**. *PLoS Genet.* 2014, **10**:e1004126.

86. Trapnell C: **Defining cell types and states with single-cell genomics [Internet]**. *Genome Res.* 2015, **25**:1491–1498.