New Insights into Haematopoietic Differentiation Landscapes from scRNA-seq

Sam Watcham, Iwo Kucinski*, Berthold Gottgens*

Department of Haematology and Wellcome & MRC Cambridge Stem Cell Institute, Hills Road, Cambridge CB2 0XY, UK

* Corresponding Authors (BG - bg200@cam.ac.uk, IK – idk25@cam.ac.uk)

Address for Correspondence:

Cambridge University Department of Haematology, Cambridge Institute for Medical Research & Wellcome and MRC Cambridge Stem Cell Institute Hills Road, Cambridge CB2 0XY, UK

Abstract

Single cell transcriptomics has recently emerged as a powerful tool to analyse cellular heterogeneity, discover new cell types and infer putative differentiation routes. The technique has been rapidly embraced by the haematopoiesis research community and like other technologies before, single cell molecular profiling is widely expected to make important contributions to our understanding of the haematopoietic hierarchy. Much of this new interpretation relies on inference of the 'transcriptomic landscape' as a representation of existing cellular states and associated transitions between them. Here we review how this model allows, under certain assumptions, charting of 'time-resolved' differentiation trajectories with unparalleled resolution and how the landscape of multipotent cells may be rather devoid of discrete structures, challenging our preconceptions about stem and progenitor cell types and their organisation. Finally, we highlight how promising technological advances may convert static differentiation landscapes into a dynamic cell flux model and thus provide a more holistic understanding of normal haematopoiesis and blood disorders.

Introduction

Haematopoiesis research spanning over 150 years has been significantly driven by technological breakthroughs. Microscopy-based observations in the 19th century established that blood is composed of two bone marrow-derived cell lineages: myeloid and lymphoid, perhaps sharing a common 'stem cell' origin[1]. It wasn't until the 1950s when bone marrow transplantation rescue of lethally irradiated mice[2–4] confirmed this hypothesis. Subsequently, *in vitro* haematopoietic colony assays provided functional evidence for intermediate stages between haematopoietic stem cells (HSCs) and terminally differentiated cells[5] – ranging from multipotent to unipotent progenitor cells. These findings arose from the shadow cast by the destructive effects of radiation on the blood

system following the first use of nuclear weapons in the 1940's[6], with the first successful human bone marrow transplant reported already in 1959[7,8]. This approach remains the only curative therapy for a number of haematopoietic malignancies to this date[9]. While these practical applications were developed early on, our biological understanding of haematopoiesis has lagged behind until isolation of specific cell populations became possible.

A critical advance came from the related field of immunology, allowing the sorting of individual cells[10] and generation of monoclonal antibodies to detect surface markers[11]. At this stage, a key achievement of the haematopoietic community had begun to take form – with the establishment of the differentiation 'tree'. By the end of the 20th century, the haematopoietic tree was rooted in the long-term HSCs (LT-HSCs), followed by short-term HSCs (ST-HSCs) and multipotent progenitors (MPPs) - partitioned according to their ability to repopulate blood in transplantation assays over diminishing periods of time[12–16]. These cells were proposed to differentiate through a set of bifurcations that produced distinct progenitor cell populations with decreasing lineage potential and self-renewal activity (Figure 1A). In the past two decades, this model has been subjected to constant extensions and refinements, largely due to new evidence highlighting cellular heterogeneity obtained from single cell assays. At the same time, cell barcoding approaches mediated clonal tracking of native haematopoiesis[17–19] and stressed the importance of gaining insight into the unperturbed tissue state. The resulting evolution of the haematopoietic tree has been discussed in detail elsewhere[6,15,20,21].

We are currently witnessing another "single cell revolution", in which vast transcriptomic datasets are transforming our understanding of haematopoiesis. As a result, the idea of cellular transitions between discrete progenitor states as they differentiate has become difficult to accommodate[20]. Instead, multiple studies propose the idea of continuous differentiation landscapes with little or no

discrete differentiation stages and smooth transitions across the cell states. In this context cells within a heterogeneous pool of haematopoietic stem and progenitor cells (HSPCs) differentiate along a multitude of potential trajectories that contain poorly-defined 'branchpoints' which determine a particular cell's fate. In this review we aim to highlight recent biological insights gained into the nature of these landscapes using single-cell RNA sequencing and downstream computational tools.

scRNA-Seq – opportunities and limitations

While single cell quantification of gene expression for small numbers of genes was already achieved in the early 1990's[22], it is thanks to breakthroughs in parallelisation in the past few years that single cell transcriptomics is now chasing its conceptual predecessors flow and mass cytometry in terms of throughput[23]. However unlike flow/mass cytometry measurements, which are typically restricted to at most a few dozen pre-defined markers, scRNA-Seq can measure expression of up to 10^4 genes simultaneously in each single cell, thus offering unprecedented detail for the definition of cellular states. Two key parameters of any scRNA-Seq experiment are (i) the number of cells assayed, which determines the probability of being able to capture rare cell populations, and (ii) detection sensitivity, which dictates the number of genes available for cell-cell comparisons. These two criteria are leveraged into two classes of scRNA-Seq platforms. High-throughput methods (e.g. droplet-based like Drop-Seq, InDrops, 10X genomics)[24–26] provide in excess of 10^4 cells per experiment but detect "only" 1,000-3,000 expressed genes per cell. By contrast, low-throughput methods (e.g. plate-based like Smart-Seq2, CelSeq2, mcSCRB-Seq, RamDA-Seq)[27–30] are typically used to profile up to ~10^3 cells but detect >5,000 genes per cell. The details of specific methods and their technical considerations have been extensively reviewed[23,31,32].

For organ systems with a fast turn-over such as haematopoiesis, scRNA-Seq offers a snapshot of cells and their expression states at a particular time-point. While missing temporal information, much excitement has been generated by the idea that the distribution of single cell expression profiles in the high-dimensional space can be considered as a transcriptional landscape, which encodes information on cellular transitions enabling ordering of cells along 'pseudo-time' from immature progenitors towards the various haematopoietic lineages[33,34]. The inference of such putative differentiation trajectories requires certain assumptions and imposes some limitations: (i) *differentiation is a continuous process* – detection of 'jumps' between cell states is difficult, (ii) *cells*

differentiate asynchronously – cells are captured at multiple points along their differentiation routes, (iii) *cells differentiate using defined (non-random) trajectories*, (iv) *cells move in one direction* – this requires additional knowledge to determine start and finish, (v) *cell state information is complete*. The last point is clearly problematic because we cannot see hidden variables like epigenetics or cell location with current scRNA-Seq protocols. Furthermore, some analytic methods require stricter assumptions, such as a 'tree-like' structure of the data, where cells undergo a series of bifurcations during their differentiation, or absence of oscillations between cell states (e.g. cell cycle), which may disregard potentially relevant biological information. For a detailed discussion of these assumptions we refer readers to Weinreb et al.[35].

The past six years have seen an explosion of scRNA-Seq studies within the haematopoiesis field, starting from profiling of previously well-defined haematopoietic populations[36–39], then shifting towards less biased selection of cells[26,40–43], and finally arriving at whole tissue studies surveying $>10^5$ cells. We have summarised some of the most useful resources in Table 1. The current culmination comes from a large collaborative effort – the Human Cell Atlas[42], which aims to create reference maps for cells across >50 tissues in the human body. This includes a recently released dataset with >500,000 human bone marrow and cord blood cells, providing the most complete scRNA-Seq map yet of the human haematopoietic compartments[44]. We present a general view of this dataset in Figure 1B. Of note, we have excluded Natural Killer and T lymphocytes from the analysis, which account for ≈50% of cells and appear disconnected from the HSPC cluster. This separation is expected for the T cells, which differentiate in the thymus but may also indicate that NK progenitors are rare or even absent in the BMMC fraction. As these large-scale datasets await their full exploitation, we will focus on the main insights into haematopoietic differentiation delivered by scRNA-Seq so far.

scRNA-Seq reshapes the bottom and middle-tiers of the haematopoietic tree

scRNA-Seq has quickly proved to be a powerful tool for distinguishing discrete cell states. As studies concerning lymphoid lineages have been reviewed previously[45], we will discuss a couple of examples from the myeloid lineage. Single cell profiling allowed discovery of several dendritic cell (DC) subtypes, which respond differently to LPS (Lipopolysaccharides)[46]. A recent study laid out a new taxonomy for human DCs and monocytes in peripheral blood[47] with a new DC population responsible for T-cell activation previously misclassified as plasmacytoid DC, and a new conventional DC progenitor population ($CD100^{Hi}$, $CD34^{Int}$), functionally distinct from the $CD34^{Hi}$ HSCs. Yanez et al.[48] showed that $Ly6C^{hi}$ monocytes can be divided into: 'neutrophil-like' and monocyte-derived dendritic cell precursors arising through alternative differentiation routes in agreement with another report[49].

Recent studies have also challenged our understanding of the oligopotent progenitor compartments. Profiling of the ($Lineage^{-}$, $Sca1^{-}$, $c\text{-}Kit^{+}$) population containing the CMP/GMP/MEP populations (Figure 1A) revealed an unanticipated heterogeneity, with at least 18 distinct subtypes[36] with various degrees of lineage priming (i.e. expression of lineage specific genes) reflecting their progressive commitment. Importantly the key surface markers FcgR and CD34, typically used to distinguish the CMP/GMP/MEP populations, turned out to be poor predictors of cell identity and lineage bias. Consistently, Olsson et al.[39] showed that only a small proportion of cells sorted from conventional myeloid progenitor populations displayed gene expression profiles consistent with multi-lineage progenitor activity. This study also elegantly demonstrates how the wealth of information provided by scRNA-Seq can be exploited to infer gene regulatory information, refine cell isolation strategies and enrich for specific bipotent progenitors, by focusing on the counteracting regulators Gfi1 and Irf8.

These findings are in line with the accumulating evidence from single cell colony assays[36,39,50–53], single cell transplantations[36,54,55] and cellular barcoding[19,56] experiments performed both in human and mouse, which collectively suggest that most cell fate decisions are taken earlier than expected from the classical haematopoietic tree model. This is also consistent with the original observations that many cells within progenitor gates (e.g. CMPs or LMPPs) are already lineage restricted[50,57]. The emerging picture therefore is that unipotent cells dominate the middle-tier of haematopoietic progenitors, albeit with a small contribution of bi/oligopotent progenitors. Whether these rare cell populations are true functional intermediates or perhaps outlier cells 'spilling over' from less mature compartments remains to be seen.

Revisiting the HSC and MPP cell compartments by scRNA-Seq

The upper tier of the haematopoietic tree classically contained three multipotent subpopulations with decreasing repopulation potentials: LT-HSCs, ST-HSC (also known as MPP1) and MPPs, all three capable of generating both myeloid and lymphoid cells[12,54,58]. Before the advent of scRNA-Seq, immunophenotyping showed that MPPs have at least 4 subpopulations (MPP1-4) with distinct cell cycle characteristics[12], molecular features[37,59–62] and functional biases[58,61,62]. Subsequent molecular analysis at the single cell level, first by sc-qPCR[62] and later by scRNA-Seq[19,37,63] showed that the HSC/MPP compartment may be characterised by a continuous landscape. Furthermore transcriptional priming towards erythroid/megakaryocytic (MPP2), myeloid (MPP3) or lymphoid lineages (MPP4), suggested that many key fate decisions have already been initiated by earlier upstream progenitors (MPP1). Rodriguez-Fraticelli et al.[19] confirmed these findings *in vivo* by employing high-throughput scRNA-Seq and lineage barcoding to quantify the degree of lineage priming and lineage contributions of each of the MPP subpopulations. Importantly, the compartment structure adapts to changing conditions, as the MPP subpopulations shift towards more myeloid bias with increased animal age[37] or following external stress[62].

The 'true' LT-HSC state characterised by the most durable self-renewal has been pursued for decades[12,54,58,64–68]. Comparison of various HSC isolation strategies[38] allowed the definition of 'Molecular Overlap' (MolO) and 'Surface Marker Overlap' (SuMO) scores, which describe gene expression patterns correlating with stem cell self-renewal capacity. A major component of the MolO score are genes negatively regulating the cell cycle, consistent with the long-standing idea that dormancy marks stem cells with the highest self-renewal[12]. The transition along the dormancy axis also appears to be gradual, with the dormant end of the spectrum characterised by deep quiescence (=slow return to the cell cycle), low biosynthetic activity and high retinoic acid signalling[69], as well as quiescence and low respiration being conserved amongst the most immature human HSPCs[70].

The other aspect of HSC heterogeneity is lineage output bias, which may manifest in lineage priming. While Kowalczyk et. al.[37] found no evidence for transcriptional lineage priming in HSCs regardless of age, Grover et al. identified a subset of HSCs expressing megakaryocytic genes including the *Vwf* marker[71] in old mice. The reason for this disparity is not clear, but may be due to cell isolation techniques, animal breeding and ageing or the bioinformatic tools used. Nevertheless, the presence of megakaryocyte-biased HSCs has been confirmed through transplantation experiments[19,71–73] and tracking of native haematopoiesis[19], indicating that committing to a megakaryocytic fate may be one of the earliest fate decisions. Interestingly, these megakaryocyte-biased HSCs can still behave as multi-potent stem cells following transplantation highlighting that cell potential does not necessary reflect cell behavior under native conditions. HSC skewing towards myeloid and lymphoid lineages evident from transplantation experiments[67,74,75] remains unexplained within the transcriptional landscape, however data currently only available as preprint[76] suggest that a myeloid-biased HSC subpopulation becomes detectable in aged mice following inflammation consistent with previous functional data[77,78].

196

197 **How to navigate the haematopoietic differentiation landscape**

198 As the HSPC landscape appears to be at least in part continuous, traditional immunophenotyping
199 approaches aiming to dissect distinct populations would now seem, at least to some extent,
200 counterintuitive. Instead, broad and unbiased approaches tracking possible routes through the
201 landscape are required. Recently there has been a surge of analytical methods for differentiation
202 trajectory inference. They commonly rely on measuring the 'distance' between cells in a high
203 dimensional gene expression space, under the assumption that cells with small distance between
204 them are related, for example representing stages of the same differentiation trajectory. The aim is
205 to find a measure of distance that reflects the structure of the data, rather than just relying on a
206 shortest possible path (Figure 2A). To perform this task, most methods use dimensionality reduction
207 techniques to learn the data structure[79–82] and simplify computation of cell-cell distances (Figure 2A).
208 However, excessive dimensionality reduction will inevitably oversimplify the data (Figure 2B). Hence
209 there is a clear distinction between 'learned data' - 10-100 dimensional representation with minimal
210 information loss, and 'visualisable data' – 2-3 dimensions interpretable by humans but at the
211 expense of potentially discarding important biological information (Figure 2C). This means that
212 whilst 'learned data' is used for trajectory inference, it is not advisable to interpret complex
213 differentiation pathways simply from a 2D representation of a dataset (tSNE, UMAP etc.).

214

215 Early methods for trajectory inference focused on the idea of 'pseudotime', where the calculated
216 distances are used to order cells along a putative trajectory according to their distance from a
217 predefined starting point[33,34,83–87]. This ordering allows recovery of the pseudo-dynamic gene
218 expression along a trajectory. However, the need to assign cells to unidirectional trajectories
219 restricts its applicability. Attempts to improve upon this by implementing methods that can discover
220 putative 'branchpoints' in a dataset may provide a solution to this, but their ability to produce

bifurcations consistent with functional data in an unsupervised manner has been limited[87–89]. Alternative methods are aiming to identify the likely connection between clustered cells (StemID[90], PAGA[91]) or stepwise cell classification working back towards the stem cells (FateID[92]). Population balance analysis takes a physically-motivated approach[35], attempting to constrain the problem of trajectory inference into a set of differential equations describing the flux of cells through the transcriptional landscape. These methods and others like them[93] successfully recapitulate trajectories through the unipotent haematopoietic states, but can also elucidate routes of differentiation amongst the heterogeneous HSPCs, as highlighted in the following section.

An evolving view of the HSPC hierarchy

An early analysis of the multipotent stem/progenitor compartment examined 1,600 cells spanning 10 classically defined HSPC populations[63]. While the surface marker information served as reference points for this and future studies, scRNA-Seq analysis identified three broad trajectories in the differentiation landscape: lymphoid, erythroid and granulocytic/monocytic together with their gene expression signature. A more complete view was provided by high-throughput scRNA-Seq platforms, which allowed dense cell sampling of large populations (typically >4,000 cells), overcoming the restriction of narrow sorting gates. Consequently, trajectories towards: megakaryocytes, erythrocytes, monocytes/dendritic cells, lymphoid cells, neutrophils, and rare populations of basophils, mast cells and eosinophils are now beginning to be defined[40,41]. The emerging cellular hierarchy is largely consistent with the one inferred from recent *in situ* barcoding studies[18,19], with the exception of the erythroid lineage. This branch appears to be coupled with the megakaryocytic fate in scRNA-Seq experiments[41,91] while barcoding data indicate closer linkage with Monocytic/Granulocytic lineages instead. Further experiments focusing on native haematopoiesis will be required to resolve this issue.

246 The identified trajectories in scRNA-Seq data contain a wealth of information. Using gene expression
247 correlations it is possible to extract putative regulators of fate decision and drivers or inhibitors of
248 differentiation pathways, which will greatly facilitate future studies[41,63]. In depth analysis of a single
249 trajectory can also help explain dynamics of differentiation. Tusi et al. delineated stages of early
250 erythroid differentiation within the transcriptional landscape including traditionally defined CFU-E,
251 BFU-E populations and described cell cycle remodelling at putative cell amplification stages during
252 differentiation. This provided a platform to analyse the effects of EPO stimulation on the progenitor
253 population structure and their cell cycle profiles. Global analysis of the progenitor landscape
254 revealed a surprising pattern of multiple types of mature cells arising via more than one trajectory.
255 Analysis of c-Kit$^+$ cells for example suggested that monocytes have two origins, one coupled with the
256 DC branch and one with the neutrophil branch[41]. Similarly, megakaryocytes are predicted to arise
257 directly from Mk-biased HSCs as well as through a multipotent progenitor intermediate[19]. Finally, the
258 yet unpublished PAGA analysis of Lin$^-$ c-Kit$^+$ population suggests that basophils can originate from
259 the neutrophil/monocyte branch or from the Ery/Mk branch[91] with the latter supported by recent
260 computational results[41]. Although the complexity of megakaryocyte and monocyte trajectories is
261 supported by experimental data[19,48,71], firmly establishing the existence of alternative basophil
262 trajectories would still benefit from further experimental validation.

263

264 While mice proved to be an excellent model system, there is accumulating evidence suggesting that
265 human progenitors may be organised differently. A 2016 study used clonal assays and xenograft
266 transplants to show that in human adult bone marrow, unlike the foetal liver, previously described
267 oligopotent progenitors (MPPs, CMPs, multi-lymphoid progenitors) are predominantly unipotent and
268 arise directly from the multi-potent stem cell compartment[55]. Subsequently, a detailed scRNA-Seq
269 study analysed the most immature HSPC compartments (Lin$^-$, CD34$^+$, CD38$^-$, and Lin$^-$, CD34$^+$, CD38$^+$)
270 and revealed that while the unipotent progenitor populations described above indeed form discrete

subpopulations, the stem cell compartment appears to form a rather continuous structure[70]. From lack of obvious priming in the immature populations (with exception of a minor lymphoid/myeloid vs Mk/Ery bias), the authors concluded that HSCs exist in a fluid "cloud" state, which gives rise directly to committed progenitors without much cell hierarchy in between. Furthermore even the least primed populations can give rise to single lineages *in vitro*, indicating that true multipotent cells constitute only a small fraction of the conventional stem cell population. These findings are at odds with evidence of the HSPC hierarchy from the murine system, where the upper tiers, while difficult to resolve molecularly, have functionally multipotent output.

Although differences between the human and mouse haematopoietic landscapes are bound to exist, drawing conclusions about the underlying reasons remains very challenging, because the current scRNA-Seq data analyses produce a rather 'flat' HSPC landscape with limited ability to resolve distinct cellular states[19,37,38,41,63]. Some aspects of population structure may be drowned in the noise caused by processes such as cell cycle or metabolism. Furthermore, current scRNA-Seq data represent an incomplete view of cellular states, due to the failure to detect lowly expressed genes, and the absence of entire domains of information, such as protein levels or epigenetic status. Importantly, the current mouse and human datasets have been analysed using different bioinformatic methods and careful cross-analysis is required before drawing any strong conclusions.

Outlook: extending the paradigm

In light of the accumulated scRNA-Seq data, the concept of clearly demarcated stem and progenitor cell types becomes questionable. Barring the caveat of hidden variables (proteins, epigenetics, cell location etc.), the landscape encompassing the most primitive HSPCs appears mostly continuous and flat. It is difficult to argue for a reason why the progenitor hierarchy needs to be discrete, other than simplicity or aesthetics. Indeed, even the very surface markers used for progenitor isolation commonly exhibit continuous, rather than discrete, levels. The notion of progenitor types has been

historically dictated by technical limitations: ability to observe cells using only a small number of markers and limited number of parallel functional assays. By contrast, in a transcriptomic landscape each cell is positioned using information from several thousand genes.

Of note, a continuous and flat shape does not imply lack of embedded features or information. On the contrary, it is evident that positions/territories within this space are functionally relevant and are associated with key functional qualities, such as durability of self-renewal[38,69] or cell output[19,36,41,94], measured using a variety of techniques including cell barcoding[19], HSC transplantation assays[36,38,69] or in vitro clonal assays[41]. Because of this continuous nature however, a typical flow cytometry gating approach provides an arbitrary section through the progenitor landscape providing a mixture of cells, covering a range of functional outputs (Figure 3A). While undoubtedly useful, isolation of specific populations offers a fundamentally restricted insight into the organisation and dynamics of stem and progenitor cells.

Transcriptomic data offer a more complex and likely more faithful representation, without the need for subjective categories. Moreover, if necessary for experimental purposes, discretisation is assisted by rich expression information. While scRNA-Seq data is static, it encodes information on cell states as they exist *in vivo*, and at deep coverage has the potential to capture molecular states representative of cellular transitions. This means that for each location within the landscape (=cellular state) it may become possible to infer transition directions and probabilities associated with them under native conditions. Quantitative description of this cell flux through the multidimensional space will constitute a major advance (Figure 3B). We highlight four directions, which will facilitate this process and combined will advance our understanding of haematopoiesis.

Towards complete cell state information. The single cell landscape constitutes an essential framework but is currently limited to a subset of mRNA information and lacks potential heterogeneity hidden at protein, epigenetic or tissue location level. Further development of scRNA-Seq technology will offer denser sampling[95], increased coverage/accuracy of transcriptional profiling[30] and combined with new techniques enable simultaneous detection of proteins (CITE-Seq or REAP-Seq)[96,97] or chromatin status[98–101]. Imaging based transcriptomics is also being developed to complement the data with spatial information[102,103]. Altogether, this near-complete information will provide precise locations of cells in a multi-dimensional feature space and tie together information at molecular, cellular and tissue levels.

Advanced analytic tools. As increasingly complex datasets accumulate, the analysis becomes more challenging. This includes trajectory inference, identification of branching points and extraction of gene regulatory information. Quantitative description of cell fluxes through the progenitor space is still in its infancy but newly developed numerical frameworks already attempt to approximate cell transitions from snapshot data (see section "How to navigate the haematopoietic differentiation landscape")[35,91,104]. Nonetheless parallel experimental information remains critical to provide directionality and real-time information for cell differentiation processes.

Real-time cell flux through the progenitor space. Rather than discretising the HSPCs by immunophenotyping, a more promising approach would be to link cell positions within the landscape with experimentally derived cell output (Figure 3B), under both native and transplantation conditions. Thus, each position would encode information on cellular flow in particular directions, quantifying differentiation and self-renewal. New technologies allowing simultaneous *in vivo*/*in vitro* barcoding and single cell transcriptional profiling (where a barcode can be assigned to cells within the scRNA-Seq data) are emerging, either using transposon tagging[105] or CRISPR scarring[106,107]. This should enable integration of real-time cell ancestry information with transcriptomic landscapes, and

potentially allow for the identification of transcriptomic signals generated only transiently during real-time differentiation. For finer time scales and insight into cell cycle related effects, pulse-chase experiments may become important[108].

Molecular mechanisms driving cell transitions. To manipulate haematopoiesis we need to understand the underlying molecular mechanisms. Correlations inferred from expression data provide ample hypotheses for regulatory mechanisms but experimental testing is essential. Gene regulatory networks and chromatin state can now be efficiently interrogated with recently developed techniques combining CRISPR screening with scRNA-Seq[109–114]. Targeting multiple genes and observing the effects globally will help us understand how RNA, proteins, epigenetics and extrinsic signals establish the shape of differentiation landscape and drive cellular fluxes.

Altogether, these approaches will provide a reference framework with computational modelling capacity, a promising starting point for understanding abnormal haematopoiesis. Even at this early stage, single cell transcriptomics provides valuable insight. As exemplified in Dahlin et al.[40] global analysis of the progenitor compartment revealed that c-Kit defective signalling reshapes the top of the haematopoietic hierarchy and blocks the mast cell fate at an early stage in mice. Combining scRNA-Seq with detection of mutations in single cells opens the possibility to analyse samples directly from leukaemia patients. Giustachini et al.[115] applied this approach to identify cell subpopulations persisting prolonged chemotherapy and related to blast crisis in CML patients. A global comparison of single cell landscapes between leukemic and wild-type states will reveal new cellular states or changes in the cellular fluxes associated with e.g. changes in self renewal or enhanced/reduced differentiation in particular lineages (Figure 3C). Combining our knowledge on cell flux with regulatory mechanisms will enable a more informed development of future therapies.

Thus we expect that scRNA-Seq analysis will soon shed new light on leukaemia pathogenesis and become closely relevant to the clinical setting.

Authorship

SW, IDK and BG wrote the paper together. SW and IDK generated the diagrams. All authors approved the final manuscript. The authors declare that there are no conflicts of interest.

Acknowledgements:

Work in the Gottgens Laboratory is funded by grants from Wellcome; Bloodwise; Cancer Research UK; National Institutes of Health (NIDDK DK106766); and core support grants by Wellcome to the Wellcome–Medical Research Council Cambridge Stem Cell Institute. SW is the recipient of a Medical Research Council Studentship.

Table 1. A selection of scRNA-Seq datasets and resources relevant to haematopoiesis. Dedicated online resources are hyperlinked. Raw and processed data can be obtained from the EBI Single Cell Expression Atlas or Gene Expression Omnibus databases.

Ref.	Cell number	Cell populations	Platform	Summary	Online Resource?
		HSCs			
Grover et al. 2016[71]	>130	BM: LT-HSC (LSK, $CD150^+$ $CD48^-$)	C1 + SMARTer	Profiling of young and old mice reveals expansion of platelet-biased HSCs	No
Kowalczyk et al. 2015[37]	>1100	BM: LT-HSC (LSK, $CD150^+$, $CD48^-$), ST-HSC (LSK, $CD150^-$, $CD48^-$), MPP (LSK, $CD150^-$, $CD48^+$)	SMART-Seq	Analysis of most immature stem and progenitor populations from old and young mice	Yes
Wilson et al. 2015[38]	>90	BM: LT-HSC (LSK, $CD34^-$, $Flt3^-$, $CD48^-$, $CD150^-$)	SMART-Seq2	Analysis of the LT-HSCs reveals subpopulations with highest repopulation potential. Surface marker data allows immunophenotyping.	No
Cabezas-Wallscheid et al. 2017[69]	>310	BM: LT-HSC(LSK, $CD34^-$, $CD48^-$, $CD150^+$, $CD135^-$), LRC/non-LRC	C1 + SMARTer	Comparison of active and dormant HSC populations	No
		Narrow gate datasets (HSPCs)			
Paul et al. 2015[36]	>2700	BM: Lin^-, Kit^+, $Sca1^-$ cells	MARS-Seq	Dissection of heterogeneity and lineage bias within CMP, MEP, GMP. Surface marker data allows immunophenotyping.	No
Nestorowa et al. 2016[63]	>1600	BM: HSPCs including: LT-HSC, LMPP, MPP, MEP, CMP, GMP	SMART-Seq2	Reconstruction of three differentiation trajectories (erythroid, granulocytic-monocytic, lymphoid). Surface marker data allows immunophenotyping.	Yes
Herman, Sagar and Grun 2018[92]	>2800	BM: LSK, LMPP, CLP + unipotent progenitors (i.e. B, DC, NK, Neu and Ery)	CEL-Seq2	Highlights trajectories towards B, Neutrophil/monocyte, DC and erythrocytes	Yes
Olsson et al. 2016[39]	>380	BM: LSK, CMP, GMP, (Lin^- Kit^+, $CD34^+$)	C1 + SMARTer	Dissection of heterogeneity and lineage bias in the intermediate progenitors	Yes
Rodriguez-Fraticelli et al. 2018[19]	>4900	BM: LT-HSC, ST-HSC, MPP2, MPP3, MPP4	inDrops	Highlights lineage priming in the multipotent progenitor compartment and direct HSC-megakaryocyte differentiation trajectory	Yes
Velten et al. 2017[70]	>1400	BM: HSPCs ($CD34^+$, Lin^-)	Quartz-Seq	Transcriptomics/functional data suggest 'cloud-HSC' state in human followed by unipotent progenitors. Surface marker data allows immunophenotyping.	Yes
		Broad gate datasets (HSPCs + differentiated cells)			
Villani et al. 2017[47]	>2300	PB: A range of DC and monocyte populations	SMART-Seq2	Analysis proposes a new taxonomy for human DCs and monocytes	Yes
Han et al. 2018[43]	>400,000 BM: >38,000	51 mouse tissues, including total BM and c-Kit^+ fraction	Microwell-Seq	Large scale overview of the bone marrow and its progenitor compartment, includes also polymorphonuclear cells	Yes

Tusi et al. 2018[41]	>7300	BM: c-Kit$^+$	inDrops	Highlights Basophil, Megakaryocyte and Erythrocyte differentiation branches and effects of EPO stimulation	Yes
Dahlin et al. 2018[40]	>58000	BM: LSK, LK	10X	Multiple differentiation trajectories and identification of basophil/mast cells progenitors, comparison with c-Kit defective haematopoietic system	Yes
Zheng et al. 2017[26]	>180,000	PBMC (WT) + BMMC (AML patients + controls)	10X	Overview of the PBMCs (WT patients) and BMMCs before and after transplantation	No
Peterson et al. 2017[97]	>15,400	PBMC: CD3$^+$ T cells, CD11b$^+$ myeloid cells, CD19$^+$ B cells	10X + REAP-Seq	Simultaneous analysis of the whole transcriptome and protein levels for 45 surface markers	No
Regev et al. 2017[42,44]	>270,000	BMMC	10X	The largest dataset of human bone marrow yet	Yes
(10X Genomics)	>8300	PBMC	10x	PBMCs from a from a healthy human donor	No
Gierahn et al. 2017[116]	>3600	PBMC	Seq-Well	Resolves major cell types and highlights heterogeneity in the monocyte population	No
Stoeckius et al. 2017[96]	>8000	CBMCs	10X + CITE-Seq	Simultaneous analysis of the whole transcriptome and protein levels for 13 surface markers	No
Zheng et al. 2018[117]	>21300	CB: CD34$^+$	Drop-Seq	Analysis reconstructs trajectories towards four distinct cell fates in the most immature compartment	Yes
Regev et al. 2017[42]	>270,000	CBMCs	10X	The largest dataset of human cord blood yet	No
Tabula Muris Consortium[118]	>8,000	BM: unfractionated + LSK, B cells, T/NK cells, granulocyte and monocyte fractions	SMART-Seq2 + 10X	>100,000 cells profiled from 20 mouse organs	Yes
Gene perturbations and diseases					
Dixit et al. 2016[109]	~70,000	BM: dendritic cells	10X	CRISPR perturbations provide insight into regulators of dendritic cells	Yes
Jaitin et al. 2016[112]	>10,000	BM: myeloid cells (CD11c+) cells and HSPCs (LSK)	MARS-Seq	CRISPR perturbations provide insight into regulators of myeloid cells in vitro and in vivo	No
Giustacchini et al. 2017[115]	>2,000	BM: Lin$^-$, CD34$^+$, CD38$^-$ HSPCs	SMART-Seq2	Simultaneous analysis of leukemic and normal cells from patients with chronic myeloid leukaemia. Modification to SMART-Seq2 protocol allows mutations detection	No
Zhao et al. 2017[119]	>970	BM: CD34$^+$	C1 + SMARTer	Use of scRNA-Seq to detect aneuploidy in patients	No

Mouse/Human

BM - bone marrow, PB - peripheral blood, BMMC - bone marrow mononuclear cells, PBMC - bone marrow mononuclear cells, Lin - lineage markers cocktail, CB - cord blood, CBMC - Cord blood mononuclear cells, LSK - (Lin$^-$, Kit$^+$, Sca1$^+$), LK - (Lin$^-$, Kit$^+$), LRC - label-retaining cells, EPLM - Early progenitor with lymphoid and myeloid potential

References

1. Boisset JC, Robin C. On the origin of hematopoietic stem cells: Progress and controversy. *Stem Cell Res.* 2012;8(1):1–13.

2. Jacobson L, Simmons EL, Marks EK, Eldredge JH. Recovery from Radiation Injury. *Science*. 1951;113:510–511.

3. Lorenz E, Uphoff D, Reid T, Shelton E. Modification of Irradiation Injury in Mice and Guinea Pigs by Bone Marrow Injections. *JNCI*. 1951;12(1):197–201.

4. Urso P, Congdon C. The Effect of the Amount of Isologous Bone Marrow Injected on the Recovery of Hematopoietic Organs, Survival and Body Weight after Lethal Irradiation Injury in Mice. *Blood*. 1957;12(3):251–260.

5. Iscove N, Till J, McCulloch E. The Proliferative States of Mouse Granulopoietic Progenitor Cells. *Exp. Biol. Med.* 1970;134(1):33–36.

6. Eaves C. Hematopoietic stem cells: concepts, definitions, and the new reality. *Blood*. 2015;125(17):2605–2614.

7. Thomas ED, Lochte HL, Cannon JH, Sahler OD, Ferrebee JW. Supralethal whole body irradiation and isologous marrow transplantation in man. *J. Clin. Invest.* 1959;38(10):1709–1716.

8. Thomas E. A history of haematopoietic cell transplantation. *Br. J. Haematol.* 1999;105(3):330–339.

9. Passweg JR, Baldomero H, Bader P, et al. Hematopoietic stem cell transplantation in Europe 2014: More than 40 000 transplants annually. *Bone Marrow Transplant.* 2016;51(6):786–792.

10. Julius MH, Masuda T, Herzenberg LA. Demonstration That Antigen-Binding Cells Are Precursors of Antibody-Producing Cells After Purification with a Fluorescence-Activated Cell

Sorter. *Proc. Natl. Acad. Sci.* 1972;69(7):1934–1938.

11. Köhler G, Milstein C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature*. 1975;256(5517):495–497.

12. Wilson A, Laurenti E, Oser G, et al. Hematopoietic Stem Cells Reversibly Switch from Dormancy to Self-Renewal during Homeostasis and Repair. *Cell*. 2008;135(6):1118–1129.

13. Orkin SH, Zon LI. Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. *Cell*. 2008;132(4):631–644.

14. Doulatov S, Notta F, Laurenti E, Dick JE. Hematopoiesis: A human perspective. *Cell Stem Cell*. 2012;10(2):120–136.

15. Nimmo RA, May GE, Enver T. Primed and ready: Understanding lineage commitment through single cell analysis. *Trends Cell Biol.* 2015;25(8):459–467.

16. Morrison SJ, Weissman IL. The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity*. 1994;1(8):661–673.

17. Sun J, Ramos A, Chapman B, et al. Clonal dynamics of native haematopoiesis. *Nature*. 2014;514(7522):322–327.

18. Pei W, Feyerabend TB, Rössler J, et al. Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature*. 2017;548(7668):456–460.

19. Rodriguez-Fraticelli AE, Wolock SL, Weinreb CS, et al. Clonal analysis of lineage fate in native haematopoiesis. *Nature*. 2018;553(7687):212–216.

20. Laurenti E, Göttgens B. From haematopoietic stem cells to complex differentiation landscapes. *Nature*. 2018;553(7689):418–426.

21. Ema H, Morita Y, Suda T. Heterogeneity and hierarchy of hematopoietic stem cells. *Exp. Hematol.* 2014;42(2):74–82.

22. Hu M, Krause D, Sharkies S, et al. Multilineage gene expression preceded commitment in the hemopoietic system. *Genes Dev.* 1997;11:774–785.

23. Svensson V, Vento-Tormo R, Teichmann SA. Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* 2018;13(4):599–604.

24. Macosko EZ, Basu A, Satija R, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell.* 2015;161(5):1202–1214.

25. Klein AM, Mazutis L, Akartuna I, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell.* 2015;161(5):1187–1201.

26. Zheng GXY, Terry JM, Belgrader P, et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 2017;8(14049):doi:10.1038/ncomms14049.

27. Picelli S, Faridani OR, Björklund ÅK, et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 2014;9(1):171–181.

28. Hashimshony T, Senderovich N, Avital G, et al. CEL-Seq2: Sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* 2016;17(1):1–7.

29. Bagnoli JW, Ziegenhain C, Janjic A, et al. Sensitive and powerful single-cell RNA sequencing using mcSCRB-seq. *Nat. Commun.* 2018;9(2937):doi:10.1038/s41467-018-05347-6.

30. Hayashi T, Ozaki H, Sasagawa Y, et al. Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs. *Nat. Commun.* 2018;9(619):doi:10.1038/s41467-018-02866-0.

31. Svensson V, Natarajan KN, Ly LH, et al. Power analysis of single-cell rnA-sequencing experiments. *Nat. Methods.* 2017;14(4):381–387.

32. Ziegenhain C, Vieth B, Parekh S, et al. Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol. Cell.* 2017;65(4):631–643.e4.

33. Trapnell C, Cacchiarelli D, Grimsby J, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 2014;32(4):381–386.

34. Bendall SC, Davis KL, Amir EAD, et al. Single-cell trajectory detection uncovers progression and regulatory coordination in human b cell development. *Cell*. 2014;157(3):714–725.

35. Weinreb C, Wolock S, Tusi BK, Socolovsky M, Klein AM. Fundamental limits on dynamic inference from single-cell snapshots. *Proc. Natl. Acad. Sci.* 2018;115(10):2467–2476.

36. Paul F, Arkin Y, Giladi A, et al. Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell*. 2015;163(7):1663–1677.

37. Kowalczyk MS, Tirosh I, Heckl D, et al. Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.* 2015;25(12):1860–1872.

38. 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.

39. Olsson A, Venkatasubramanian M, Chaudhri VK, et al. Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. *Nature*. 2016;537(7622):698–702.

40. Dahlin JS, Hamey FK, Pijuan-Sala B, et al. A single-cell hematopoietic landscape resolves 8 lineage trajectories and defects in Kit mutant mice. *Blood*. 2018;131(21):e1–e11.

41. Tusi BK, Wolock SL, Weinreb C, et al. Population snapshots predict early haematopoietic and erythroid hierarchies. *Nature*. 2018;555(7694):54–60.

42. Regev A, Teichmann S, Lander S, et al. The Human Cell Atlas. *Elife*. 2017;6:1–30.

43. Han X, Wang R, Zhou Y, et al. Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell*.

2018;172(5):1091–1107.e17.

44. Hay S, Ferchen K, Chetal K, Grimes HL, Salomonis N. Human Cell Atlas bone marrow single-cell interactive web portal. *Exp. Hematol.* 2018;In Press:

45. Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat. Rev. Immunol.* 2018;18(1):35–45.

46. Jaitin D, Kenigsberg E, Keren-Shaul H, et al. Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types. *Science*. 2014;343(February):776–779.

47. Villani A-C, Satija R, Reynolds G, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*. 2017;356(6335):eaah4573.

48. Yáñez A, Coetzee SG, Olsson A, et al. Granulocyte-Monocyte Progenitors and Monocyte-Dendritic Cell Progenitors Independently Produce Functionally Distinct Monocytes. *Immunity*. 2017;47(5):890–902.e4.

49. Menezes S, Melandri D, Anselmi G, et al. The Heterogeneity of Ly6Chi Monocytes Controls Their Differentiation into iNOS+ Macrophages or Monocyte-Derived Dendritic Cells. *Immunity*. 2016;45(6):1205–1218.

50. Adolfsson J, Månsson R, Buza-Vidas N, et al. Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential: A revised road map for adult blood lineage commitment. *Cell*. 2005;121(2):295–306.

51. Görgens A, Radtke S, Möllmann M, et al. Revision of the Human Hematopoietic Tree: Granulocyte Subtypes Derive from Distinct Hematopoietic Lineages. *Cell Rep.* 2013;3(5):1539–1552.

52. Pronk CJH, Rossi DJ, Månsson R, et al. Elucidation of the Phenotypic, Functional, and Molecular Topography of a Myeloerythroid Progenitor Cell Hierarchy. *Cell Stem Cell*.

2007;1(4):428–442.

53. Alberti-Servera L, von Muenchow L, Tsapogas P, et al. Single-cell RNA sequencing reveals developmental heterogeneity among early lymphoid progenitors. *EMBO J.* 2017;36(24):3619–3633.

54. Yamamoto R, Morita Y, Ooehara J, et al. Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. *Cell*. 2013;154(5):1112–1126.

55. Notta F, Zandi S, Takayama N, et al. Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science*. 2016;351(6269):aab2116.

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

57. Akashi K, Traver D, Miyamoto T. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature*. 2000;404(March):193–197.

58. Oguro H, Ding L, Morrison SJ. SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. *Cell Stem Cell*. 2013;13(1):102–116.

59. Bock C, Beerman I, Lien WH, et al. DNA Methylation Dynamics during In Vivo Differentiation of Blood and Skin Stem Cells. *Mol. Cell*. 2012;47(4):633–647.

60. Gazit R, Garrison BS, Rao TN, et al. Transcriptome analysis identifies regulators of hematopoietic stem and progenitor cells. *Stem Cell Reports*. 2013;1(3):266–280.

61. 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.

529 62. Pietras EM, Reynaud D, Kang YA, et al. Functionally Distinct Subsets of Lineage-Biased
530 Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions.
531 *Cell Stem Cell*. 2015;17(1):35–46.

532 63. Nestorowa S, Hamey FK, Pijuan Sala B, et al. A single-cell resolution map of mouse
533 hematopoietic stem and progenitor cell differentiation. *Blood*. 2016;128(8):e20–e31.

534 64. Kiel MJ, Yilmaz ÖH, Iwashita T, et al. SLAM family receptors distinguish hematopoietic stem
535 and progenitor cells and reveal endothelial niches for stem cells. *Cell*. 2005;121(7):1109–
536 1121.

537 65. Kent DG, Copley MR, Benz C, et al. Prospective isolation and molecular characterization of
538 hematopoietic stem cells with durable self-renewal potential. *Blood*. 2009;113(25):6342–
539 6350.

540 66. Muller-Sieburg CE, Sieburg HB. The GOD of hematopoietic stem cells: a clonal diversity model
541 of the stem cell compartment. *Cell Cycle*. 2006;5(4):394–398.

542 67. Dykstra B, Kent D, Bowie M, et al. Long-Term Propagation of Distinct Hematopoietic
543 Differentiation Programs In Vivo. *Cell Stem Cell*. 2007;1(2):218–229.

544 68. Upadhaya S, Reizis B, Sawai CM. New genetic tools for the in vivo study of hematopoietic
545 stem cell function. *Exp. Hematol.* 2018;61:26–35.

546 69. Cabezas-Wallscheid N, Buettner F, Sommerkamp P, et al. Vitamin A-Retinoic Acid Signaling
547 Regulates Hematopoietic Stem Cell Dormancy. *Cell*. 2017;169(5):807–823.e19.

548 70. Velten L, Haas SF, Raffel S, et al. Human haematopoietic stem cell lineage commitment is a
549 continuous process. *Nat. Cell Biol.* 2017;19(4):271–281.

550 71. Grover A, Sanjuan-Pla A, Thongjuea S, et al. Single-cell RNA sequencing reveals molecular and
551 functional platelet bias of aged haematopoietic stem cells. *Nat. Commun.*

2016;7(11075):doi:10.1038/ncomms11075.

72. Sanjuan-Pla A, Macaulay IC, Jensen CT, et al. Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature*. 2013;502(7470):232–236.

73. Carrelha J, Meng Y, Kettyle LM, et al. Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. *Nature*. 2018;554(7690):106–111.

74. Muller CE, Cho RH, Thoman M, Adkins B, Sieburg HB. Deterministic regulation of hematopoietic stem cell self-renewal and differentiation. *Blood*. 2002;100(4):1302–1309.

75. Benz C, Copley MR, Kent DG, et al. Hematopoietic stem cell subtypes expand differentially during development and display distinct lymphopoietic programs. *Cell Stem Cell*. 2012;10(3):273–283.

76. Mann M, Mehta A, Boer C De, et al. Heterogeneous Responses of Hematopoietic Stem Cells to Inflammatory. *bioRxiv*. 2017;doi:10.1101/163402.

77. Pietras EM, Mirantes-Barbeito C, Fong S, et al. Chronic interleukin-1 exposure drives haematopoietic stem cells towards precocious myeloid differentiation at the expense of self-renewal. *Nat. Cell Biol.* 2016;18(6):607–618.

78. Pang WW, Price E a, Sahoo D, et al. Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. *Proc. Natl. Acad. Sci.* 2011;108(50):20012–20017.

79. Van Der Maaten LJP, Hinton GE. Visualizing high-dimensional data using t-sne. *J. Mach. Learn. Res.* 2008;9:2579–2605.

80. Haghverdi L, Buettner F, Theis FJ. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics*. 2015;31(18):2989–2998.

81. Weinreb C, Wolock S, Klein AM. SPRING: A kinetic interface for visualizing high dimensional

single-cell expression data. *Bioinformatics*. 2018;34(7):1246–1248.

82. McInnes L, Healy J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arxiv*. 2018;doi:1802.03426v1.

83. Ocone A, Haghverdi L, Mueller NS, Theis FJ. Reconstructing gene regulatory dynamics from high-dimensional single-cell snapshot data. *Bioinformatics*. 2015;31(12):i89–i96.

84. Marco E, Karp RL, Guo G, et al. Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. *Proc. Natl. Acad. Sci.* 2014;111(52):5643–5650.

85. Setty M, Tadmor MD, Reich-zeliger S, et al. Articles Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nat. Biotechnol.* 2016;34(6):614–637.

86. Shin J, Berg DA, Christian KM, et al. Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades underlying Adult Neurogenesis Resource Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades underlying Adult Neurogenesis. *Cell Stem Cell*. 2015;17(3):360–372.

87. Haghverdi L, Büttner M, Wolf FA, Buettner F, Theis FJ. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods*. 2016;13(10):845–848.

88. Welch JD, Hartemink AJ, Prins JF. SLICER: Inferring branched, nonlinear cellular trajectories from single cell RNA-seq data. *Genome Biol.* 2016;17(1):1–15.

89. Giecold G, Marco E, Garcia SP, Trippa L, Yuan GC. Robust lineage reconstruction from high-dimensional single-cell data. *Nucleic Acids Res.* 2016;44(14):1–7.

90. Grün D, Muraro MJ, Boisset JC, et al. De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data. *Cell Stem Cell*. 2016;19(2):266–277.

91. Wolf FA, Hamey F, Plass M, et al. Graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *bioRxiv*. 2017;doi:10.1101/208819.

92. Herman JS, Sagar, Grün D. FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data. *Nat. Methods*. 2018;15(5):379–386.

93. Lummertz E, Rowe RG, Lundin V, et al. Reconstruction of complex single-cell trajectories using CellRouter. *Nat. Commun.* 2018;9(892):doi:10.1038/s41467-018-03214-y.

94. Tusi BK, Socolovsky M. High-throughput single-cell fate potential assay of murine hematopoietic progenitors in vitro. *Exp. Hematol.* 2018;60:21–29.e3.

95. Rosenberg A, Roco C, Muscat R, et al. Single-cell profiling of the developing mouse brain and spinal cord wit split-pool barcoding. *Science*. 2018;doi: 10.1126/science.aam8999.

96. Stoeckius M, Hafemeister C, Stephenson W, et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods*. 2017;14(9):865–868.

97. Peterson VM, Zhang KX, Kumar N, et al. Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* 2017;35(10):936–939.

98. Buenrostro JD, Corces MR, Lareau CA, et al. Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell*. 2018;173(6):1535–1548.e16.

99. Lay F, Kelly T, Jones P. Nucleosome Occupancy and Methylome Sequencing (NOMe-seq). *Methods Mol. Biol.* 2018;3rd editio:267–284.

100. Cusanovich DA, Daza R, Adey A, et al. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science*. 2015;348(6237):910–914.

101. Clark SJ, Argelaguet R, Kapourani CA, et al. ScNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat. Commun.* 2018;9(781):10.1038/s41467-018-03149-4.

102. Crosetto N, Bienko M, Van Oudenaarden A. Spatially resolved transcriptomics and beyond.

Nat. Rev. Genet. 2015;16(1):57–66.

103. Wang X, Allen WE, Wright MA, et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science*. 2018;361(6400):10.1126/science.aat5691.

104. Schiebinger G, Shu J, Tabaka M, et al. Reconstruction of developmental landscapes by optimal-transport analysis of single-cell gene expression sheds light on cellular reprogramming. *bioRxiv*. 2017;doi:10.1101/191056.

105. Wagner DE, Weinreb C, Collins ZM, et al. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science*. 2018;360(6392):981–987.

106. Raj B, Wagner DE, McKenna A, et al. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* 2018;36(5):442–450.

107. Spanjaard B, Hu B, Mitic N, et al. Simultaneous lineage tracing and cell-type identification using CrIsPr-Cas9-induced genetic scars. *Nat. Biotechnol.* 2018;36(5):469–473.

108. Akinduro O, Weber TS, Ang H, et al. Proliferation dynamics of acute myeloid leukaemia and haematopoietic progenitors competing for bone marrow space. *Nat. Commun.* 2018;9(519):doi:10.1038/s41467-017-02376-5.

109. Dixit A, Parnas O, Li B, et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell*. 2016;167(7):1853–1866.e17.

110. Adamson B, Norman TM, Jost M, et al. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell*. 2016;167(7):1867–1882.e21.

111. Datlinger P, Rendeiro AF, Schmidl C, et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods*. 2017;14(3):297–301.

112. Jaitin DA, Weiner A, Yofe I, et al. Dissecting Immune Circuits by Linking CRISPR-Pooled

644 Screens with Single-Cell RNA-Seq. *Cell*. 2016;167(7):1883–1896.e15.

645 113. Giladi A, Paul F, Herzog Y, et al. Single-cell characterization of haematopoietic progenitors and
646 their trajectories in homeostasis and perturbed haematopoiesis. *Nat. Cell Biol.* 2018;20:836–
647 846.

648 114. Xie S, Duan J, Li B, Zhou P, Hon GC. Multiplexed Engineering and Analysis of Combinatorial
649 Enhancer Activity in Single Cells. *Mol. Cell*. 2017;66(2):285–299.e5.

650 115. Giustacchini A, Thongjuea S, Barkas N, et al. Single-cell transcriptomics uncovers distinct
651 molecular signatures of stem cells in chronic myeloid leukemia. *Nat. Med.* 2017;23(6):692–
652 702.

653 116. Gierahn TM, Wadsworth MH, Hughes TK, et al. Seq-Well: Portable, low-cost rna sequencing
654 of single cells at high throughput. *Nat. Methods*. 2017;14(4):395–398.

655 117. Zheng S, Papalexi E, Butler A, Stephenson W, Satija R. Molecular transitions in early
656 progenitors during human cord blood hematopoiesis. *Mol. Syst. Biol.* 2018;14(3):e8041.

657 118. The Tabula Muris Consortium. Single-cell transcriptomics of 20 mouse organs creates a
658 Tabula Muris. *Nature*. 2018;562:367–372.

659 119. Zhao X, Gao S, Wu Z, et al. Single-cell RNA-seq reveals a distinct transcriptome signature of
660 aneuploid hematopoietic cells. *Blood*. 2017;130(25):2762–2773.

661

662

663

FIGURE LEGENDS

Figure 1. Comparison of a haematopoietic tree diagram with a single-cell transcriptomic landscape. (A) Schematic showing one of the classic views of the haematopoietic cell hierarchy. Dashed boxes show three compartments encompassing cells of different potency: multipotent cells on top, bi/oligopotent cells in the middle and terminally differentiated (unipotent) cells at the bottom. (B) A dimensionality reduction projection (UMAP algorithm) of single cell transcriptomes from the bone marrow mononuclear cell fraction. Arrows indicate main directions of differentiation, inferred from analysis of typical marker genes. Grey indicates 'unassigned' cells, which identity based on markers is unclear (dataset downloaded from HCA data portal and processed by I.K). LT-HSC, long-term haematopoietic stem cell; ST-HSC, short-term haematopoietic stem cell; MPP, multipotent progenitor; LMPP, lymphoid-primed multipotent progenitor; CMP, common myeloid progenitor; MEP, megakaryocyte-erythroid progenitor; GMP, granulocyte-monocyte progenitor; CLP, common lymphoid progenitor; HSPC, haematopoietic stem and progenitor cell; Mk, megakaryocyte.

Figure 2. Distances and data dimensionality. (A) A set of single cells expressing 3 genes arranged along a curved shape has been simulated. There are two measures of distance between the blue and red cells: Whilst D_1 represents the shortest possible distance between the two cells, D_2 is the distance between the cells through the structure of the data ('manifold'). The two arms of the curved shape may represent continuous transition processes (e.g. cell differentiation) thus distance D_2 is the important distance measure. A dimensionality reduction technique (here tSNE) should capture such features. (B) Excessive reduction in dimensionality causes important information to be lost. In this case, a 2D representation of the data incorrectly suggests that the green cell is further from the yellow cell than the orange cell, because information has been lost about axis 2. (C) To infer cellular trajectories from scRNA-seq data, dimensionality reduction is used to 'learn' the structure of the data ('learned data'), which captures the important distances between cells in a suitable number

of dimensions, typically 10-100. Trajectory inference can then be attempted from this learned data. For visualisation, the dimensionality of the data needs to be reduced to either 2 or 3, but this will inevitably lose some of the important biological information rendering data unsuitable for trajectory inference.

Figure 3. From transcriptomic landscapes to tissue function. The three diagrams depict 2D landscapes with single cells as points, where a pool of progenitors gives rise to two differentiated populations X and Y. (A) A classical approach to identify a specific bipotent progenitor population, an immunophenotypic gate drawn based on expression of a surface marker (green) sections through the landscape and yields a heterogeneous population of cells upon functional analysis. (B) Normal haematopoiesis; dynamic information about cellular fluxes (arrows) indicates directions of differentiation throughout the landscape and degree of self-renewal. (C) An example of how cell flux analysis reveals mechanisms of a disease. In this case the stem cell pool is exhausted (low self-renewal) compromising production of Y cells, while there is an increased production of X cells due to acquired increased self-renewal of progenitors X.

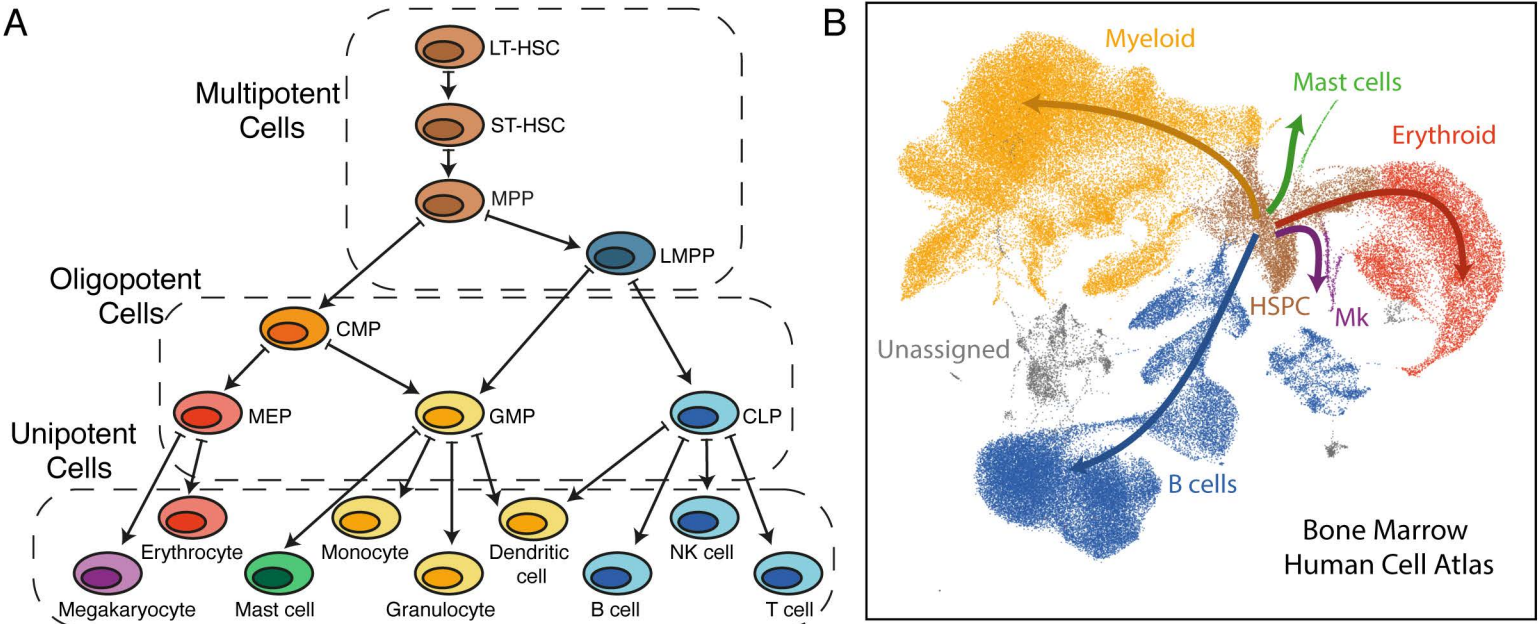
A
Multipotent Cells
LT-HSC
ST-HSC
MPP
LMPP
Oligopotent Cells
CMP
MEP
GMP
CLP
Unipotent Cells
Erythrocyte
Megakaryocyte
Mast cell
Monocyte
Granulocyte
Dendritic cell
B cell
NK cell
T cell
B
Myeloid
Mast cells
Erythroid
HSPC
Mk
Unassigned
B cells
Bone Marrow
Human Cell Atlas

Figure 1

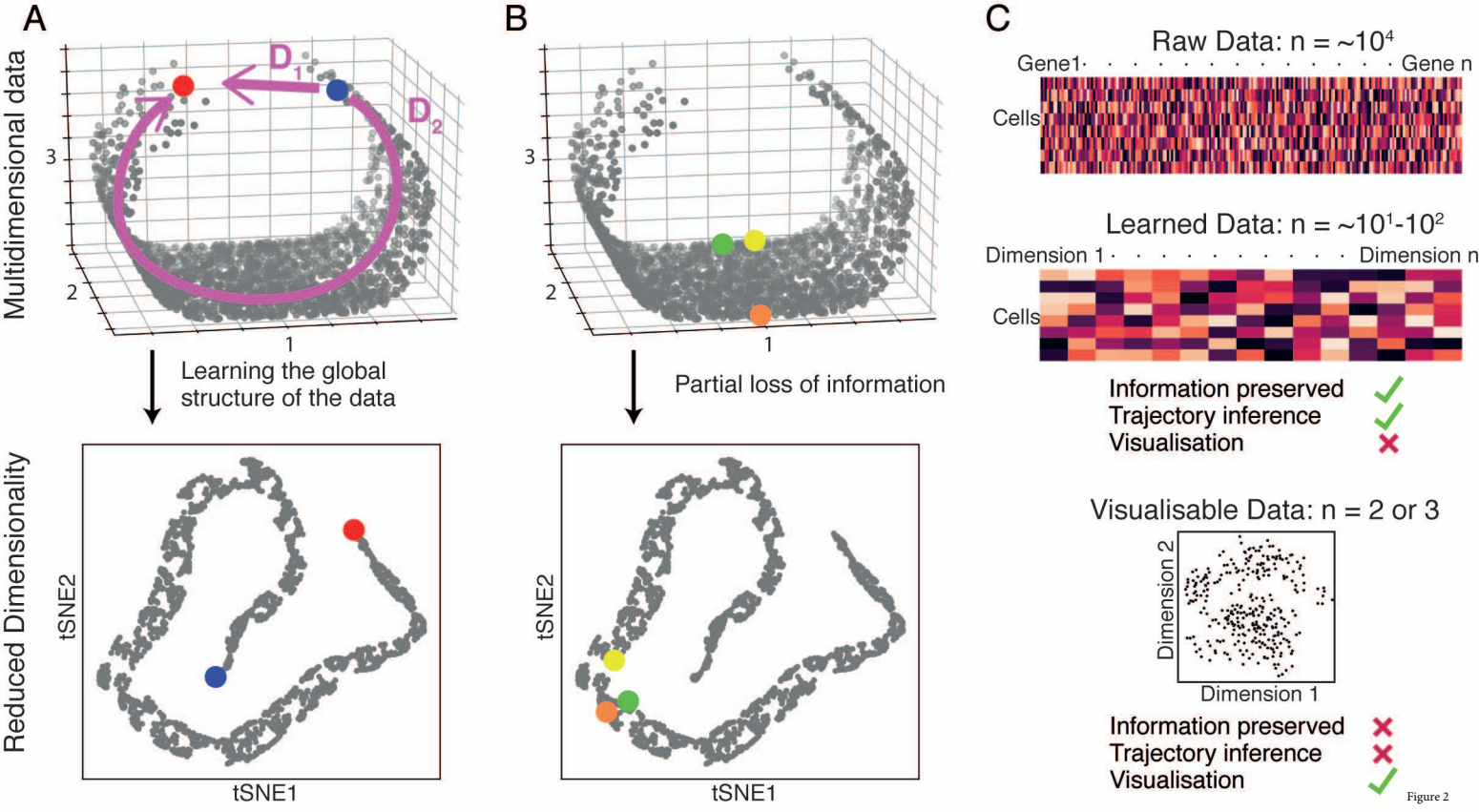
A
B
C
Multidimensional data
Reduced Dimensionality
D1
D2
3
2
1
Learning the global structure of the data
Partial loss of information
tSNE1
tSNE2
Raw Data: n = ~10^4
Gene1
Gene n
Cells
Learned Data: n = ~10^1-10^2
Dimension 1
Dimension n
Cells
Information preserved
Trajectory inference
Visualisation
Visualisable Data: n = 2 or 3
Dimension 2
Dimension 1
Information preserved
Trajectory inference
Visualisation

Figure 2

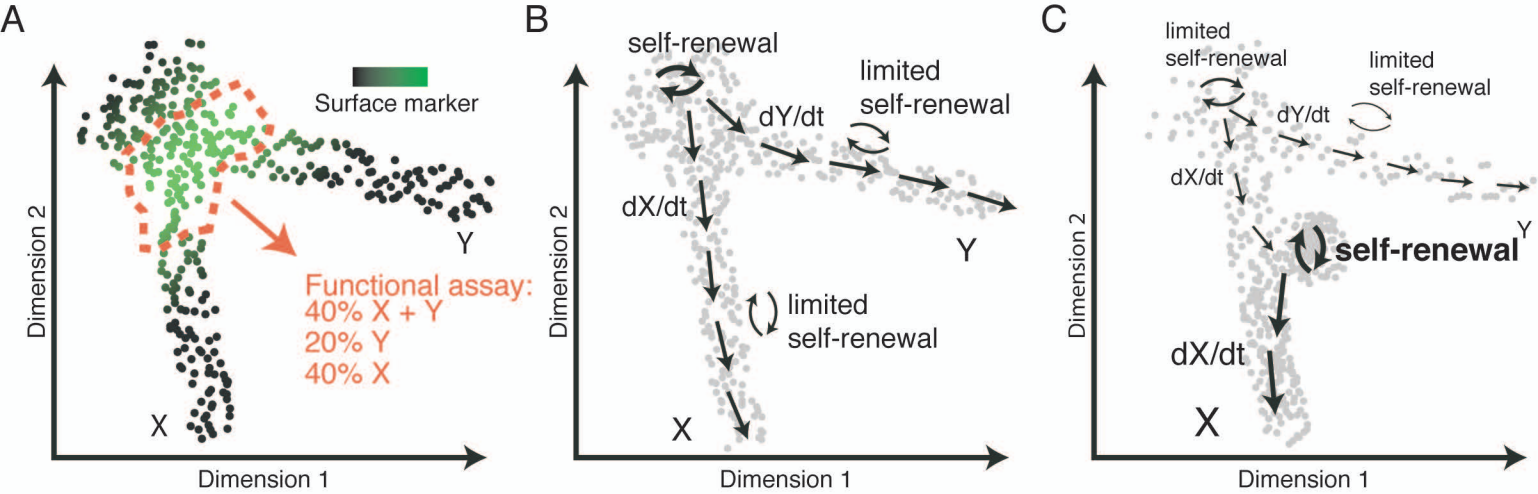
A
Surface marker
Dimension 2
Dimension 1
Y
X
Functional assay:
40% X + Y
20% Y
40% X
Isolation of bipotent progenitors
B
self-renewal
dY/dt
limited
self-renewal
dX/dt
Y
limited
self-renewal
X
Dimension 2
Dimension 1
Cellular flux through the landscape
native conditions
C
limited
self-renewal
limited
self-renewal
dY/dt
dX/dt
self-renewal
Y
dX/dt
X
Dimension 2
Dimension 1
Cellular flux through the landscape
disease scenario

Figure 3