Supplementary Information

Methods

Mice

C57BL/6 wild-type (CD45.1, CD45.1/2 and CD45.2), TdT^{hCD4}, TdT^{iCre}, $Rosa26^{LSL\text{-}YFP}$ and $Rosa26^{mTmG}$ mice[1,2] were bred and maintained in our animal facility under specific pathogen free conditions according to institutional guidelines (Veterinäramt BS, license number 2786_26606 and ASP Number: 19-896). All mice used as donors in transplantations and for analysis were 6-10 and recipient mice were 8-15 weeks old, and all were of the C57BL/6 strain.

TdT^{hCD4} and TdT^{iCre} mice were generated at the Center for Transgenic Models in Basel using Cas9/CRISPR technology. All Cas9 reagents were purchased from IDT. Briefly, RNPs consisting of Cas9 protein (40 ng/µl), trcrRNA (20 ng/µl) and crRNAs (10 ng/µl each) targeting the last exon of the *Dntt* gene just before the stop codon, together with a single stranded DNA template (IDT) encoding the P2A self-cleaving peptide [3] in front of the human *CD4* or *iCre* coding sequence flanked by 200 base pair long homology arms, were microinjected into C57BL/6 zygotes essentially as described in [4]. Embryos that survived the DNA and Cas9 RNP microinjections were transferred into pseudo-pregnant females generated by mating with genetically vasectomized males [5] and the offspring were allowed to develop to term. Extended Data 1a illustrates the strategy used to generate the TdT^{hCD4} and TdT^{iCre} mice by Cas9 mediated homology directed repair. Genotyping was performed by PCR using different sets of primers. To detect hCD4 and iCre integration forward and reverse primers were located within the transgenes: PCR1: hCD4 FW1 + hCD4 RV1 (200bp product); iCre FW1 + iCre RV1 (258bp product) (Supplementary Table 1). To distinguish between homozygous and heterozygous mice a forward primer located in the Dntt gene right before the transgenes and a reverse primer located in the untranslated region of the Dntt gene right after the transgenes were used: PCR2: Dntt FW1 + Dntt RV1 (291bp product) (see Table 1). In mice heterozygous for hCD4 or iCre insertion both PCRs are positive, while for homozygous animals PCR2 is negative (product too large for amplification). Furthermore, combinations of the primers allowed to confirm transgene integration at the designated site: PCR3+4: hCD4 FW1 or

iCre FW1 + Dntt RV1; Dntt FW1 + hCD4 RV1 or iCre RV1. PCRs were performed with GoTaq Green Master Mix (Promega) according to the manufacturer`s instructions.

Cell harvest and flow cytometry

For analysis and sorting, bone marrow cells were flushed or extracted through fragmentated with a mortar and pestle from femurs and/or tibiae and/or pelvic bones of the two hind legs of mice with FACS buffer (PBS containing 0.5% BSA and 5 mM EDTA) and single-cell suspensions of spleen and thymus cells were made. Debris was removed by filtration through a 70 μm strainer. Red blood cells were lysed with ACK lysis buffer. Cells were counted and stained in FACS buffer with antibodies of interest (Table 1) for 30 min at 4°C. Cells were additionally stained with propidium iodide or 7AAD to exclude dead cells. For blood cell analysis 5 μL of blood were used for platelet and 50 μL for B cell and myeloid cell staining. After 30 min at room temperature 2 mL of FACS buffer were added to the platelet staining, which were then readily analyzed. To lyse red blood cells 2 mL of FACS lysing solution (BD Biosciences) were added to the B cell and myeloid cell staining before analysis. For intra-cellular staining, cells were fixed and permeabilized after cell-surface staining using a Fix/Perm buffer set (Invitrogen) according to the manufacturers protocol. Enrichment of progenitor cell populations prior to sorting was performed by Magnetic-Activated Cell Sorting (Milteny Biotec) using biotin labeled antibodies directed against lineage markers (CD3, CD19, B220, Ter119, NK1.1, and Ly6G) and anti-biotin MicroBeads (Milteny Biotec) according to the manufacturers protocol. For cell sorting, a BD FACSAria IIu instrument (BD Biosciences) with a custom built-in violet laser was used. Cells were sorted into Iscove's modified Dulbecco's medium (IMDM) supplemented with 5% fetal bovine serum, 5 x 10-5 M β-mercaptoethanol, 1mM glutamine, 0.03% (wt/vol) Primatone, 100 units/mL penicillin, and 100 μg/mL streptomycin. Cell purities of at least 95% were confirmed by post-sort analysis. Cells were analyzed on a BD LSR Fortessa instrument (BD Biosciences), and data were analyzed with FlowJo X software (TreeStar).

B- and T-cell progenitor populations were gated as previously described [6, 7] (Fig. 1a,d,e; Extended Data 1e). For the identification of the hematopoietic stem cell and multipotent progenitor compartment BM cells were gated as lineage negative (CD3, CD19, B220, CD11b, CD11c, GR-1, Ter119, and NK1.1), Sca-1^+ and $cKit^{high}$ (LSK compartment). LSK cells were further separated into $FLT3^-CD48^-CD150^+$ LT-HSC, $FLT3^-CD48^-CD150^-$ ST-HSC, $FLT3^-CD48^+CD150^+$ MPP2,

FLT3$^-$CD48$^+$CD150$^-$ MPP3, and FLT3$^+$ MPP4 (Fig. 2c)[8]. GMP, CFU-E and MkP progenitor populations were identified as lineage negative (see LSK compartment) and cKit$^+$Sca-1$^-$CD127$^-$. GMPs were further gated as CD41$^-$CD16/32^{high}CD150$^-$, CFU-E as CD41$^-$CD16/32^{low}CD150$^-$ CD105$^+$ and MkP as CD41$^+$ (Extended Data 2f) [9]. MDP, CDP and cMoP progenitor populations were identified by excluding cells stained positive for the following lineage markers: CD3, CD19, B220, Ter119, and NK1.1. MDPs and CDPs were further defined as Ly6C$^-$FLT3$^+$CD115$^+$ and distinguished as cKit^{high} and cKit$^{low/int}$, respectively, while cMoPs were defined as cKit^{high}Ly6C$^+$CD115$^+$ (Extended Data 2g). Mature cell populations were defined as the following: B cells (CD3$^-$CD19$^+$), NK cells, (CD3$^-$CD19$^-$NK1.1$^+$), CD4 T cells (CD3$^+$CD4$^+$CD8$^-$), CD8 T cells (CD3$^+$CD4$^-$CD8$^+$), γδ T cells (CD3$^+$CD4$^-$CD8$^-$TCRγδ$^+$), pro-erythrocytes (CD3$^-$CD19$^-$ Ter119$^+$CD71^{high}CD105$^+$), platelets (FSC^{low}Ter119$^-$CD41$^+$CD61$^+$), pDCs (CD3$^-$CD19$^-$CX3CR1$^-$ Siglec-H$^+$ and/or Bst2$^+$), cDCs (CD3$^-$CD19$^-$CD11c^{high}MHCII^{high}, if indicated cDCs were split into XCR1$^+$ cDC1 and Sirpα$^+$ or CD11b$^+$ cDC2), monocytes (CD3$^-$CD19$^-$CD11b$^+$Ly6C^{high}), and granulocytes (CD3$^-$CD19$^-$CD11b$^+$Ly6C^{low}) (Extended Data 2a-d).

Transplantations

For transplantation experiments recipient mice were either sub-lethally (600 rad) or lethally (900 rad) irradiated using a Cobalt source (Gammacell 40, Atomic Energy of Canada, Ltd) ~3 hours prior to transplantation. Indicated numbers of purified donor cells were injected intravenously. At indicated timepoints blood was collected from the tail vein (50-75 μL) and stained for platelet, myeloid cell and B-cell reconstitution. Recipient mice were euthanized at indicated timepoints after cell transfer and their spleen and bone marrow were analyzed for the presence of donor cells.

Limiting dilution assays

Limiting dilution assays were adapted from[10]. In brief, ST2 [11] or OP9 [12] stromal cells were plated at a concentration of 4000 cells per well in a 96-well flat-bottom plate one day prior to plating. One day later, the s.e.m.i-confluent stromal cells were γ-irradiated with 2000 rad using a Cobalt source (Gammacell 40, Atomic Energy of Canada, Ltd). Populations of interest were sorted and plated at different concentrations (3, 6, 12, 24, or 48 cells per well). Cultures were maintained in supplemented IMDM, for ST2 co-cultures, or Opti-MEM (Gibco) supplemented with 10% fetal

bovine serum, 100 units/mL penicillin, 100 µg/mL streptomycin, 50 ng/mL murine IL-7 (PeproTech), 50 ng/mL human FLT3-ligand (produced in-house) and 25 ng/mL murine stem cell factor (produced in-house) for OP9 co-cultures, at 37°C in a humidified atmosphere containing 10% CO2 in the air. After 14 or 18 days in culture, for ST2 or OP9 co-cultures, respectively, wells were inspected under an inverted microscope, and wells containing colonies of more than 50 cells were scored as positive.

Methylcellulose cultures

For BFU-E methylcellulose assays, 500-2000 cells in 1 mL SF M3436 (StemCell Technologies) supplemented with 100 units/mL penicillin, 100 µg/mL streptomycin were cultured in a 3 cm petri dish. For simultaneous assessment of multilineage CFU-GEMM, CFU-GM, CFU-G, CFU-M and CFU-E colonies, 200 cells were cultured in 1 mL M3231 (StemCell Technologies) supplemented with 5% FBS, L-Glutamine (2 mM), 100 units/mL penicillin, 100 µg/mL streptomycin, and the following cytokines: SCF (25 ng/mL), FLT3-ligand (25 ng/mL), GM-CSF (10 ng/mL), EPO (25 ng/mL), TPO (25 ng/mL), IL-3 (10 ng/mL), and IL-11 (25 ng/mL). Colonies were counted after 10 days of culture under an inverted microscope. Colonies are defined as CFU-GEMM (colonies forming units containing granulocytes, macrophages and erythrocytes or megakaryocyte progenitors), CFU-GM (mixed granulocyte and macrophage colonies), CFU-G (granulocyte colonies), CFU-M (macrophage colonies), and CFU-E (erythroid colonies).

Quantitative RT-PCR

Total RNA was extracted using RNAqueous Micro Kit (Invitrogen) followed by cDNA synthesis using GoScript reverse transcription (Promega) according to the manufacturer's protocols. Quantitative PCR was performed using SYBR green PCR Master Mix (Applied Biosystems), and samples were run on an Applied Biosystems StepOnePlus qPCR machine.

Cellular indexing of transcriptomes and epitopes by sequencing (CITE-Seq)

Bone marrow cells from four $TdT^{hCD4/YFP}$ double reporter mice were isolated and enriched for progenitor cells by MACS by the usage of antibodies directed against CD3, CD19, B220, Ter119, and Ly6G. Subsequently cells were stained with antibodies directed against additional lineage markers (CD11b, CD11c, NK1.1, GR-1), Sca-1, and CD117 in order to identify LSK cells.

In addition, cells were stained with antibodies coupled to oligonucleotides directed against hCD4, FLT3, CD48, CD150, CD9, CD41, CD55, CD105, CD115, CXCR4, and ESAM (Biolegend, see Table 1). LSK cells were sorted and an estimate of 4'000-6'000 cells per mouse were loaded on one well each of a single 10x Genomics Chromium Single Cell Controller. Single-cell capture and cDNA and library preparation were performed at the Genomics Facility Basel of the ETH Zurich, Basel, with a Single-Cell 3' v3 Reagent Kit (10x Genomics) according to the manufacturer's instructions with the changes as described in [13] to capture cDNA and produce libraries from antibody derived oligos (ADT). Sequencing was performed on 4 lanes (2 flow-cells) of an Illumina NovaSeq 6000 instrument, with a mix of 90% cDNA library and 10% ADT library for the 2 first lanes, and 95% cDNA library and 5% ADT library for the 2 last lanes, to produce 91nt-long R2 reads.

The dataset was analyzed by the Bioinformatics Core Facility, Department of Biomedicine, University of Basel. Read quality was controlled with the FastQC tool (version 0.11.5). Sequencing files of both cDNA and ADT libraries were jointly processed using the Cell Ranger Software (v3.1.0), and the "Feature Barcoding Analysis" instructions (https://support.10xgenomics.com/single-cell-gene expression/software/pipelines/latest/using/feature-bc-analysis) were followed to perform quality control, sample demultiplexing, cell barcode processing, alignment of cDNA reads to the mm10 genome with STAR (version 2.6.1.a) [14] and counting of UMIs for cDNAs and CITE-Seq antibody barcodes. Default parameters were used for Cell Ranger, except for the STAR parameters *outSAMmultNmax* set to 1 and *alignIntronMax* set to 10000. The reference transcriptome refdata-cellranger-mm10-3.0.0 using Ensembl 93 gene models (https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest) was used, and supplemented by the sequences of the YFP, human CD4 and iCre constructs from the $TdT^{hCD4/YFP}$ double reporter mice.

Filtering for high-quality cells was done based on library size (at least 1,000 UMI counts per cell), the number of detected genes (at least 1,000 genes detected) and the percentage of reads mapping to mitochondrial genes (larger than 0% and lower than 7%), based on the distribution observed across cells. Low-abundance genes with average counts per cell lower than 0.015 were filtered out. After quality filtering, the resulting dataset consisted of UMI counts for 12,165 genes and 20,595 cells, ranging from 3,932 to 6,286 per sample.

Further analyses were performed using R (version 3.6), and Bioconductor (version 3.10) packages, notably dropletUtils (version 1.6.1) [15], scran (v1.14.6) [16] and scater (v1.14.6) [17], and the Seurat package (v4.0.5), [18] mostly following the steps of the workflow presented at https://osca.bioconductor.org/ (Amezquita et al., 2019). Clustering of cells was performed on normalized [16] and denoised log-count values with hierarchical clustering on the Euclidean distances between cells (with Ward's criterion to minimize the total variance within each cluster [19]; package cluster version 2.1.0). The number of clusters used for following analyses was identified by applying a dynamic tree cut (package dynamicTreeCut, version 1.63-1) [20], resulting in 12 clusters and an average silhouette width of 0.09. As complementary clustering approach we used the Seurat graph-based clustering, using the FindNeighbors() function on the 10 first principal components of the PCA results, and a k of 20, followed by calling the FindClusters() function with a resolution of 0.6 (Data not Shown).

Cell cycle phase was assigned to each cell using the *cyclone* function from the scran package and the available pre-trained set of marker pairs for mouse [21]. The vast majority of the cells classified in G2M or S phase belonged to a subset of three clusters, so to best eliminate the effects of cell-cycle we filtered out cells from these clusters, and in the other clusters only retained the cells classified in G1 phase (Extended Data 4a,b,d). Cells from an additional cluster were filtered out because it was heterogenous and composed of cells with elevated percentage of reads mapping to mitochondrial genes (e.g., likely of lower quality; Extended Data 4c). The final filtered dataset was composed of 15,853 cells, ranging from 3,081 to 4,849 per sample. Re-clustering of these cells resulted in 8 clusters and an average silhouette width of 0.1. The *findMarkers* function of the scran package was used to find markers (genes, constructs or CITE-Seq antibodies) up-regulated in any of the clusters. The top 30 markers for each cluster were extracted and pooled to from a list of 104 markers (Fig. 4d). DEG are displayed in Table 2.

The Bioconductor package SingleR (version 1.0.5) was used for cell-type annotation of the cells [22] using as reference the relevant samples from the Immunological Genome Project (ImmGen) mouse RNA-seq dataset (LTHSC.34-.BM", "LTHSC.34+.BM", "STHSC.150-.BM", "MPP2.150+48+.BM", "MPP3.48+.BM" and "MPP4.135+.BM") [23, 24, 25, 26, 27, 28] and the HSC, MPP1, MPP2, MPP3, and MPP4 bulk RNA-seq samples from Cabezas-Wallscheid et al. [29]. For the visualization of SingleR scores across cells on heatmaps, the scores were scaled between 0 and 1 across populations for each cell and cubed to improve dynamic range next to 1 [22]. *A posteriori*

gating of cells to the LT-HSC, ST-HSC, MPP2, MPP3 and MPP4 subpopulations was performed based on the surface protein signal from the CITE-Seq antibodies (except for FLT3/CD135 which displayed a continuous gradient, leading us to use also the *Flt3* transcript expression level to recover gating results most similar to the FACS analyses as shown in Extended Data Fig. 4e,f. For classification of $YFP^{+/-}$, $hCD4^{+/-}$ and $ESAM^{+/-}$ cells, a similar thresholding approach was used, and the *findMarkers* function of the scran package was used to find differentially expressed markers between positive and negative populations at a false discovery rate (FDR) of 1% (in both directions).

A uniform manifold approximation and projection (UMAP) dimensionality reduction was used for visualizing single cells on 2 dimensions [30], calculated using the *runUMAP* function from the scater package and default parameters (using the 10 components of the denoised principal component analysis as input, the 500 most variable genes, and a neighborhood size of 15). For visualization, the y-axis coordinates were adjusted which led to exclusion of 8 cells separating from the bulk of other cells on the second dimension. Contour lines displaying the 2D cell density on the UMAP space were calculated with the MASS package (version 7.3-51.5).

Trajectory analysis was performed with the Bioconductor package Slingshot (version 1.4.0) [31], a choice based on the very good performances of this tool in a recent benchmark of 45 single-cell trajectory inference [32]. We ran the analysis using the UMAP coordinates and the hierarchical clustering labels. Cluster 8 (HSCs) was set up as the start cluster. The cluster-based minimum spanning tree and the reconstructed smooth curves are shown in Fig. 5g. We compared this trajectory to the Monocle 3 results, where a cell from cluster 8 was also set as starting point of the trajectory (Fig 5g) [33].

Integration of our dataset with a scRNA-seq dataset of sorted subsets from Rodriguez-Fraticelli et al [28, 34] was done using the findIntegrationAnchors function from the Seurat package [35]. A newly generated UMAP projection of the joint dataset is shown as Extended Data Fig. 4i.

Statistical analysis

A two-tailed unpaired Student`s t test was performed comparing frequency of YFP^{+} subsets in BM and spleen at steady state and following sublethal irradiation (Fig. 7a). *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

A multiple two-tailed unpaired Student`s t test was performed for Experiments shown in Fig. (6e) *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$. Error bars indicate s.e.m.

Supplemental References

1. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev Biol* **1**, 4 (2001).

2. Muzumdar, M.D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent Cre reporter mouse. *Genesis* **45**, 593-605 (2007).

3. Trichas, G., Begbie, J. & Srinivas, S. Use of the viral 2A peptide for bicistronic expression in transgenic mice. *BMC Biol* **6**, 40 (2008).

4. Jacobi, A.M. *et al.* Simplified CRISPR tools for efficient genome editing and streamlined protocols for their delivery into mammalian cells and mouse zygotes. *Methods* **121-122**, 16-28 (2017).

5. Haueter, S. *et al.* Genetic vasectomy-overexpression of Prm1-EGFP fusion protein in elongating spermatids causes dominant male sterility in mice. *Genesis* **48**, 151-160 (2010).

6. Klein, F. *et al.* The transcription factor Duxbl mediates elimination of pre-T cells that fail beta-selection. *J Exp Med* **216**, 638-655 (2019).

7. Klein, F. *et al.* Accumulation of Multipotent Hematopoietic Progenitors in Peripheral Lymphoid Organs of Mice Over-expressing Interleukin-7 and Flt3-Ligand. *Front Immunol* **9**, 2258 (2018).

8. Pietras, E.M. *et al.* Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions. *Cell Stem Cell* **17**, 35-46 (2015).

9. Pronk, C.J. *et al.* Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. *Cell Stem Cell* **1**, 428-442 (2007).

10. von Muenchow, L. *et al.* Permissive roles of cytokines interleukin-7 and Flt3 ligand in mouse B-cell lineage commitment. *Proc Natl Acad Sci U S A* **113**, E8122-E8130 (2016).

11. Ogawa, M. *et al.* B cell ontogeny in murine embryo studied by a culture system with the monolayer of a stromal cell clone, ST2: B cell progenitor develops first in the embryonal body rather than in the yolk sac. *EMBO J* **7**, 1337-1343 (1988).

12. Nakano, T., Kodama, H. & Honjo, T. Generation of lymphohematopoietic cells from embryonic stem cells in culture. *Science* **265**, 1098-1101 (1994).

13. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* **14**, 865-868 (2017).

14. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

15. Griffiths, J.A., Richard, A.C., Bach, K., Lun, A.T.L. & Marioni, J.C. Detection and removal of barcode swapping in single-cell RNA-seq data. *Nat Commun* **9**, 2667 (2018).

16. Lun, A.T., Bach, K. & Marioni, J.C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol* **17**, 75 (2016).

17. McCarthy, D.J., Campbell, K.R., Lun, A.T. & Wills, Q.F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179-1186 (2017).

18. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021).

19. Murtagh, F.L.P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?
. *Journal of Classification* **31**, 274-295 (2014).

20. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719-720 (2008).

21. Scialdone, A. *et al.* Computational assignment of cell-cycle stage from single-cell transcriptome data. *Methods* **85**, 54-61 (2015).

22. Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* **20**, 163-172 (2019).

23. Gazit, R. *et al.* Transcriptome analysis identifies regulators of hematopoietic stem and progenitor cells. *Stem Cell Reports* **1**, 266-280 (2013).

24. Heng, T.S., Painter, M.W. & Immunological Genome Project, C. The Immunological Genome Project: networks of gene expression in immune cells. *Nat Immunol* **9**, 1091-1094 (2008).

25. Yoshida, H. *et al.* The cis-Regulatory Atlas of the Mouse Immune System. *Cell* **176**, 897-912 e820 (2019).

26. Biddy, B.A. *et al.* Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564**, 219-224 (2018).

27. Dong, F. *et al.* Differentiation of transplanted haematopoietic stem cells tracked by single-cell transcriptomic analysis. *Nat Cell Biol* **22**, 630-639 (2020).

28. Rodriguez-Fraticelli, A.E. *et al.* Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis. *Nature* **583**, 585-589 (2020).

29. Cabezas-Wallscheid, N. *et al.* Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. *Cell Stem Cell* **15**, 507-522 (2014).

30. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* (2018).

31. Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).

32. Saelens, W., Cannoodt, R., Todorov, H. & Saeys, Y. A comparison of single-cell trajectory inference methods. *Nat Biotechnol* **37**, 547-554 (2019).

33. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502 (2019).

34. Rodriguez-Fraticelli, A.E. *et al.* Clonal analysis of lineage fate in native haematopoiesis. *Nature* **553**, 212-216 (2018).

35. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821 (2019).