Clonal analysis of lineage fate in native haematopoiesis

Alejo E. Rodriguez–Fraticelli1,2, Samuel L. Wolock3, Caleb S. Weinreb3, Riccardo Panero4, Sachin H. Patel1, Maja Jankovic1, Jianlong Sun1,2,†, Raffaele A. Calogero4, Allon M. Klein3 & Fernando D. Camargo1,2

Haematopoiesis, the process of mature blood and immune cell production, is functionally organized as a hierarchy, with self-renewing haematopoietic stem cells and multipotent progenitor cells sitting at the very top1,2. Multiple models have been proposed as to what the earliest lineage choices are in these primitive haematopoietic compartments, the cellular intermediates, and the resulting lineage trees that emerge from them3–10. Given that the bulk of studies addressing lineage outcomes have been performed in the context of haematopoietic transplantation, current models of lineage branching are more likely to represent roadmaps of lineage potential than native fate. Here we use transposon tagging to clonally trace the fates of progenitors and stem cells in unperturbed haematopoiesis. Our results describe a distinct clonal roadmap in which the megakaryocyte lineage arises largely independently of other haematopoietic fates. Our data, combined with single-cell RNA sequencing, identify a functional hierarchy of unilineage- and oligolineage-producing clones within the multipotent progenitor population. Finally, our results demonstrate that traditionally defined long-term haematopoietic stem cells are a significant source of megakaryocyte-restricted progenitors, suggesting that the megakaryocyte lineage is the predominant native fate of long-term haematopoietic stem cells. Our study provides evidence for a substantially revised roadmap for unperturbed haematopoiesis, and highlights unique properties of multipotent progenitors and haematopoietic stem cells in situ.

To probe native lineage relationships in the fully unperturbed bone marrow, we used the Sleeping Beauty lineage-tracing model and TARIS (T7-amplification mediated recovery of integration sites), an improved transposon integration sequencing technique (Fig. 1a and Extended Data Figs 1 and 2)11. Our analysis relies on comparing tags across multiple differentiated populations at different time points to understand the dynamics of lineage coupling, without the need to isolate and transplant prospective progenitor subsets (Fig. 1b). We pulsed adult Sleeping beauty mice with doxycycline (Dox) for two days and, at one, two, four, and eight weeks after induction, sorted transposon-labelling (DsRed+) nucleated erythroblasts, megakaryocyte progenitors (MkPs), granulocytes, monocytes, and B-cell progenitors (Fig. 1c). Notably, control experiments demonstrated that negligible amounts of transposition occur one day after removal of Dox (Extended Data Fig. 3).

We observed that blood lineages were mostly segregated for the first two weeks, suggesting their replacement by unilineage progenitors during this period (Fig. 1d). Later, we began to detect a significant number of shared tags across lineages, revealing the activity of oligolineage progenitors (Fig. 1d and Extended Data Fig. 4). At four weeks, 40.5% (±8.4%) of all monocyte-detected tags (approximately 289 ± 89 clones) were also found in the granulocyte compartment, confirming their well-established common origin (Fig. 1e). Unexpectedly, a similar proportion of erythroblast clones were also found shared with granulocyte/monocyte (myeloid) tags (Fig. 1d, e), revealing a common progenitor for erythrocytes, granulocytes, and monocytes at this stage. Remarkably, we detected virtually no MkP clones that were shared exclusively with erythroblasts during the whole period of observation, which would have been predicted had a megakaryocyte–erythroid progenitor (MEP)-like cell existed (Fig. 1d, e and Extended Data Fig. 4b)12,13. At eight weeks, our analysis revealed the activity of a set of multilineage clones (239 ± 58), with lymphoid (B-cell progenitor), granulocyte/monocyte, and erythroid contribution, but still with no presence in MkP, indicating the existence of megakaryocyte-deficient lympho-erythromyeloid progenitors (Fig. 1d, e). We did observe a very small (9.7 ± 2.8), yet increasing, number of MkP tags shared with multiple lineages after eight weeks (Fig. 1e and Extended Data Fig. 4a, b), suggesting that clonal megakaryocyte-lineage production can also be associated with multilineage outcomes, although at lower frequencies. Spearman’s rank correlation analyses of tag-read distribution between lineage pairs showed a progressive association of granulocyte/monocyte, erythro-myeloid, and lympho-myeloid progenitors, segregated from MkP (Fig. 1f, g). To address potential sampling and sensitivity limitations, we performed independent TARIS amplifications (Extended Data Fig. 5) and clone-specific PCRs (not-shown). Taken together, our results provide evidence for novel lineage couplings during unperturbed haematopoiesis, in which the megakaryocyte lineage is produced largely independently from the other haematopoietic lineages, and argue for the robust activity of erythromyeloid and lympho-erythromyeloid progenitor clones.

We next aimed to identify ancestral relationships by comparing the clonal repertoires of differentiated cells and previously defined progenitor populations. Classically, oligopotent progenitors reside in the common myeloid progenitor (CMP), granulocyte–monocyte progenitor (GMP), and MEP phenotypic gates (referred together as myeloid progenitors, or MyPs)14. Our data revealed largely unilineage outcomes for detected MyPs (89.0 ± 0.8%), suggesting that these populations represent a collection of lineage-restricted progenitors, functionally validating predictions from single-cell expression profiling (Extended Data Fig. 6)14–18. We next focused on the multipotent progenitors (MPPs), the cellular subset proposed to be upstream of MyPs. At 1 and 2 weeks, we observed a small number of ‘active’ MPP tags (overlapping with Lin− tags), which aligned mostly with single lineages (1 week: 75.8 ± 5.0%; 2 weeks: 66.3 ± 6.1%), suggesting the existence of a small population of lineage-committed MPPs that rapidly produce differentiated progeny (Fig. 2a, b and Extended Data Fig. 7a). MPP output significantly increased at 4–8 weeks for all lineages (9.35 ± 0.6% of all MPP tags at 8 weeks), consisting mostly of oligolineage erythromyeloid clones (79.2 ± 5.3% of active MPP clones). A robust number of lympho-erythromyeloid MPP clones (12 ± 2) were detected beginning at eight weeks (Fig. 2a), consonant with our analysis of Lin− fractions (Fig. 1f). Although we also observed oligolineage

1Stem Cell Program, Boston Children’s Hospital, Boston, Massachusetts 02115, USA. 2Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, USA. 3Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. 4Molecular Biotechnology Center, Department of Clinical and Biological Sciences, University of Torino, Torino 10126, Italy. †Present address: School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China.
MkP-producing MPP clones, MkP overlap was more lineage-restricted than any other lineage, even after eight weeks (MkP 67.8 ± 8.0% versus other 22.1 ± 4.6%; Fig. 2a, b and Extended Data Fig. 7b), indicating that at least a subset of MPPs is responsible for a stable restricted contribution to the megakaryocyte lineage.

Our analyses also provided relative quantitative information about the dynamics of lineage replacement by MPPs. For instance, the average clone size of MPP-derived erythromyeloid clones at eight weeks was 18.3 ± 7.7-fold larger than non-MPP-derived clones, suggesting a significant cellular amplification, in contrast to the B-cell progenitor lineage (1.2 ± 0.4-fold; Fig. 2c). In addition, we found that the erythroid lineage was replaced at the fastest rate, with at least 35% of all erythroblast reads overlapping with MPPs after just two weeks, from just a handful of erythroid-committed MPPs (Fig. 2d, e). By comparison, the granulocyte/monocyte-producing MPPs achieved similar levels of replacement only after two months. Considering that our analysis cannot measure the contribution of MPP clones that disappear from the MPP pool (that is, by cell death or differentiation), our results probably underestimate the overall MPP contribution.

To provide further insight into the heterogeneity and hierarchy of the haematopoietic stem cell (HSC)/MPP compartment, we sorted subsets within these populations using previously described surface markers and interrogated their single-cell gene expression landscape using inDrop (Fig. 3a–c)17. Louvain–Jaccard clustering analysis of transcriptomes resulted in 12 reproducibly distinct clusters (Fig. 3b).

Most analysed cells (78.9% of all subsets combined) fitted into one of three major clusters that we labelled as unprimed (‘C1’, ‘C2’, ‘C3’) on the basis of the lack of expression of lineage-restricted gene signatures (Supplementary Table 2 and Extended Data Figs 8 and 9). We also identified several primed clusters (21.1% of HSCs/MPPs) that formed branches defined by progressive expression of genes associated with lineage commitment (Fig. 3b–d, right). Predictably, cells indexed as long-term (LT)-HSCs and MPP1s (also known as short-term (ST)-HSCs) mostly fitted into the ‘C1’ (67.9%) and ‘C2’ (78.3%) clusters, respectively. By contrast, other MPP subsets displayed different degrees of heterogeneity. MPP2s contained the largest proportion of primed cells (59.3%), and MPP4s the least (13.2%) (Fig. 3c, d). MPP2s comprised a larger number of erythroid-primed (18.7%) and megakaryocyte-primed (21.9%) cells, whereas MPP3s contained a larger number of granulocyte/monocyte-primed cells (20.8%) (Fig. 3c, d and Extended Data Fig. 8b). Using transposon tracing, we confirmed that MPP2s presented a preference for MkP production, and generated less multilineage output (5 ± 5% of all active clones) within the first week, where their immediate progeny is likely to be measured, compared with MPP3s and MPP4s (40.17 ± 11.4%) (Fig. 3e, f). Analysis of tags not arising from upstream progenitors at four weeks revealed similar findings (Fig. 3g, h).

On the contrary, MPP4s produced more lympho-erythromyeloid and multilineage clones (Fig. 3h) and preferentially overlapped with MPP1/ST-HSCs, suggesting that at least a fraction of MPP4s represent direct activated progeny of MPP1/ST-HSCs (Fig. 3i). Combined, our data support the notion that a functional hierarchy, consisting of progenitors at varying degrees of lineage priming, exists already within HSCs/MPPs.

Figure 1 | Clonal analysis of haematopoietic lineage fates in the native bone marrow. a, M2/HSB/Tn (transcriptional activator M2/hyperactive Sleeping Beauty/transposon) mouse model. Addition of Dox induces random transposition of the transposon, and concomitant cell labelling with DsRed. The transposon insertion site is stable after removal of Dox. b, Transposon lineage tracing. Shared tags can be detected between a self-renewing progenitor stem cell and its progeny, or between two different mature cell populations. c, Experimental design. M2/HSB/Tn mice were labelled with Dox for 2 days and five blood lineages were isolated from bone marrow after different periods of time. Transposon insertion tag libraries were prepared and sequenced for each population. d, Alignment of transposon tags from different lineage-committed (Lin−) blood cell populations in the bone marrow at 1–8 weeks. Tags are coloured by frequency in each lineage, and organized by rank. Each chart is representative of three independent experiments. MkP, megakaryocyte progenitors; Er, erythroblasts; Gr, granulocytes; Mo, monocytes; B, B-cell progenitors. e, Percentage of clonal overlap between designated lineage pairs (left), and quantification of total number of detected bi/tri-lineage clones at 1–8 weeks (right). Abbreviation My refers to either granulocyte or monocyte lineage. Mean ± s.d. from three independent mice. f, Spearman’s correlation coefficient (ρ) matrices for all Lin− tags at 1–8 weeks. Each matrix is the average from three independent experiments per time point. g, Hierarchical clustering of blood lineages using (1 − ρ) as the distance measure (4 and 8 weeks after labelling).
**Figure 2** | Functional heterogeneity of MPP lineage fates in steady-state haematopoiesis. a. The alignment of all active MPP tags together with the five analysed blood lineages at each time point (all tags collected from three mice per time point). LT-HSC tags were analysed in parallel and excluded from the analysis to represent only MPP behaviour. b. Fraction of active MPP tags that overlap with a single lineage (calculated independently for each lineage). Values are mean ± s.e.m. from three mice. *P_{MPP,Er} = 0.13, P_{MPP,Gr} = 0.03, P_{MPP,Mb} = 0.03, P_{MPP,B} = 0.001 (8 weeks). Abbreviations as Fig. 1. c. Distribution of Lin⁻ clone sizes comparing tags overlapping with MPP versus non-overlapping at eight weeks. Values are median and interquartile ranges of all detected clones from three mice. *Kolmogorov–Smirnov P_{MPP,Er} = 0.03, P_{MPP,Gr} = 0.03, P_{Gr,Mb} = 0.003. d. Fraction of each lineage replaced by MPPs calculated as the percentage of total MPP-overlapping lineage reads over time. Values are mean ± s.e.m. from three independent mice. *P_{Er–Gr/Mb} = 0.04, P_{Er–MPP} = 0.03 (2 weeks) and P_{Er–Gr/Mb} = 0.03, P_{Gr,Mb} = 0.04 (8 weeks). e. Average number of detected active MPP clones per lineage per mouse at different time points (normalized for percentage DsRed labelling efficiency).

**Figure 3** | Transcriptional and functional hierarchy of HSC and MPP subsets. a. Experimental design for inDrops experiment (left). Transcriptional fate map of combined fluorescence-activated cell sorting (FACS) subsets using the SPRING representation (subsampled in silico to represent proportions of the Lin⁻Sca1⁺Kit⁺ (LSK) gate). Points represent a single HSC/MPP distributed according to their similarity using gene expression variation. b. In silico identification of different cell populations within all combined HSC and MPP subsets. Non-primed clusters 1–3 (C1–C3, left) and lineage-primed clusters (right) are presented separated and labelled according to their primed lineage signatures: Neu, neutrophil; DC, dendritic cell; T, T-cell progenitor; B, B-cell progenitor; Er, erythroid; Mk, megakaryocyte; Mo1 and Mo2 represent two monocyte-like signatures. c. Plots showing localization of each sorted HSC/MPP subset within the combined SPRING plot. Top right, fraction of cells from each sorted HSC/MPP subtype (and LSK cells) that group within primed or non-primed clusters. d. Hierarchical clustering (Ward) of sorted HSC/MPP subsets. For each FACS-sorted experiment, the fraction of cells corresponding to each cluster was used to analyse the similarity between subsets. The arrow points out the megakaryocyte-primed cluster within the LT-HSC gate. e. Fraction of lineage-restricted MPP-overlapping clones corresponding to each lineage, for each MPP subset at one week. Values are mean of three independent mice. NS, not significant. f. Fraction of oligolineage output of each MPP subset after 1 week. Values are mean ± s.e.m. of three independent mice. NS, not significant. g. Alignment of Lin⁻ progeny tags of different MPP subsets (excluding tags present in HSCs/MPP1s) at four weeks. h. Fraction corresponding to each MPP subset for each representative lineage fate (including restricted, oligolineage, and multilineage output) at four weeks (all tags detected from four mice). i. Frequency of MPP2/3/4 tags (and LT-HSC tags) overlapping with MPP1 at 1–8 weeks (average of three mice per time point).
Our single-cell RNA sequencing data also revealed that a subset of marker-defined LT-HSCs exhibited megakaryocyte-lineage priming (Fig. 3c, d and Extended Data Fig. 9). This is in line with previous reports of multipotent, yet platelet-biased, subsets of LT-HSCs in the context of transplantation10,18–21. However, the physiological relevance of this observation in native haematopoiesis is unknown. With these precedents, we analysed the Lin− transposon tag overlap of sorted LT-HSCs. Although only a very small number of LT-HSC clones was active four weeks after labelling (5.5 ± 2.3%), remarkably a large majority of these clones were found exclusively in the MkP population (Fig. 4a, b and Extended Data Fig. 10a). This megakaryocyte-restricted output of LT-HSCs was more pronounced after 30 weeks post-labelling (MkP: 13.3 ± 5.6%; lymphoid/erythroid/myeloid: 3.2 ± 1.0%) (Fig. 4c).

Quantitatively, LT-HSCs accounted for replacing at least 31% of the total MkP pool, compared with just 3.8% of granulocyte/monocyte and erythroblast reads (Fig. 4d). Among all MkP that had a detectable tag in primitive populations, approximately half demonstrated overlap with LT-HSCs and the other half with MPPs (where no LT-HSC tag was detected) (Extended Data Fig. 10b). MPP-overlapping clones contributed to the megakaryocyte lineage to a similar extent as LT-HSCs, markedly differing from lympho-erythromyeloid output, which is predominantly MPP-driven (Fig. 4e and Extended Data Fig. 10c).

Our single-cell RNA sequencing data also revealed that a subset of LT-HSCs, MPPs, and Lin− cells were purified from bone marrow at 4 and 30 weeks and their transposon tag content was analysed. Only the LT-HSC tags overlapping with detectable Lin− progeny are shown. Abbreviations as Fig. 1. b, Distribution of types of progeny detected from LT-HSCs at 4 weeks and 30 weeks after labelling. Data are pooled from four independent M2/HSB/Tn mice per time point. Ly, lymphocyte.

c, Percentage of labelled LT-HSC clones producing progeny at 1–8 weeks. Our analyses also revealed that many LT-HSCs contribute to MkP in the absence of any intermediates in the MPP compartment (Fig. 4a), suggesting that at least a subset of LT-HSCs generates megakaryocyte lineage cells through a ‘direct’ pathway.

Previous studies have shown that the commonly used LT-HSC gate contains unilineage CD41+ megakaryocyte-restricted progenitors as
assayed by transplant or culture. To rule out potential contamination by such cells, we aimed to determine whether megakaryocyte-producing LT-HSC clones in situ had properties of classic LT-HSCs in the context of transplantation. For this, we transplanted clonally labelled LT-HSCs isolated from mice four weeks after induction, and at 16 weeks post-transplantation we purified mature lineages from recipients and compared their transposon repertoires with those of cells initially isolated from the donor (Fig. 4f). We observed that six out of eight detected megakaryocyte lineage-restricted LT-HSC clones in the donor were able to generate multilineage progeny in recipients (Fig. 4g, i). We reached similar conclusions when evaluating the culture potential of in situ MkP-producing LT-HSC clones (Extended Data Fig. 10d, e). Additionally, our results demonstrate that MkP production is not exclusive to the CD41+ LT-HSC fraction (Extended Data Fig. 10f, g). Thus, we conclude that most megakaryocyte lineage-producing clones residing in the LT-HSC gate are not simply megakaryocyte-restricted progenitors, but clones that can exhibit multipotency upon transplantation.

Our work here uncovers critical features of the native haematopoietic process. In our model, as much as half of the megakaryocyte lineage is produced independently of other lineages by cells at the top of the haematopoietic ladder (Fig. 4j). A heterogeneous hierarchy of lineage-restricted and oligolineage progenitors, historically classified as MPPs, produce other haematopoietic lineages with selective lineage couplings. Although our work still supports a model for progressive restriction of developmental potential, it suggests that these events are clonally heterogeneous and occur much earlier in the haematopoietic hierarchy, in line with recent data. We postulate that while megakaryocyte–erythrocyte bipotential exists in transplant or culture settings, this fate is not substantially manifested in unperturbed conditions. Alternatively, such cellular behaviour might be too transient to be captured with our technology.

Our data demonstrate that at least a fraction of LT-HSCs behave as a potent source of MkP, indicating that the megakaryocyte fate is the predominant fate of HSCs in situ. However, these same cells exhibit potential for multilineage outcomes following transplantation. Thus, our findings highlight the critical differences between studying native fate versus potential in stem cell biology. Although we are unable to conclude whether a particular subset or all LT-HSCs will eventually display megakaryocyte-producing behaviour, we favour the idea that most LT-HSC clones transition through a megakaryocyte-primed state with age. Our data also suggest that an MPP population (within MPP2) is involved in megakaryocyte production. It remains to be determined whether these represent two different pathways for megakaryocyte production or whether LT-HSCs are upstream of MPPs. Finally, our results are still consonant with the idea that adult LT-HSCs have a limited lympho-erythroid-myeloid output during steady state, although this finding has been debated. Future work with second-generation cell labelling in combination with Cre-based labelling will be needed to elucidate full lineage histories and determine the mechanisms of fate restriction.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to this finding have been debated and will not be included in full lineage histories and determine the mechanisms of fate restriction.

Received 18 November 2016; accepted 21 November 2017.

Published online 3 January 2018.

1. Morrison, S. J., Wasylishen, A. M., Hemmati, H. D., Wright, D. E. & Weissman, I. L. Identification of a lineage of multipotent hematopoietic progenitors. Development 124, 1929–1939 (1997).
2. Morrison, S. J. & Weissman, I. L. The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. Immunity 1, 661–673 (1994).
3. Adolfsson, J. et al. Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential: a revised road map for adult blood lineage commitment. Cell 121, 295–306 (2005).
4. Akashi, K., Traver, D., Miyamoto, T. & Weissman, I. L. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature 404, 193–197 (2000).
5. Ceredig, R., Rolink, A. G. & Brown, G. Models of haematopoiesis: seeing the wood for the trees. Nat. Rev. Immunol. 9, 293–300 (2009).
6. Forsberg, E. C., Serwald, T., Kogan, S., Weissman, I. L. & Passegué, E. New evidence supporting megakaryocyte-erythrocyte potential of Flk2/Flt3+ multipotent hematopoietic progenitors. Cell 126, 415–426 (2006).
7. Nota, F. et al. Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. Science 351, aab2116 (2016).
8. Periê, L., Dufy, K., Kok, L. de Boer, R. J. & Schumacher, T. N. The branching point in erythro-myoelid differentiation. Cell 163, 1655–1662 (2015).
9. Pautou, E. M. et al. Functional distinction of distinct subsets of lineage-biased multipotent progenitors control blood production in normal and regenerative conditions. Cell Stem Cell 17, 35–46 (2015).
10. Yamamoto, R. et al. Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. Cell 154, 1112–1126 (2013).
11. Sun, J. et al. Clonal dynamics of native haematopoiesis. Nature 514, 322–327 (2014).
12. Delibl, N. et al. Characterization of a bipotent erythro-megakaryocytic progenitor in human bone marrow. Blood 88, 1284–1296 (1996).
13. Sanada, C. et al. Adult human megakaryocyte-erythroid progenitors are in the CD34+CD38− fraction. Blood 128, 923–933 (2016).
14. Foege, F. et al. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. Cell 163, 1663–1677 (2015).
15. Pronk, C. J. et al. Elucidation of the phenotypic, functional, and molecular topography of a myeloiderythroid progenitor cell hierarchy. Cell Stem Cell 1, 409–442 (2007).
16. Velten, L. et al. Human haematopoietic stem cell lineage commitment is a continuous process. Nat. Cell Biol. 19, 271–281 (2017).
17. Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell 161, 1187–1201 (2015).
18. Calaminus, S. D. et al. Lineage tracing of P4Cre marks hematopoietic stem cells and their progeny. PLoS ONE 7, e51361 (2012).
19. Gekas, C. & Graf, T. CD41 expression marks myeloid-biased adult hematopoietic stem cells and increases with age. Blood 121, 4463–4472 (2013).
20. Haas, S. et al. Inflammation-induced emergency megakaryopoiesis driven by hematopoietic stem cell-like megakaryocyte progenitors. Cell Stem Cell 17, 422–434 (2015).
21. Nishiki, H. et al. Unipotent megakaryopoietic pathway bridging hematopoietic stem cells and mature megakaryocytes. Stem Cells 33, 2196–2207 (2015).
22. Roch, A., Trachsel, V. & Lutolf, M. P. Single-cell analysis reveals cell division-independent emergence of megakaryocytes from phenotypic hematopoietic stem cells. Stem Cells 33, 3152–3157 (2015).
23. Sanjuan-Pia, A. et al. Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. Nature 502, 232–236 (2013).
24. Vannucci, A. M. et al. Identification and characterization of a bipotent (erythroid and megakaryocytic) cell precursor from the spleen of phenylhydrazine-treated mice. Blood 95, 2559–2568 (2000).
25. Busch, K. et al. Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. Nature 518, 542–546 (2015).
26. Sawai, C. M. et al. Hematopoietic stem cells are the major source of multilineage hematopoiesis in adult animals. Immunity 45, 597–609 (2016).
27. Junker, J. P. et al. Massively parallel clonal analysis using CRISPR/Cas9-induced genetic scars. Preprint at https://doi.org/10.1101/056499 (2017).
28. Raj, B. et al. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain by scGENTAL. Preprint at https://doi.org/10.1101/205534 (2017).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to members of the Carmargo and Klein laboratory for comments. A.R.F. is a Merck Fellow of the Life Sciences Research Foundation and a non-stipendiary European Molecular Biology Organization postdoctoral fellow. This work was supported by National Institutes of Health grants HL118850-01A1 to F.D.C. and PO1HL128850 to F.D.C. which is a Leukemia and Lymphoma Society and a Howard Hughes Medical Institute Scholar. A.M.K. is supported by a Burroughs-Wellcome Fund CASI award, and by the Edward J. Malinckrodt Fellowship.

Author Contributions A.R.F. and F.D.C. designed the study, analysed the data, and wrote the manuscript. A.R.F. performed and analysed the experiments, assisted by M.J., S.P., and J.W. which is a Leukemia and Lymphoma Society and a Howard Hughes Medical Institute Scholar. A.M.K. designed and analysed inDrops experiments and transcriptome data. F.D.C. supervised the study.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to F.D.C. (Fernando.Carmargo@childrens.harvard.edu).

Reviewer Information Nature thanks B. Gottgens and the other anonymous reviewer(s) for their contribution to the peer review of this work.
METHODS
No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Mice. The M2/HSB/Tn mice were generated as previously described. To induce transposon mobilization, 8- to 10-week-old male or female mice with the M2/HSB/Tn genotype were fed with 2 mg ml−1 Dox together with 5 mg ml−1 sucrose in drinking water for 48 h. Thereafter, Dox was removed and successful labelling was verified by retro-orbital sinus peripheral blood cell digestion and analysis (70%L) after 1 week. All animal procedures were approved by the Boston Children’s Hospital Institutional Animal Care and Use Committee. Previous studies have estimated that most haematopoietic lineages are replaced by MPPs within 1–2 months after labelling. Thus, for Lin− lineage couples, M2/HSB/Tn mice were analysed within the first 8 weeks after labelling. Since MyPs have limited self-renewal capacity and are rapidly replaced by MPPs, we performed the MyP analysis at short time points after labelling (1 week) and only considered transposon tags not simultaneously present in MPPs.

Bone marrow preparation. After euthanasia, whole bone marrow (excluding the cranium) was immediately isolated in 2% fetal bovine serum in phosphate buffered saline, and erythrocytes were removed with red blood cell lysis buffer, CD45.1 (Ly5.1) mice were used as transplantation recipients (B6.SJL-Ptprca Pep3b/Joyl, stock 002014, the Jackson Laboratory).

FACS. Lineage depletion was performed using magnetic-assisted cell sorting (Miltenyi Biotec) with anti-biotin magnetic beads and the following biotin-conjugated lineage markers: CD3e, CD19, Gr1, Mac1, and Ter119. Cell populations from bone marrow were purified through four-way sorting using FACSaria (Becton Dickinson) and six-way sorting using MoFlo XDP (Beckman Coulter). The following combinations of cell surface markers were used to define these cell populations: Erythroleukaemia: Ly6G−/4 B220− Ter119−; monocytes: Ly6G−/4 B220−/2 Ter119−; pro-/pre-B cells: Ly6G− B220+ IL7Rα−; MPP: Lin− Kit+ Scal− CD150+ CD41+; MPP1/ST-HSC: Lin− Kit+ Scal− Flt3− CD150− CD48−; MPP2: Lin− Kit+ Scal− Flt3− CD150− CD48−; MPP3: Lin− Kit+ Scal− Flt3− CD150− CD48−; MPP4: Lin− Kit+ Scal− Flt3+ CD150− CD48+; LT-HSC: Lin− Kit+ Scal− Flt3+ CD150− CD48+ (±CD41). Other populations are defined in Supplementary Table 1. Representative examples of sorted populations are shown in Supplementary Figs 1–3. Flow cytometry data were analysed with FlowJo (Tree Star). For transposon tag correlation analysis, we FACS-sorted all the available cells from the whole bone marrow extract using purity modes (approximately 98% purity) at about 75–80% efficiency. The antibodies (their clone number, the commercial house, and concentration) were as follows: Ly6B.2 FITC (7/4, Miltenyi, 1:100), Ly6G Alexa Fluor 700 (IA8, ebioscience), 1:50, Ter119 APC (TER119, ebioscience), 1:100, CD71 BV510 (C2, BD biosciences, 1:100), CD45R(B220) eFluor 450 (RA3-6B2, eBioscience, 1:100), CD19 APC/ Cy7 (1D3, ebioscience, 1:50), CD127(IL-7R) PE/Cy7 (2H7, Biolegend, 1:25), CD48 APC/Cy7 (10F11, Biolegend, 1:100), CD19 APC/Cy7 (1D3, ebioscience, 1:100), CD3e biotin (145-2C11, eBioscience, 1:100), CD19 biotin (MB19-1, eBioscience, 1:100), Gr1 biotin (RB6-865, eBioscience, 1:100), CD11b (Mac1) biotin (M1/70, eBioscience, 1:100), Ter119 biotin (TER119, ebioscience, 1:100), streptavidin eFluor 450 (eBioscience, 1:200), FcγRIII/II FcεR1 (eBioscience, 1:100), CD34 FITC (RAM, eBioscience, 1:25), CD42 APC (HIP1, Becton, 1:100), CD9 PE (M3Z, Becton, 1:200).

Transplantation assays. Whole bone marrow cells or sort-purified LT-HSCs from M2/HSB/Tn mice were transplanted in 150 μl of PBS (Gibco, Thermo Fisher Scientific) through retro-orbital injection into nude irradiated recipient mice (split dose obtaining 9–10 Lethal irradiation). Nude mice were used we observed that orthotopic splenic and bone marrow reconstitution with 2 h interval). Donor cell engraftment and label frequency were analysed after 16 weeks using LSRII equipment (Becton Dickinson).

HSC culture assays. One thousand sort-purified LT-HSCs from M2/HSB/Tn mice were cultured together with 10,000 MS-5 stromal cells in round-bottom 96-well plates together with SCF (100 ng ml−1), TPO (100 ng ml−1), Flt3L (50 ng ml−1), IL7 (20 ng ml−1), IL3 (10 ng ml−1), IL11 (10 ng ml−1), and GM-CSF (20 ng ml−1) in PBS with 1% penicillin/streptomycin and 10% FCS (Thermo Fisher) for 2 weeks, changing the medium 24 h after sort and then every 48 h (Becton Dickinson). Myeloid and lymphoid HSC progeny was FACS-sorted after labelling with Gr-1, Mac1, B220, and CD19 and 180 antibodies for precise and accurate principal component analysis (PCA), retaining the top 50 principal components. The cells were then visualized using SPRING, a graph-based single-cell viewing interface. Visual inspection of the SPRING plot revealed a strong cell-cycle signature defined by high expression of genes associated with the G2/M phase (Cenbl1, Pk1, Cdk2, Aurka, Cenpf, Cenpa, Cenpb2, Birc5, Bub1, Bub1b, Cenab2, Cks2, E2f5, Cdkn2b). Hypothesizing that this instructions. Samples with more than 10,000 cells were purified by a QIAamp DNA Micro kit (56304, Qiagen).

TARIS. Our original technique for molecular identification of transposon integration sites was based on ligation-mediated (LM) PCRs. Others and we have observed significant tag amplification biases with this method, which limit the quantitative potential of the clonal data obtained. To improve the current technique, we developed a method based on TARIS (Extended Data Fig. 1). This method provided similar sensitivity levels as LM-PCR but more quantitatively and reproducibly captures the clonal composition of complex samples (Extended Data Fig. 2). For TARIS, the total purified DNA was subjected to enzymatic restriction with 10 U of HindIII-HF (NEB) overnight. TARIS adaptor primer was hybridized and extended using U Klenow DNA polymerase (NEB) for 2 h. Then, total DNA was cleaned up using Ampure XP SPRI beads (Beckman Coulter) and used as a template for a 20 μl T7 RNA polymerization reaction (NEB, High Yield Hifi PCR kit). Solexa sequencing was performed on HiSeq 2000 (Illumina) at the Tufts Genomics Core. Tag identification and alignment was performed as previously described. In brief, we extracted the transposon-containing reads from each fastq file, trimmed the adaptor and transposon sequences, and aligned the integration sites to the reference mouse genome (Ensembl mm9) using Bowtie 2. Then, reads were normalized between samples (per million reads). Sequences were always compared with at least one additional independently labelled mouse, with libraries prepared in parallel and sequenced in the same HiSeq lane to account for contamination. All tags present in the control mouse samples were filtered out (contaminating reads). Then, read frequencies were column-normalized, and graphs were coloured using a logarithmic scale. For hierarchical clustering based on transposon tag distribution, we first determined the Spearman’s correlation matrix for the compared populations and then performed agglomerative clustering (single method) using (1 – correlation coefficient) as the distance metric. Curve fitting was performed with the Lowess function. All indicated statistical tests were performed using Welch’s s.d. correction (exceptions are mentioned where appropriate).

Data visualization and statistical analysis was performed using Microsoft Excel R (version 3.3.1), and GraphPad Prism (version 7). Primers used were TARIS adaptor primer (5′–GGTACTGAGGCGGCGGGAATTAATACGCTACTGAGGAGTGCTAATAAGCGATGATCACT-3′), Tn-1 primer (5′–CTTGGTGTCATCGAAAGATGATGTC-3′), MAP-Tn-1F primer (5′–ACACTCTTTCCTCCACACGCGCTCTTTCCATGTNNNCGATTTATGACTCACAAT-3′), and MAP-polY primer (5′–GAGACTGCTTTGACAGAGGCTCTTGGC-3′). All primers were ordered from IDT DNA technologies, at 100 mmole scale and HPLC-purified.

Single-cell RNA sequencing and low-level data processing. Transcriptional barcoding and preparation of libraries for single-cell mRNA sequencing was performed using the most up-to-date inDrop protocol.

Micro kit (56304, Qiagen).

Data visualization using SPRING. We combined mRNA count matrices from five simultaneously processed and indexed libraries (LTHSC-2A, LTHSC-2A, MPP2-4A, MPP2-3A, MPP2-2A). Cells with fewer mRNA counts (<1,000 unique molecular identifiers) and stressed cells (mitochondrial gene set Z-score > 1) were filtered out. The remaining high-quality cells (4,248) were total-counts normalized. We next filtered genes, keeping those that were well detected (mean expression > 0.05) and highly variable (CV > 2). Finally, we reduced dimensionality of each gene by projecting onto the top 50 principal components. The cells were then visualized using SPRING, a graph-based single-cell viewing interface.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
cell-cycle signature could affect high-dimensional distances between cells in a way that obscures their segregation by lineage-specific genes, we attempted to remove it\textsuperscript{37}. Specifically, we filtered from the analysis genes that were significantly correlated with the sum Z-score of G2/M genes ($P < 10^{-4}$, Bonferroni corrected; 401 genes total, resulting in 28,205 remaining genes). PCA and clustering analysis were repeated using the reduced gene list.

**Clustering of single-cell profiles.** We performed unsupervised clustering of the processed single-cell data with the Louvain–Jaccard method package from ref. 38. To assess cluster stability and choose the value of $k$, we downsampled 85% of cells and applied the Louvain–Jaccard method using 50 principal components. We tested $k$ values from 10 to 30 and for each $k$ we compared 100 times the randomly downsampling clustering using the Jaccard-index measurement in the R package fpc (Flexible Procedures for Clustering). We considered a Jaccard-index minimum of 0.75 as sufficiently robust and selected values of $k > 30$, which resulted in the identification of 11–12 clusters\textsuperscript{39}. Differential expression analysis was performed using the method package from ref. 38 (results are included in Supplementary Table 2).

**Data availability.** The Gene Expression Omnibus accession number is GSE90742. Additional data files will be made available upon reasonable request from the corresponding author. SPRING plots (with and without removal of the G2/M cell-cycle signature) are available for inspection at the following links: https://kleintools.hms.harvard.edu/tools/springViewer.html?cgi-bin/client_datasets/ARF2017_combinednocycle and https://kleintools.hms.harvard.edu/tools/springViewer.html?cgi-bin/client_datasets/ARF2017_combined.
Extended Data Figure 1 | TARIS. Illustration of the TARIS procedure. The procedure is described in detail in the Methods.
Extended Data Figure 2 | See next page for caption.
Extended Data Figure 2 | Evaluation of the TARIS method. a, Design for the detection limit experiment. Spike-ins of a known number of HEK293 cells carrying unique transposon integration tags were used in a mix of 10,000 DsRed⁺ peripheral blood cells from a freshly induced HSB mouse. b, Detection limit chart. Values represent the read number for each clone and for each number of input cells. Both axes are in log10 scale. Values represent the sum of two independent experiments. c, Comparison of the average read number value between TARIS and the LM-PCR method. Values represent mean ± s.d. of five different transposon tag clones. d, Reproducibility analysis in a non-whole-genome amplified sample with high complexity (2 × 10⁶ bone marrow granulocytes 4 weeks after pulse). e, Reproducibility in a whole-genome amplified sample with low complexity (863 LT-HSCs 4 weeks after pulse). f, Venn diagram showing overlapping transposon tag reads between two TARIS replicates from the same sample high-complexity sample (2 × 10⁵ bone marrow monocytes at 4 weeks after induction). g, Venn diagram showing overlapping transposon tag reads between two TARIS replicates from the same low-complexity sample (863 LT-HSCs at 4 weeks after induction). h, Contamination analysis between samples from two different mice. The plot represents the read numbers of tags from Lin⁺ populations from mouse 1, and their read number values in Lin⁺ populations in mouse 2. High-confidence tags are selected as those tags with more than 25 reads, and at least 10 times higher read count compared with any of the samples from a separate mouse.
Extended Data Figure 3 | Analysis of residual HSB activity after Dox withdrawal. a, Experimental design. Residual HSB activity after Dox removal was assayed by transplantation into CD45.1 mice. Sub-lethally irradiated recipients were treated with Dox for 48 h. Dox was removed 12 h before transplantation. Ten million whole bone marrow cells were transplanted and mice were allowed to recover for 2 weeks. As a positive control, mice were continuously treated with Dox until 48 h after transplant. As a negative control, cells were transplanted into non-Dox treated mice. DsRed labelling was analysed as a proxy for HSB activity in granulocytes, erythroblasts, monocytes, and B cells. b, FACS plots showing the negligible labelling of CD45.2 M2/HSB/Tn cells in transplanted recipients 24 h after Dox removal.
Extended Data Figure 4 | Additional representations and analyses of Lin\(^{-}\) tags. a, Lin\(^{-}\) clones of the second and third mice used for quantifications in Fig. 1d–g. b, Scale-adjusted binary (presence/absence) representation of all detected MkP and erythroblast tags in the experiments from Fig. 1d–g. c, Relative quantification of scale-normalized clone sizes for each lineage, comparing unilineage versus oligo/multilineage clones. Values are interquartile range and median from three independent mice at four weeks and eight weeks after induction.
Extended Data Figure 5 | Validation of granulocyte/monocyte/erythroblast and megakaryocyte-restricted tags. a, Three independent transposon tag libraries were prepared and sequenced from 2-week-, 4-week-, and 8-week-chased mice. Reads from the three libraries were then pooled together for each lineage.
Extended Data Figure 6 | Lineage fate of myeloid progenitors. **a**, Two M2/HSB/Tn mice were induced and chased for one week, and then myeloid progenitors (GMP, MEP, and CMP) and Lin⁺ cells were isolated from bone marrow and their transposon tag content was analysed. Chart is a binary representation of all Lin⁺ tags overlapping with any myeloid progenitor tag ranked by lineage. **b**, Quantification of relative lineage contribution of GMPs, MEPs, and CMPs as a fraction of lineage-specific/total lineage-overlapping clones for each MyP subset. Values are mean of the two analysed mice. **c**, An additional M2/HSB/Tn mouse was induced and chased for three weeks, and then processed as in **a**. **d**, Quantification of relative lineage contribution of GMPs, MEPs, and CMPs at three weeks after labelling.
Extended Data Figure 7 | Additional analyses of MPP clonal outcomes.

a, Quantification of the percentage of MPP clones that produced any Lin⁺ progeny at different time points. Values are average ± s.d. from three mice.

b, Three independent transposon tag libraries were prepared and sequenced for all the populations from one bone marrow at 2, 4, and 8 weeks after labelling. Each column in the charts represents the combined tags detected in any of the three libraries for each population.
Extended Data Figure 8 | Single-cell heterogeneity of HSC/MPPs.

a, SPRING plots showing selected differentially expressed markers. Scale represents amount of detected mRNA copies (normalized) of each marker gene. b, Enrichment score analysis for single cells in each FACS-sorted population compared with previously obtained bulk transcriptional signatures of bone marrow populations sorted using traditional markers (from the Immgen database).
Extended Data Figure 9 | Differentially expressed markers for clusters C1, C2, C3, and megakaryocyte. a, FACS plots showing heterogeneity in expression of cluster markers within the analysed HSC/MPP subsets. b, FACS plots showing expression of different megakaryocyte-primed cluster markers (CD41, CD42, and CD9) within the LT-HSC gate. c, The expression value (nTrans) and percentage of expressing cells from each cluster (% Exp). The top ten highest expressed genes that distinguish each cluster are shown.
Extended Data Figure 10 | See next page for caption.
Extended Data Figure 10 | Additional data on clonal origin of MkP.

**a,** Three independent transposon tag libraries were prepared and sequenced for LT-HSC, MPP, and the five Lin− populations, from one mouse at four weeks. Each column represents the combined tags detected from three amplicon libraries prepared for each population, to facilitate visualization of the smallest clones. Tags are coloured by frequency in each lineage, and organized by rank. **b,** Origin of megakaryocytes. Alignment of all MkP clones that had detectable tags in HSC/MPPs from a mixed library combining three independent sequencing reactions. Tags are coloured by frequency in each lineage (except for MkP), and organized by rank. Arrows indicate tags verified by clone-specific PCR. **c,** Alignment of transposon tags from all Lin− populations, LT-HSCs, and MPPs collected from 30-week-chased mice. Tags are coloured by frequency in each lineage, and organized by rank. **d,** Experimental design for testing *in vitro* myeloid and lymphoid potential from sorted LT-HSCs. **e,** *In vitro* myeloid potential of LT-HSCs. Alignment of donor Lin− tags with transposon tags obtained from myeloid and lymphoid cells derived from donor LT-HSCs after two weeks in culture. **f,** Clonal output of CD41hi and CD41lo LT-HSCs at four weeks after labelling. **g,** Quantification of megakaryocyte lineage replacement by CD41hi versus CD41lo LT-HSCs (measured as the percentage of overlapping/total MkP reads) at four weeks after labelling. Values are mean ± s.e.m. of three independent mice.
# Experimental design

1. **Sample size**

   Describe how sample size was determined.

   For sufficient statistical power, sample size was defined by analyzing Tn tag data available from a previous study (Sun et al. Nature 2014). No further determination was made a posteriori.

2. **Data exclusions**

   Describe any data exclusions.

   Some data were excluded based on the quality of library sequences. Samples that presented sequencing results with poor quality (based on FastQC score), DNA libraries that presented mostly sequences without transposon (>70%) and samples with less than 10% of the predicted number of Tn tags were excluded.

3. **Replication**

   Describe whether the experimental findings were reliably reproduced.

   Experimental findings were reliably reproduced. We did find a large variability in some of our observations, and these are indicated where appropriate.

4. **Randomization**

   Describe how samples/organisms/participants were allocated into experimental groups.

   Weaned mice from a Tn/Tn x M2/M2 HSB/HSB cross were selected randomly and separated into male and female cages. These mice were labeled with doxycyclin and randomly chosen at different time points for BM isolation.

5. **Blinding**

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   No blinding was performed during data collection.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. **Statistical parameters**

   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

   - **n/a**
   - **Confirmed**

   - The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)

   - A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly

   - A statement indicating how many times each experiment was replicated

   - The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)

   - A description of any assumptions or corrections, such as an adjustment for multiple comparisons

   - The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted

   - A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)

   - Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.
7. Software

Describe the software used to analyze the data in this study.

Custom code required for Tn tag integration analysis has been previously published (Sun et al. Nature 2014).

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). Nature Methods guidance for providing algorithms and software for publication provides further information on this topic.

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

N/A

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

All antibodies were purchased from Biolegend or BD Biosciences, and are well characterized and validated by providers. Antibodies used and their concentrations are described in methods (page 10).

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

293T cells were obtained ATCC and were used as previously described (Sun et al. Nature 2014).

b. Describe the method of cell line authentication used.

N/A

c. Report whether the cell lines were tested for mycoplasma contamination.

Cell lines were regularly tested for mycoplasma contamination.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

Cells were used mainly for the purpose of validation of the technique of Tn integration quantitation, and not for describing any physiological features.

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Tn and HSB mice were generated as previously described (Sun et al. Nature 2014). Tn/Tn homozygous mice (backcrossed into the C57BL/6J background) were crossed to M2/M2 HSB/HSB homozygous mice (mixed C57BL/6j x 129/SvJ background). To induce Tn mobilization, 8-10 weeks old male or female mice with the M2/HSB/Tn genotype were fed with 2mg/ml Dox together with 5mg/ml sucrose in drinking water for 48h. Thereafter, Dox was removed and successful labelling was verified by retroorbital sinus peripheral blood collection and analysis (70 ul) after 1 week. All animal procedures were approved by the Boston Children’s Hospital Institutional Animal Care and Use Committee.

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

N/A

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines.

Policy information about studies involving human research participants.
Flow Cytometry Reporting Summary

Data presentation

- For all flow cytometry data, confirm that:
  - 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
  - 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
  - 3. All plots are contour plots with outliers or pseudocolor plots.
  - 4. A numerical value for number of cells or percentage (with statistics) is provided.

Methodological details

5. Describe the sample preparation.
   - Cells were prepared from whole bone marrow, lineage-depleted using quadroMACS (LS columns), and filtered (40 um, BD) before FACS.

6. Identify the instrument used for data collection.
   - FACSAria Ilu (BD), LSRII (BD), Astrios XP (Beckman)

7. Describe the software used to collect and analyze the flow cytometry data.
   - FlowJo (TreeStar)

8. Describe the abundance of the relevant cell populations within post-sort fractions.
   - Cells were sorted with Purity modes at 75-80% efficiency. Post sort fractions analyzed were at least 98% pure.

9. Describe the gating strategy used.
   - This information is included in the Methods section (page 9-10) and Supplementary Information (Figures 1-3).

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☒