

Annals of Stem Cell Research

Review Article

Single-Cell RNA Sequencing of Haematopoietic Progenitor Heterogeneity in Human Bone Marrow

Bizjak N^{1*}, Barbu A² and Szymanski A³

¹Department of Biomedical Engineering, University of Ljubljana, Ljubljana, Slovenia

²Department of Experimental Oncology, Carol Davila University of Medicine and Pharmacy, Bucharest, Romania

³Department of Clinical Immunology, Jagiellonian University, Krakow, Poland

***Address for Correspondence:** Nina Bizjak, Department of Biomedical Engineering, University of Ljubljana, Ljubljana, Slovenia; E-mail: nina.bizjak@universityofljubljana.edu

Received: 4 February 2025; **Accepted:** 20 May 2025; **Published:** 31 August 2025; **DOI:** 10.5281/zenodo.22274651

Citation of this article: Bizjak N, Barbu A, Szymanski A (2025) Single-Cell RNA Sequencing of Haematopoietic Progenitor Heterogeneity in Human Bone Marrow. Ann Stem Cell Res, 8(1): 01-08.

Copyright: © 2025 Nina Bizjak, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Human haematopoietic progenitors were long mapped as discrete marker-defined compartments. Single-cell RNA sequencing has replaced that tree with a continuous transcriptional landscape.

Objective: To assess how securely the continuum interpretation holds against technical artefact.

Methods: Narrative review of index-sorted and droplet single-cell studies, reference atlases and clonal-tracing work.

Findings: Index-sorted profiling of 1,413 cells from two donors showed a continuum rather than discrete oligopotent stages, and clonal marks independent of expression agreed. Atlases detect only 525 to 880 genes per cell, while ambient RNA at 2 to 6% and multiplets at 1.6 to 3.1% push data toward apparent continuity.

Conclusions: A continuous progenitor landscape is supported; finer substructure within it is not, since noise of comparable magnitude can generate it.

Keywords: Haematopoietic progenitor cells; Single-cell transcriptomics; Bone marrow; Cellular heterogeneity; Clonal lineage tracing

Introduction

For three decades the human blood system was drawn as a tree of discrete compartments, each defined by a surface-marker gate and validated by the colonies its cells produced. Single-cell RNA sequencing did not add branches to that tree; it dissolved most of them. The progenitor compartment now reads as a continuous transcriptional landscape in which lineage priming accumulates gradually, and the sharp intermediate populations of the older maps appear as regions of a cloud rather than as cell types [1,2].

That reinterpretation rests on measurements with a specific and uncomfortable property: the transcriptional differences used to argue for continuity are of the same magnitude as the technical artefacts of the assays that reveal them. Ambient RNA, cell multiplets and dissociation-induced stress each affect a few percent of a droplet dataset, and a few percent of intermediate-looking cells is exactly what a continuum claim needs to be robust to [3,4]. The literature is therefore best read as two intertwined stories: a genuine and well-replicated collapse of the discrete hierarchy, and a slowly improving understanding of how finely heterogeneity can be resolved before noise sets the floor.

This review traces both. It covers the compartment map and the assays that built it, the single-cell studies that displaced it, the reference atlases that traded depth for scale, the measured magnitudes of the principal artefacts, and the clonal and genealogical work that tests transcriptional continuity against cell behaviour in vivo.

The compartment map and how it was built

The compartment scheme was constructed by prospective isolation. Akashi and colleagues identified a clonogenic common myeloid progenitor in mouse bone marrow that gave rise to all myeloid lineages [5], and the human counterpart was isolated two years later on a Lin-CD34+CD38+IL-3Ralpha-lo CD45RA- gate [6].

Doulatov and colleagues then revised the human map, placing the origin of macrophages and dendritic cells in the early lymphoid branch rather than the myeloid one, which already indicated that the gates were separating function imperfectly [7].

The strongest form of the argument came from single-cell transplantation. Notta and colleagues purified single human haematopoietic stem cells capable of long-term multilineage engraftment using CD49f, showing that a marker-defined population could be resolved to a functionally homogeneous, self-renewing unit [8]. But the same laboratory later found that the routes of lineage development are reshaped across ontogeny, with the adult marrow behaving as a two-tier hierarchy in which multipotent cells give rise directly to unipotent progenitors and the classical oligopotent intermediates are largely absent [9].

Two limitations of that programme matter for what followed. Marker gates are defined by a handful of proteins and validated by an assay — colony formation or xenotransplantation — that reports what a cell can do under stress, not what it does at steady state. A gate can therefore be reproducible, clonogenic and still enclose a mixture. And because each new subdivision required a new marker, the resolution of the map was bounded by the number of parameters a cytometer could read, which is why the hierarchy grew by adding tiers rather than by dissolving them.

The transcriptional evidence pointed the same way. In mouse myeloid progenitors, Paul and colleagues found that a marker-defined compartment contains cells already committed to distinct downstream fates, with transcriptional heterogeneity preceding any change in immunophenotype [10]. By the time the human single-cell studies arrived, the question was no longer whether the gates were pure, but whether the compartments they were meant to capture existed at all.

Single-cell transcriptomes dissolve the compartments

Velten and colleagues answered directly. Index-sorted Lin-CD34+ cells from two healthy adult donors were profiled by a modified Smart-seq2 protocol, with 379 cells from the first individual and 1,034 from the second passing quality control — 1,413 in total [1]. Because every surface marker used for sorting was recorded per cell, transcriptome and immunophenotype could be aligned directly. The result was a continuum of low-primed undifferentiated cells rather than a set of discrete oligopotent stages: lineage priming appeared as a gradient acquired early and continuously, and the CMP, GMP and MLP gates cut across it rather than enclosing it. The protocol choice mattered, since the modified Smart-seq2 approach strongly decreased the dropout rate relative to QUARTZ-Seq, and dropout is precisely what makes weakly primed cells look alike [1].

Karamitros and colleagues reached a compatible conclusion for the lympho-myeloid branch, reporting that LMPP, GMP and MLP gates each contain functionally and transcriptionally heterogeneous cells and that a continuum, not a series of compartments, executes lympho-myeloid differentiation [11]. Belluschi and colleagues sharpened the timing using index-sorted single cells: 435 HSC/MPP-pool cells from six cord blood donors and 819 CD49f+ HSCs from four independent cord bloods were assayed in single-cell culture, and 192 single CD49f+ HSCs were taken for scRNA-seq with 88% passing QC [12]. Fewer than 1% of cells produced quadrilineage colonies, and roughly 10% of multilineage colonies were myeloid/NK/erythroid, placing myelo-lymphoid restriction inside the HSC compartment itself, upstream of the LMPP stage the hierarchy assigned it to.

Pellin and colleagues extended the picture to droplet scale, sequencing 6,011 single CD34+ cells and 15,401 single Lin- cells from healthy adult bone marrow on the

inDrops platform, resolving 11 transcriptional groups in the CD34+ map and 15 in the Lin- map [13]. The clusters were continuous rather than separated, and priming toward single lineages appeared early. Chromatin accessibility gave an independent readout of the same geometry, with integrated single-cell analysis mapping a continuous regulatory landscape across human haematopoietic differentiation [14]. Trajectory methods then formalised the continuum by assigning each cell a probability of reaching each terminal fate rather than a discrete label [15], and the field's synthesis moved from hierarchies to landscapes [2].

Table 1: Major single-cell landscape studies of human haematopoietic progenitors, by cells profiled, donors, platform and headline conclusion [1,13,16,17,20,32].			
Study	Cells / donors	Platform	Headline conclusion
velten2017	379 + 1,034; 2 adults	Index-sorted Smart-seq2	Priming is continuous, not oligopotent
pellin2019	6,011 CD34+; 15,401 Lin-	inDrops	Continuous landscape, early unipotent priming
oetjen2018	76,645; 20 donors 24-84 y	10x 3' v2	Donor-diverse marrow baseline, 880 genes/cell
hay2018	>100,000; 8 donors	10x, HCA census	35 coherent groups in a public reference
ainciburu2023	34,590 young; 40,641 elderly	10x v3.1, 30k reads/cell	Same continuum, age-perturbed
weng2024	7,104 CD34+; 4 donors 26-78 y	ReDeeM mtDNA multi-omic	78 clonal groups among 5,393 HSCs

Reference atlases and the depth-scale trade-off

The studies that dissolved the compartments were small and deep. The studies that made human marrow a reference tissue are large and shallow, and the difference is not incidental.

Oetjen and colleagues profiled bone marrow from 20 healthy volunteers aged 24 to 84 years, median 57, evenly

split by sex, capturing more than 90,000 cells and retaining 76,645 after filtering at 500 genes per cell and 8% mitochondrial reads [16]. The mean number of genes detected per cell was 880, with a per-sample range of 575 to 1,390. That figure is the honest measure of what a droplet atlas sees: at roughly 20,000 reads per cell, the founding 10x characterisation reported a median of 525 genes and 1,300 UMIs per cell across a 68,000-cell peripheral blood dataset, within a study spanning some 250,000 cells across 29 samples [3]. Cluster boundaries in data of this depth are soft because most of the transcriptome is unobserved, not because the underlying states are necessarily continuous.

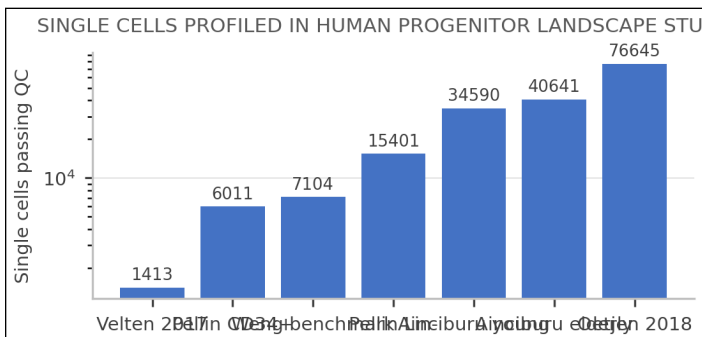


Figure 1: Single cells passing quality control in seven human haematopoietic landscape datasets, spanning 1,413 index-sorted cells from 2 donors to 76,645 marrow cells from N = 20 donors [1,13,32,20,16].

The trade-off is systematic. The datasets that displaced the discrete hierarchy profiled 1,413 cells with a protocol chosen to suppress dropout [1]; the datasets that serve as population references profile tens of thousands of cells at a depth where a gene expressed in a minority of a cluster is likely to be missed entirely. Neither design alone settles a question about substructure: the deep studies lack the cell numbers to detect rare states, and the shallow studies lack the per-cell information to separate adjacent ones.

The Human Cell Atlas bone marrow census assembled more than 100,000 immune cells from eight healthy donors into 35 transcriptionally coherent groups with an interactive portal [17], within the broader atlas

programme's aim of a comprehensive reference of human cell types [18]. Developmental references followed the same logic, with roughly 140,000 fetal liver cells plus about 74,000 skin, kidney and yolk sac cells profiled to decode prenatal haematopoiesis [19]. Ainciburu and colleagues applied the approach to ageing, profiling 34,590 CD34+ cells from five young donors aged 18 to 20 years and 40,641 from three donors over 65, on 10x NEXTGEM 3' v3.1 chemistry at an average of 30,000 reads per cell, resolving 14 subpopulations in the young dataset [20]. The elderly landscape was perturbed rather than restructured — the same continuum, differently weighted.

The technical noise floor

Ambient RNA is the first floor. In a barnyard experiment about 1% of observed transcripts were cross-species, a lower bound from which the true contamination rate was inferred at roughly 2%; in a peripheral blood dataset the same method estimated about 6% contamination automatically and about 5% manually from immunoglobulin genes [4]. Model-based background removal is now available as an alternative correction [21], but the magnitude to be corrected is set by the tissue and the protocol, not by the tool.

Multiplets are the second. In a 50:50 293T:Jurkat titration, roughly 50% of loaded cells were captured across some 100,000 droplet partitions per six-minute run, and the inferred multiplet rate was 1.6% at 1,012 recovered cells and 3.1% at approximately 3,000, rising linearly between 1,200 and 9,500 recovered cells [3]. Simulated-doublet detection identifies a substantial fraction of these computationally [22], but a doublet of a stem cell and a committed progenitor is, by construction, an artificial intermediate state — the exact object a continuum argument counts as evidence.

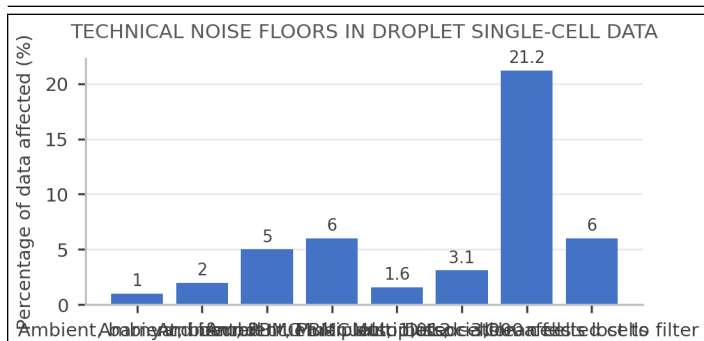


Figure 2: Measured magnitudes of ambient RNA, multiplet and dissociation artefacts, each of the same order as the intermediate states used to argue for a continuum; dissociation values are from a 665-cell CEL-Seq dataset and multiplet values from a 293T:Jurkat titration at N = 1,012 and ~3,000 recovered cells [4,3,23].

Dissociation is the third and the least tractable. In a CEL-Seq muscle stem cell dataset of 665 cells, 21.2% were classified as dissociation-affected, distinguished by 55 differentially expressed genes; filtering cells with at least 5.75% of transcripts from those genes removed all affected cells while discarding 6% of unaffected ones [23]. Comparing 37 °C collagenase with 6 °C cold-active protease across 155,165 cells, O'Flanagan and colleagues defined a core signature of 512 stress-response genes, 507 up and 5 down at 37 °C, and found 11,975 of 19,464 retained genes (62%) differentially expressed at an absolute log fold change of at least 1.5 [24]. Marrow aspiration and CD34 enrichment impose comparable handling.

These magnitudes are not independent of the biological claim. Ambient RNA transfers erythroid or myeloid transcripts into progenitor droplets, which reads as early priming; doublets place two transcriptomes at an intermediate coordinate, which reads as a transitional state; dissociation stress imposes a shared program across cells that would otherwise separate, which compresses the landscape. Each artefact pushes the data toward continuity rather than away from it, so the continuum interpretation is the one that a careless analysis would reach regardless of the biology.

Donor structure compounds all three. Mutual nearest

neighbours, Harmony and anchor-based integration each remove batch structure effectively [25,26,27], but each also removes real between-donor biology by the same operation, and no published percentage separates the two.

Clonal genealogies as an independent test

Because transcriptional continuity can be manufactured by noise, the strongest tests come from lineage marks that are independent of expression. Somatic mutations in single-cell-derived colonies gave the first population-scale estimate: 129,582 mutations across 140 colonies from one healthy 59-year-old man implied 50,000 to 200,000 HSCs actively producing white cells at any time [28]. Extending this across the lifespan, 3,579 whole genomes from colonies of ten subjects aged 0 to 81 years showed HSC/MPPs accumulating a mean of 17 mutations and losing 30 bp of telomere per year, with a stable pool of 20,000 to 200,000 contributing evenly below 65; above 75, 12 to 18 independent clones accounted for 30 to 60% of haematopoiesis, each contributing 1 to 34%, and only 22% of expanded clones carried a known driver [29]. Native murine barcoding had already shown that steady-state output is dominated by progenitors rather than by the transplantation-defined HSC [30].

Mitochondrial mutations allow genealogy and transcriptome to be read in the same cell [31]. Applied with enrichment to 7,104 CD34+ HSPCs from one healthy young donor, this yielded 4,831 high-confidence mtDNA mutations, a median of 9 per cell against 1 without enrichment [32]. Across four donors aged 26, 31, 76 and 78 years, with 11,009, 15,101, 9,519 and 14,715 haematopoietic cells respectively and 17 major clusters, 78 HSC clonal groups were resolved among 5,393 profiled HSCs, and 14 HSC subpopulations within a single young donor. Clonal identity aligned with position on the transcriptional landscape rather than with discrete compartments — continuity is a property of the cells, not

only of the assay. Multi-omic linkage of accessibility to expression, yielding 91,601 peak-to-gene links across healthy and leukaemic haematopoiesis, supplies the regulatory layer of the same map [33].

Table 2: Sources of technical noise in droplet marrow data, reported magnitude and correction method [4,21,3,22,23,24,25,26,27].		
Artefact	Reported magnitude	Correction
Ambient RNA, barnyard	1% observed, ~2% inferred	SoupX
Ambient RNA, PBMC	~6% auto, ~5% manual	SoupX; CellBender
Multiplets	1.6% at 1,012; 3.1% at ~3,000 cells	Simulated-doublet detection
Cell capture	~50% of loaded cells	None
Dissociation stress	21.2% of cells; 512-gene signature	5.75% filter; cold protease
Donor batch structure	No published percentage	MNN; Harmony; anchors

These clonal results also constrain what a transcriptional snapshot can mean. A pool of 20,000 to 200,000 HSCs contributing evenly [29] is far larger than the few thousand HSCs any single scRNA-seq experiment captures, so absence of a state from an atlas is weak evidence of its absence from the marrow, and the age-dependence of clonal structure means that donor age is a variable rather than a nuisance in every human progenitor dataset [20,32].

Conclusion

The discrete-compartment scheme did not survive single-cell measurement, and the replacement is a continuum on which lineage priming accumulates early and gradually. That conclusion is supported by index-sorted plate-based data, droplet landscapes, chromatin accessibility and, most importantly, by clonal marks that owe nothing to expression. What remains unsettled is resolution. Ambient RNA at 2 to 6%, multiplets at 1.6 to 3.1% and dissociation stress affecting a fifth of cells in an unfiltered dataset define a floor at the same scale as the intermediate states under discussion, and reference atlases detecting 525 to 880 genes per cell

cannot see beneath it. Finer substructure within the human progenitor cloud should be claimed only where ambient and doublet correction have been shown not to erase it.

References

1. Velten, L., Haas, S.F., Raffel, S., Blaszkiewicz, S., Islam, S., Hennig, B.P., et al. (2017) Human haematopoietic stem cell lineage commitment is a continuous process. *Nature Cell Biology*, 19(4): 271-281.
2. Laurenti, E., Gottgens, B. (2018) From haematopoietic stem cells to complex differentiation landscapes. *Nature*, 553(7689): 418-426.
3. Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., et al. (2017) Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, 8(1): 14049.
4. Young, M.D., Behjati, S. (2020) SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *GigaScience*, 9(12): giaa151.
5. Akashi, K., Traver, D., Miyamoto, T., Weissman, I.L. (2000) A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature*, 404(6774): 193-197.
6. Manz, M.G., Miyamoto, T., Akashi, K., Weissman, I.L. (2002) Prospective isolation of human clonogenic common myeloid progenitors. *Proceedings of the National Academy of Sciences*, 99(18): 11872-11877.
7. Doulatov, S., Notta, F., Eppert, K., Nguyen, L.T., Ohashi, P.S., Dick, J.E. (2010) Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nature Immunology*, 11(7): 585-593.
8. Notta, F., Doulatov, S., Laurenti, E., Poepl, A., Jurisica, I., Dick, J.E. (2011) Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science*, 333(6039): 218-221.

9. Notta, F., Zandi, S., Takayama, N., et al. (2016) Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science*, 351(6269): aab2116.
10. Paul, F., Arkin, Y., Giladi, A., Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., et al. (2015) Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Cell*, 163(7): 1663-1677.
11. Karamitros, D., Stoilova, B., Aboukhalil, Z., Hamey, F., Reinisch, A., Samitsch, M., et al. (2018) Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. *Nature Immunology*, 19(1): 85-97.
12. Belluschi, S., Calderbank, E.F., Ciauro, V., Pijuan-Sala, B., Santoro, A., Mende, N., et al. (2018) Myelo-lymphoid lineage restriction occurs in the human haematopoietic stem cell compartment before lymphoid-primed multipotent progenitors. *Nature Communications*, 9(1): 4100.
13. Pellin, D., Loperfido, M., Baricordi, C., Wolock, S.L., Montepeloso, A., Weinberg, O.K., et al. (2019) A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. *Nature Communications*, 10(1): 2395.
14. Buenrostro, J.D., Corces, M.R., Lareau, C.A., Wu, B., Schep, A.N., Aryee, M.J., et al. (2018) Integrated single-cell analysis maps the continuous regulatory landscape of human hematopoietic differentiation. *Cell*, 173(6): 1535-1548.e16.
15. Setty, M., Kisieliovas, V., Levine, J., Gayoso, A., Mazutis, L., Pe'er, D. (2019) Characterization of cell fate probabilities in single-cell data with Palantir. *Nature Biotechnology*, 37(4): 451-460.
16. Oetjen, K.A., Lindblad, K.E., Goswami, M., Gui, G., Dagur, P.K., Lai, C., et al. (2018) Human bone marrow assessment by single-cell RNA sequencing, mass cytometry, and flow cytometry. *JCI Insight*, 3(23): e124928.
17. Hay, S.B., Ferchen, K., Chetal, K., Grimes, H.L., Salomonis, N. (2018) The Human Cell Atlas bone marrow single-cell interactive web portal. *Experimental Hematology*, 68: 51-61.
18. Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., et al. (2017) The Human Cell Atlas. *eLife*, 6: e27041.
19. Popescu, D.-M., Botting, R.A., Stephenson, E., Green, K., Webb, S., Jardine, L., et al. (2019) Decoding human fetal liver haematopoiesis. *Nature*, 574(7778): 365-371.
20. Ainciburu, M., Ezponda, T., Berastegui, N., Alfonso-Pierola, A., Vilas-Zornoza, A., San Martin-Uriz, P., et al. (2023) Uncovering perturbations in human hematopoiesis associated with healthy aging and myeloid malignancies at single-cell resolution. *eLife*, 12: e79363.
21. Fleming, S.J., Chaffin, M.D., Arduini, A., Akkad, A.-D., Banks, E., Marioni, J.C., et al. (2023) Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. *Nature Methods*, 20(9): 1323-1335.
22. Wolock, S.L., Lopez, R., Klein, A.M. (2019) Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Systems*, 8(4): 281-291.e9.
23. van den Brink, S.C., Sage, F., Vertesy, A., Spanjaard, B., Peterson-Maduro, J., Baron, C.S., et al. (2017) Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nature Methods*, 14(10): 935-936.
24. O'Flanagan, C.H., Campbell, K.R., Zhang, A.W., Kabeer, F., Lim, J.L.P., Biele, J., et al. (2019) Dissociation of solid tumour tissues with cold active protease for single-cell RNA-seq minimizes conserved collagenase-associated stress responses. *Genome Biology*, 20(1): 210.
25. Haghverdi, L., Lun, A.T.L., Morgan, M.D., Marioni, J.C. (2018) Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nature Biotechnology*, 36(5): 421-427.
26. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., et al. (2019) Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods*, 16(12): 1289-1296.

27. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., et al. (2019) Comprehensive integration of single-cell data. *Cell*, 177(7): 1888-1902.e21.
28. Lee-Six, H., Obro, N.F., Shepherd, M.S., Grossmann, S., Dawson, K., Belmonte, M., et al. (2018) Population dynamics of normal human blood inferred from somatic mutations. *Nature*, 561(7724): 473-478.
29. Mitchell, E., Spencer Chapman, M., Williams, N., Dawson, K.J., Mende, N., Calderbank, E.F., et al. (2022) Clonal dynamics of haematopoiesis across the human lifespan. *Nature*, 606(7913): 343-350.
30. Rodriguez-Fraticelli, A.E., Wolock, S.L., Weinreb, C.S., Panero, R., Patel, S.H., Jankovic, M., et al. (2018) Clonal analysis of lineage fate in native haematopoiesis. *Nature*, 553(7687): 212-216.
31. Ludwig, L.S., Lareau, C.A., Ulirsch, J.C., Christian, E., Muus, C., Li, L.H., et al. (2019) Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell*, 176(6): 1325-1339.e22.
32. Weng, C., Yu, F., Yang, D., Poeschla, M., Liggett, L.A., Jones, M.G., et al. (2024) Deciphering cell states and genealogies of human haematopoiesis. *Nature*, 627(8003): 389-398.
33. Granja, J.M., Klemm, S., McGinnis, L.M., Kathiria, A.S., Mezger, A., Corces, M.R., et al. (2019) Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nature Biotechnology*, 37(12): 1458-1465.