Human Bone Marrow Assessment by Single Cell RNA Sequencing, Mass Cytometry and Flow Cytometry

Short Title: Single Cell Analyses of Human Bone Marrow

Karolyn A. Oetjen[1], Katherine E. Lindblad[1], Meghali Goswami[1], Gege Gui[1], Pradeep K. Dagur[2], Catherine Lai[1], Laura W. Dillon[1], J. Philip McCoy[2], Christopher S. Hourigan[1]*

[1]Laboratory of Myeloid Malignancies and [2]Flow Cytometry Core, National Heart Lung and Blood Institute, 10 Center Drive, Bethesda, Maryland, USA

* To whom correspondence should be addressed:

Christopher S. Hourigan
Laboratory of Myeloid Malignancies
Hematology Branch
National Heart, Lung and Blood Institute
Room 10CRC 5-5130, 10 Center Drive
Bethesda, Maryland, 20814-1476, USA
hourigan@nih.gov

Abstract:	150 words
Text:	2469 words
References:	39
Figures:	3
Tables:	1

Abstract

New techniques for single-cell analysis have led to insights into hematopoiesis and the immune system, but the ability of these techniques to cross-validate and reproducibly identify the biological variation in diverse human samples is currently unproven. We therefore performed a comprehensive assessment of human bone marrow cells using both single-cell RNA sequencing and multiparameter flow cytometry from twenty healthy adult human donors across a broad age range. These data characterize variation between healthy donors as well as age-associated changes in cell population frequencies. Direct comparison of techniques revealed discrepancy in the quantification of T lymphocyte and natural killer cell populations. Orthogonal validation of immunophenotyping using mass cytometry demonstrated good correlation with flow cytometry. Technical replicates using single-cell RNA sequencing matched robustly, while biological replicates showed variation. Given the increasing use of single-cell technologies in translational research, this resource serves as an important reference dataset and highlights opportunities for further refinement.

Key Words

hematopoiesis, bone marrow, high-dimensional, single-cell transcriptomics, mass cytometry, CyTOF, bone marrow, reproducibility

Introduction

New technologies for characterizing cell populations are being implemented to more deeply describe the cell surface receptor phenotype and gene transcriptional signature at the single cell level (1, 2). Benefits of single cell approaches include examination of heterogeneity within the sample, and the most recent advances permit use of samples with very limited cell numbers for high dimensional characterization of cell surface phenotype or transcriptome. Single cell RNA sequencing (scRNAseq) has been used to elucidate hematopoietic differentiation (3-5) and immune cell subsets (6) including dendritic cells and monocytes (7), and innate lymphoid cells (8). Mass cytometry has been applied to the study of tissue-infiltrating immune cells (e.g. melanoma (9), renal cell (10), lung (11), and breast (12) cancers).

Expanding these new single cell approaches to patient samples requires a clear understanding of their correlation with established techniques, including flow cytometry. In order to facilitate and validate analysis of large databases of scRNAseq we set out to provide a data set of human bone marrow analyzed by both scRNAseq and deep immunophenotyping. Our reference cohort includes a broad range of donor ages in recognition of age-related variation in the healthy population.

Materials and Methods

Bone Marrow Aspirate Collection

Healthy volunteers were recruited for bone marrow aspirate collection at the National Institutes of Health. This research was approved by the National Heart, Lung and Blood Institute Institutional Review Board, and all participants provided oral and written informed consent. Using standard operating procedures, mononuclear cells from bone marrow aspirates were isolated using Ficoll density gradient separation and

cryopreserved in 90% FBS/ 10% DMSO for storage in liquid nitrogen. Assays were performed as listed in Table 1 using matched cryopreserved vials from each donor.

Single cell RNA Sequencing

scRNAseq was performed using 10X Genomics Single Cell 3' Solution, version 2 according to manufacturer's instructions (protocol rev A). Libraries were sequenced on HiSeq3000 and analyzed using Cell Ranger V2.0.0 (10X Genomics). Quality control metrics were used to select cells with mitochondrial gene percentage less than 8% and at least 500 genes detected. Samples were analyzed using Seurat (www.satijalab.org/seurat) using canonical correlation analysis with Louvain clustering, and visualized by t-distributed stochastic neighbor embedding (tSNE) (31). Developmental trajectories were created using Monocle versions 2 and 3 (32-34), the latter using Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) (35).

Flow cytometry

BMMCs were thawed in RPMI-1640 (Gibco) with 10% FBS and resuspended in cell staining buffer. Benzonase nuclease (Sigma Aldrich, catalog #E1014-25KU) was added for some samples during thawing to minimize cell clumping. Cells were blocked with Human TruStain FcX Fc receptor blocking solution (Biolegend, catalog #422302) and stained with antibodies listed in Table S1 followed by LIVE/DEAD Fixable Yellow stain (Life Technologies Corporation, Grand Island, NY) and fixation with 1% formaldehyde. Data were acquired with a Becton–Dickinson LSRFortessa (BD, San Jose, CA, USA) equipped with five lasers (355, 407, 488, 532 and 633 nm wavelengths) and 22 PMT detectors using DIVA 8 software using the high throughput sampler (BD) system at a flow rate of 2.5ul/sec in a 96 well U bottom tissue culture plate. Compensation controls were performed using single color staining of compensation beads (BD), and daily quality assurance was performed using Cytometer setup and Tracking beads (BD) as

per manufacturer's recommendation along with 1 peak rainbow Beads (BD) and 8 peak beads (Spherotec)(36, 37). Post-acquisition analysis was performed using Flowjo 9.9.6 (Treestar Inc., San Carlos, CA, USA). Analysis excluded debris and doublets using light scatter measurements, and dead cells by live/dead stain. Gating strategies used to identify immune cell subsets are provided in Figure S2.

Mass cytometry

Thawed BMMCs were stained for 37 markers using the MaxPar Complete Human T Cell Immuno-Oncology Panel Set (Fluidigm), according to manufacturer instructions. Briefly, cells were thawed, washed, incubated with cisplatin cocktail for viability, fixed in 1.6% formaldehyde and permeabilized. Cells were then stained with the antibody cocktail, incubated with intercalation solution, mixed with EQ4 element beads and acquired with a Helios mass cytometer (Fluidigm). Gating and viSNE analysis (38) were performed using Cytobank (cytobank.org). Initial analysis excluded doublets using DNA content and non-viable cells using cisplatin. CD45-positive cells were gated for viSNE analysis of 100,000 total events from all analyzed samples.

Bulk RNA sequencing

RNA was harvested from thawed cell vials of BMMCs using AllPrep kits (QIAGEN). Libraries were prepared using TruSeq Stranded Total RNA Sample Preparation Kit (Illumina) with 1ug of RNA input. Sequencing was performed by paired-end 75 nt on Illumina HiSeq 3000. Fastq files were mapped to using *kallisto*, and gene counts were tabulated using *tximport*. Deconvolution was performed using Xcell v1.1 (xcell.ucsf.edu)(16) or Cibersort using LM22 gene signature and 100 permutations (cibersort.stanford.edu)(17).

Data analysis and statistics

Data were analysis, visualization and statistical comparisons were performed R (cran.r-project.org). Bland-Altman analysis (39) was implemented in the BlandAltmanLeh package v0.3.1.

Data availability

FCS files for flow cytometry and mass cytometry data sets have been deposited in FlowRepository (*accession#*). Single cell RNA sequencing and bulk RNA sequencing datasets have been deposited in Gene Expression Omnibus (GEO) (*accession#*).

Results

Healthy donor characteristics

Twenty healthy volunteers were recruited for bone marrow aspiration procedures. The cohort consisted of 10 males and 10 females with ages ranging 24-84 years old and median age of 57 years. A second bone marrow aspiration was performed for two donors (Ck, Sk) (*"biological replicate"*) either 2 or 5 months after their first aspiration respectively. Cryopreserved cells from all twenty donors were analyzed by droplet-based scRNAseq and flow cytometry, and additional cryopreserved vials for eight donors were analyzed by mass cytometry for T cell phenotyping, as well as bulk RNA sequencing, as summarized in Table 1.

Single cell RNA sequencing

Droplet-based scRNAseq of bone marrow mononuclear cells (BMMCs) for all donor samples was performed with goal minimum sequencing depth of 50,000 reads/cell and detected a mean of 880 genes/cells (range 575-1,390 gene/cell, Table 1). Greater than 90,000 cells were captured; using quality filters of at least 500 genes per cell and less than 8% mitochondrial RNA content, 76,645 cells were analyzed in the final analysis.

To account for sample variations between donors, alignment of all samples was performed in Seurat using canonical correlation analysis (CCA) then visualized using t-

distributed stochastic neighbor embedding (t-SNE). Cell clusters were distinguished using the Louvain clustering algorithm implemented in Seurat. Compiled analysis of all donor cells is annotated in Figure 1A with the contribution of each individual donor displayed in Figure S1. All major previously identified populations of bone marrow mononuclear cells were present in the clustered scRNAseq analysis.

Single cell trajectory analysis was performed using Monocle 3. As there were potentially multiple disjoint trajectories in this complex dataset containing a large number of cells, UMAP was used for dimension reduction. The resulting development trajectories clearly display the major lymphoid, myeloid and erythroid lineages of hematopoiesis with correct ordering of developmental stages (Figure 1B). Trajectories of erythroid and myeloid lineages could also be created using an earlier, well validated, version of this software (Monocle2, see Figure S1) and were consistent with those observed for the full dataset.

Annotation of cell cluster identities was determined using a panel of canonical gene expression, with the expression patterns for a subset of these genes displayed in Figure 1C. Analysis of each donor sample individually using principal component analysis (PCA) in Seurat revealed suboptimal quantification of frequencies of some transcriptionally similar cell subsets, including those annotated as effector T cells and NK cells. Such clusters were typically well delineated for each individual sample when using CCA in the context of the entire dataset (Figure S1).

A potential use of scRNAseq is to compare across two or more samples. To confirm the validity of scRNAseq for this approach, assay reproducibility was determined by preparing duplicate, side-by-side libraries from cells thawed from the same cryopreserved vial, for a total of three cryopreserved samples. Cell subtype quantification for each of these technical replicate pairs matched robustly (Figure 1D). The optimum number of cells required to identify, using scRNAseq, sub-populations

174 within a heterogenous samples remains an area of interest (13). Technical replicates
175 ranged from 1,138 to 6,692 cells from the same sample (Table 1).

176 ***Flow cytometry***

177 13-color flow cytometry using five customized panels ("T, B, NK, Mono and DC", see
178 Table S1) designed to allow deep immunophenotyping of the predominant cell
179 populations found in human bone marrow was performed on all samples. Approximately
180 1 million cells were stained for each panel, and a median of 196,000 CD45 positive
181 events collected (25^{th}-75^{th} percentile: 100,000-278,000 events). Gating strategy is
182 shown in Table S2. Most frequent cell subtype populations observed were, in order, T
183 cells, monocytes, B cells, natural killer cells (NK), dendritic cells (DC) and hematopoietic
184 stem/progenitor cells (HSPC) (see Figure S2).

185 Paired analysis of the same sample by both transcriptome and cell surface
186 phenotype offers a powerful opportunity to compare cell population frequencies
187 determined by these methods. The proportion of major cell populations is summarized
188 for scRNAseq and flow cytometry in Figure 2A. Sample-by-sample correlations for all of
189 these populations are shown in Figure 2B. It is well established that the T memory cell
190 population increases with increasing age in humans, likely due to response to viral
191 infection (in particular CMV), and this trend was reproduced in our cohort using both
192 scRNAseq and flow cytometry (Figure S2)(14). Two subjects had a second bone
193 marrow aspiration performed at either 2 or 5 months after their first aspiration. These
194 biological replicates showed good concordance by flow cytometry but showed variation
195 by scRNAseq particularly in lower frequency cell subsets, likely from sampling error
196 (Figure S2)

197 While concordance between these two modalities was generally good, it appeared
198 that T cell frequency was elevated, and NK cell frequency decreased in scRNAseq as
199 compared with flow cytometry. This led to a more detailed examination of T cells

subsets and orthogonal validation of cell surface immunophenotyping using a third single cell modality.

Mass cytometry

In order to more deeply characterize immune populations within healthy bone marrow, and to validate our flow cytometry results, T cell phenotyping was performed by mass cytometry using a 37-marker panel for a subset of eight donors. Using Cytobank software, CD45-positive cells were visualized using viSNE across the panel of markers (Figures 3A and S3). Correlation between mass cytometry and flow cytometry for CD4- and CD8-positive T lymphocyte subsets was good as shown in Figure 3B.

To further compare mass cytometry and flow cytometry with scRNAseq of T cell populations, the frequencies of T cell subsets for this cohort of eight donors were determined using all three of these methods, shown in Figure 3C with sample correlations reported in Figure 3D. Comparing frequencies of T cell populations between mass cytometry and scRNAseq confirmed a small but persistent skewing in the identification of NK and T cells. Using Bland-Altman calculations, the mean difference between scRNAseq and mass cytometry for T cells was -6.5% (95% CI: -29% to 16%) and for NK cells was 3.2% (95% CI: -1.1% to 7.6%).

CD8 cytotoxic T cells and NK cells are known to have substantial overlap at the transcriptome level (15). To better understand systemic bias in the frequency of NK or T cells identified, we confirmed that overlapping gene signatures are found in clusters annotated as NK or T cells in this scRNAseq data set (Figure S4). The reasons for this bias are likely however multifactorial.

Bulk RNA-sequencing

Analysis of bulk sample RNA expression has been used to attempt to deconvolute the proportion of each cell subtype in human tissues (16, 17). Finally, as an additional resource, stranded whole transcriptome sequencing of RNA isolated from

thawed BMMCs was performed on samples from all eight subjects for which mass and flow cytometry and single cell RNA sequencing was available. Initial analysis using deconvolution algorithms that attempt to predict the proportion of cell subpopulations is shown in Table S4.

Discussion

Changes in the immune system (14) and hematopoiesis (18) occur during human aging. Using an unbiased approached based on unsorted human BMMCs, we describe the major cell populations of healthy human bone marrow from a cohort of donors over a wide range of adult age by multiple high-dimensional single cell techniques. This resource serves as a complement to existing data sets that have consisted primarily of younger donors without associated paired immunophenotyping. Our data set provides a resource of scRNAseq, flow cytometry and mass cytometry data for healthy control cohorts across the full range of adulthood providing not only cell population frequencies and characteristics, but also highlighting individual variation in human cohorts.

Using scRNAseq of a total of over 76,000 cells from 20 healthy donors, all the major bone marrow mononuclear populations are identified, and overall population frequencies are comparable to flow cytometry of the matched samples. A primary limitation is distinguishing cell populations such as NK cells and CD8+ effector T cells, which have overlapping transcriptional programs with a small number of distinguishing genes captured by droplet-based scRNAseq. To overcome this limitation and provide additional reference data beyond previous reports of major healthy bone marrow populations by flow cytometry (19) and mass cytometry (20, 21), we used the strength of mass cytometry for high resolution of T cell subpopulations (22), both to validate our flow cytometry results and provide quantification of rare T cell subpopulations within healthy human bone marrow.

As a data resource, these high-dimensional approaches to bone marrow characterization add valuable information on transcriptional and cell surface marker co-expression. The growing number of bioinformatics tools for mass cytometry (23) and scRNAseq (24, 25) will benefit from reference data sets for validation and integrated comparison across techniques. Future opportunities for integrating these data sets include droplet-based sequencing with oligonucleotide-tagged antibodies, including CITE-Seq (26), REAP-Seq (27), and AbSeq (28), which can be compared to this reference set of cell surface protein and transcriptome expression. As techniques (29) and repositories (30) of high-dimensional single cell human data sets are expanded, validating the observed cell identities will be a critical aspect of interpreting large data set analysis.

Additional aliquots of bone marrow aspirate from this cohort together with paired blood samples, that were not yet analyzed, have been stored. Should transformative technologies emerge over the next few years we would be willing, subject to relevant technology transfer and clinical regulatory approvals, to share remaining samples with academic investigators for additional benchmarking and validation. In summary, this resource provides a reference dataset for cell populations in healthy human bone marrow across a wide age range as assessed by multiple single-cell approaches. We show that scRNAseq quantification of marrow-resident cell populations has good concordance with immunophenotyping by flow and mass cytometry with some discrepancies in T and NK subsets. We hope this unique combined dataset will prove useful both to those seeking to refine or innovate bioinformatic algorithms for scRNAseq data and also to those investigators hoping to apply these powerful single-cell technologies in their own research.

Acknowledgments: This work was supported by the Intramural Research Program of the National Heart, Lung, and Blood Institute of the National Institutes of Health. We appreciate the technical expertise of Brian Sellers and the NIH Center for Human Immunology, Yan Luo, Yuesheng Li and the NHLBI DNA Sequencing Core, Alan Hoofring of NIH Medical Arts and NIH High-Performance Computing. We thank Sheenu Sheela, Blair DeStefano, Janet Valdez and NHLBI research nurses for bone marrow aspirate procedures. The authors would like to thank Neal Young, Cindy Dunbar (both NIH) and Elizabeth Jaffee (Johns Hopkins) for reading and comments, and Efthymia Papalexi and Rahul Satija (both New York Genome Center, NY) for help with scRNAseq analysis and comments.

Authorship Contributions: KO performed experiments, analyzed data and wrote the manuscript; KL and MG performed flow cytometry experiments and analyzed data; GG and LD analyzed data; PD and PM designed, supervised and analyzed flow cytometry experiments, CL coordinated donor recruitment; CH designed experiments, analyzed data and wrote the manuscript. All authors reviewed the final manuscript.

Declaration of Interests: CSH receives research funding from Merck Sharpe & Dohme and SELLAS Life Sciences Group AG. The other authors declare no relevant competing financial interests.

References

1. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2015;161(5):1202-14.

2. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nature communications. 2017;8:14049.

3. Buenrostro JD, Corces MR, Lareau CA, Wu B, Schep AN, Aryee MJ, et al. Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. Cell. 2018;173(6):1535-48.e16.

4. Psaila B, Barkas N, Iskander D, Roy A, Anderson S, Ashley N, et al. Single-cell profiling of human megakaryocyte-erythroid progenitors identifies distinct megakaryocyte and erythroid differentiation pathways. Genome biology. 2016;17:83.

5. Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, et al. Human haematopoietic stem cell lineage commitment is a continuous process. Nature cell biology. 2017;19(4):271-81.

6. Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. Nature reviews Immunology. 2018;18(1):35-45.

7. Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science (New York, NY). 2017;356(6335).

8. Bjorklund AK, Forkel M, Picelli S, Konya V, Theorell J, Friberg D, et al. The heterogeneity of human CD127(+) innate lymphoid cells revealed by single-cell RNA sequencing. Nature immunology. 2016;17(4):451-60.

9. Nirschl CJ, Suarez-Farinas M, Izar B, Prakadan S, Dannenfelser R, Tirosh I, et al. IFNgamma-Dependent Tissue-Immune Homeostasis Is Co-opted in the Tumor Microenvironment. Cell. 2017;170(1):127-41.e15.

10. Chevrier S, Levine JH, Zanotelli VRT, Silina K, Schulz D, Bacac M, et al. An Immune Atlas of Clear Cell Renal Cell Carcinoma. Cell. 2017;169(4):736-49.e18.

11. Lavin Y, Kobayashi S, Leader A, Amir ED, Elefant N, Bigenwald C, et al. Innate Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell Analyses. Cell. 2017;169(4):750-65.e17.

12. Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S, et al. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. Cell. 2018;174(5):1293-308.e36.

13. Zhang MJ, Ntranos V, Tse D. One read per cell per gene is optimal for single-cell RNA-Seq. bioRxiv. 2018.

14. Nikolich-Zugich J. The twilight of immunity: emerging concepts in aging of the immune system. Nature immunology. 2018;19(1):10-9.

15. Hidalgo LG, Einecke G, Allanach K, Halloran PF. The transcriptome of human cytotoxic T cells: similarities and disparities among allostimulated CD4(+) CTL, CD8(+) CTL and NK cells. American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons. 2008;8(3):627-36.

16. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome biology. 2017;18(1):220.

17. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nature methods. 2015;12(5):453-7.

18. Pang WW, Price EA, Sahoo D, Beerman I, Maloney WJ, Rossi DJ, et al. Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(50):20012-7.

19. Nombela-Arrieta C, Manz MG. Quantification and three-dimensional microanatomical organization of the bone marrow. Blood advances. 2017;1(6):407-16.

20. Bendall SC, Simonds EF, Qiu P, Amir el AD, Krutzik PO, Finck R, et al. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. Science (New York, NY). 2011;332(6030):687-96.

21. Levine JH, Simonds EF, Bendall SC, Davis KL, Amir el AD, Tadmor MD, et al. Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell. 2015;162(1):184-97.

22. Wong MT, Ong DE, Lim FS, Teng KW, McGovern N, Narayanan S, et al. A High-Dimensional Atlas of Human T Cell Diversity Reveals Tissue-Specific Trafficking and Cytokine Signatures. Immunity. 2016;45(2):442-56.

23. Kimball AK, Oko LM, Bullock BL, Nemenoff RA, van Dyk LF, Clambey ET. A Beginner's Guide to Analyzing and Visualizing Mass Cytometry Data. Journal of immunology (Baltimore, Md : 1950). 2018;200(1):3-22.

24. Zappia L, Phipson B, Oshlack A. Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. PLoS computational biology. 2018;14(6):e1006245.

25. Becht E, Dutertre C-A, Kwok IWH, Ng LG, Ginhoux F, Newell EW. Evaluation of UMAP as an alternative to t-SNE for single-cell data. bioRxiv. 2018.

26. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, et al. Simultaneous epitope and transcriptome measurement in single cells. Nature methods. 2017;14(9):865-8.

27. Peterson VM, Zhang KX, Kumar N, Wong J, Li L, Wilson DC, et al. Multiplexed quantification of proteins and transcripts in single cells. Nature biotechnology. 2017;35(10):936-9.

28. Shahi P, Kim SC, Haliburton JR, Gartner ZJ, Abate AR. Abseq: Ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding. Scientific reports. 2017;7:44447.

29. Cheung P, Vallania F, Warsinske HC, Donato M, Schaffert S, Chang SE, et al. Single-Cell Chromatin Modification Profiling Reveals Increased Epigenetic Variations with Aging. Cell. 2018;173(6):1385-97.e14.

30. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, et al. The Human Cell Atlas. eLife. 2017;6.

31. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nature biotechnology. 2018;36(5):411-20.

32. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nature biotechnology. 2014;32(4):381-6.

33. Qiu X, Hill A, Packer J, Lin D, Ma YA, Trapnell C. Single-cell mRNA quantification and differential analysis with Census. Nature methods. 2017;14(3):309-15.

34. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. Nature methods. 2017;14(10):979-82.

35. McInnes L, Healy J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv:180203426. 2018.

36. Perfetto SP, Ambrozak D, Nguyen R, Chattopadhyay PK, Roederer M. Quality assurance for polychromatic flow cytometry using a suite of calibration beads. Nature protocols. 2012;7(12):2067-79.

37. Wang L, Hoffman RA. Standardization, Calibration, and Control in Flow Cytometry. Current protocols in cytometry. 2017;79:1.3.1-.3.27.

38. Amir el AD, Davis KL, Tadmor MD, Simonds EF, Levine JH, Bendall SC, et al. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nature biotechnology. 2013;31(6):545-52.

39. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet (London, England). 1986;1(8476):307-10.

Figures and Tables

Figure 1. Single cell RNA sequencing of healthy bone marrow cells. (A) Cluster identification visualized using t-distributed stochastic neighbor embedding (t-SNE). (B) Single cell trajectory analysis using UMAP/Monocle 3. Color as in Figure 1A. (C) Examples of canonical gene expression used for annotation. (D) Reproducibility of technical replicates for single cell RNA sequencing. Linear regression line displayed in grey.

Figure 2. Comparison of single cell RNA sequencing and flow cytometry assessment of bone marrow cell type population frequencies. (A) Frequencies for major cell populations in human bone marrow shown for single cell RNA sequencing and flow cytometry. (B) Individual sample comparisons by scatter plot for each cell population. All population comparisons are shown in background in grey. Population frequencies are reported using denominator of all CD45 positive cells.

Figure 3. Comparison of single cell RNA sequencing, mass cytometry and flow cytometry assessment of T lymphocyte frequencies in human bone marrow. (A) Mass cytometry for phenotyping of T cell populations visualized using viSNE analysis with expression of key markers shown (B) Comparison of cell frequencies for each donor determined by mass cytometry and flow cytometry. (C) T cell frequencies for cell populations identified by mass cytometry, flow cytometry and single cell RNA sequencing. (D) Individual sample comparisons by scatter plot for each cell population. All population comparisons are shown in background in grey.

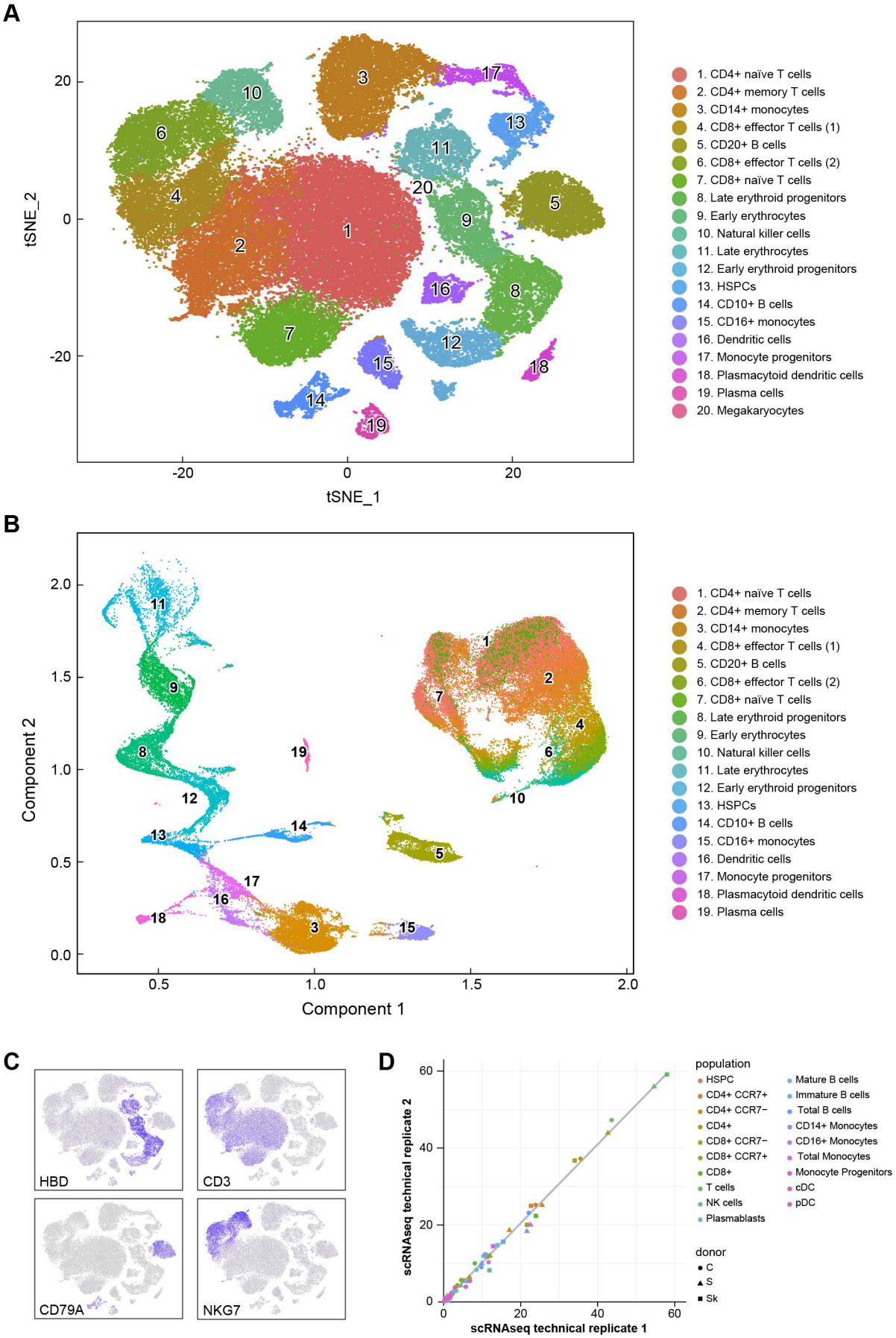
Table 1. Heathy volunteer sex and age at time of bone marrow aspiration. Biological replicate time points for a second longitudinal bone marrow aspirate from the same volunteers are shown within grey boxes. Assays from matched cryopreserved bone marrow mononuclear cell vials are indicated. Single cell RNA sequencing cell counts and sequencing depth (reads per cell and genes per cell) are listed for each donor and replicate.

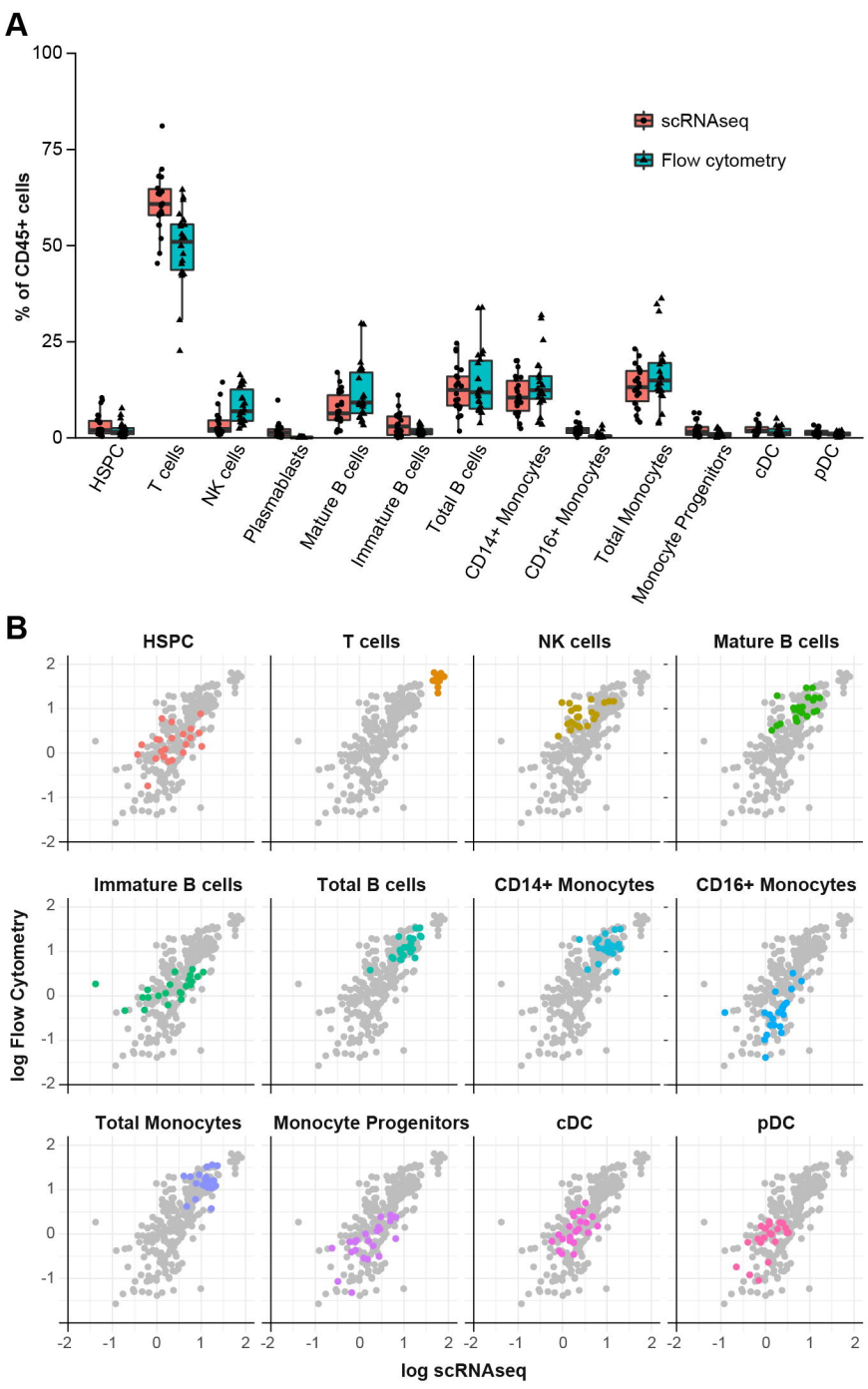
Figures S1-S4 and Tables S1-S4 can be found in the supplementary information.

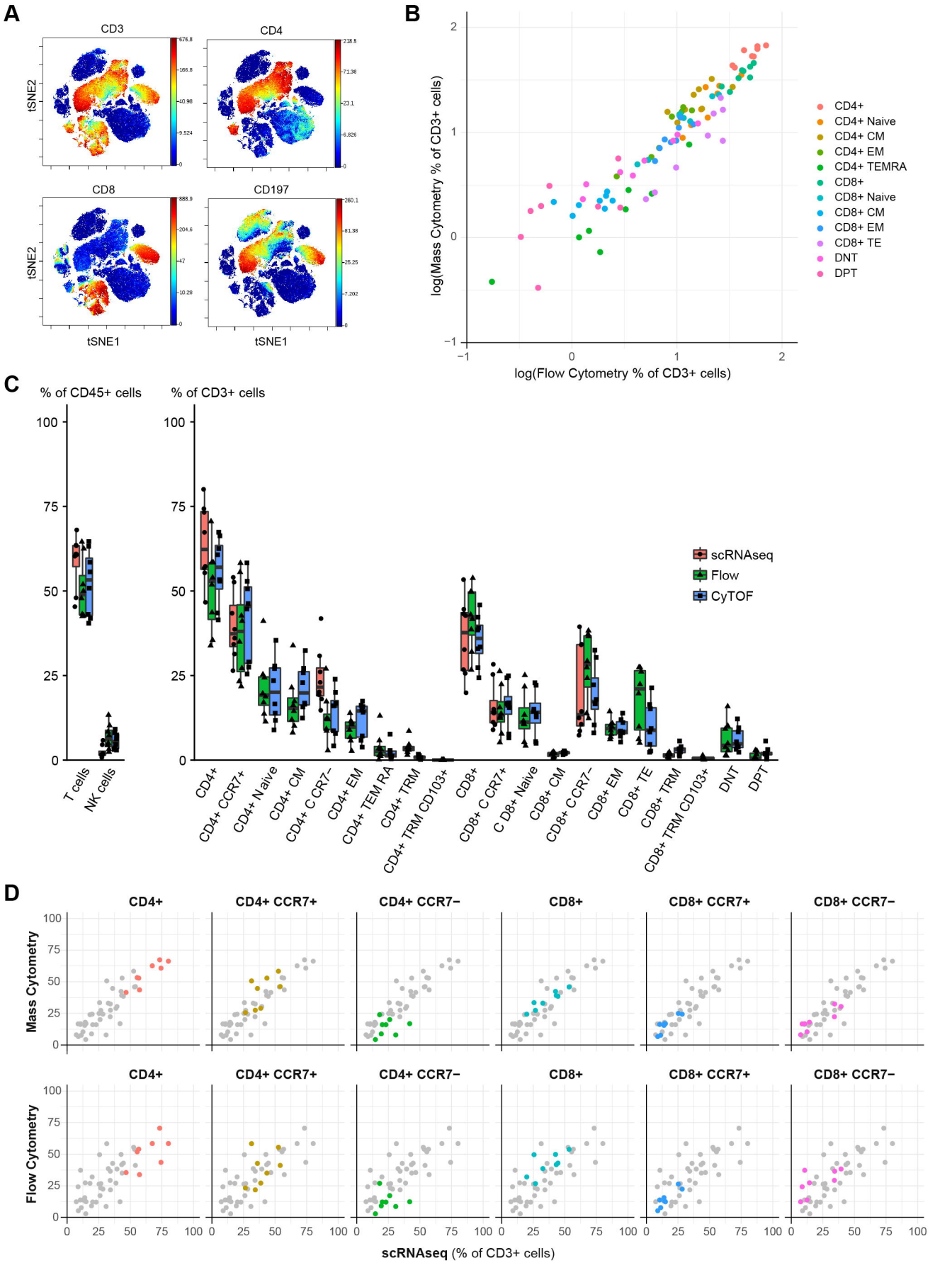
Table 1

Sample	Gender	Age	Flow Cytometry	Mass Cytometry	Bulk RNA	scRNAseq	Technical Replicate	Cells	Reads/Cell	Genes/Cell
T	F	24	x	x	x	x		4,293	88,461	1,000
W	F	28	x			x		3,643	47,132	649
E	M	30	x			x		3,939	69,061	575
R	M	31	x			x		3,593	131,257	669
F	F	41	x			x		3,746	75,712	846
J	F	43	x	x	x	x		3,446	108,679	970
U	F	46	x	x	x	x		4,118	85,295	1,143
B	M	47	x	x	x	x		3,293	99,593	794
H	F	50	x	x	x	x		5,013	63,164	883
O	M	50	x	x	x	x		4,516	47,778	851
Sk	F	55	x			x	Sk1	1,138	323,589	823
						x	Sk2	4,726	163,732	820
S		56	x			x	S1	2,437	113,302	1,089
						x	S2	2,367	83,847	1,163
L	M	57	x			x		4,548	67,199	950
P	F	58	x			x		3,383	223,652	1,390
G	M	58	x			x		4,283	89,208	667
A	F	59	x	x	x	x		2,994	159,501	1,303
Ck	F	59	x			x		1,052	349,511	761
C		60	x	x	x	x	C1	3,556	62,645	692
						x	C2	3,136	58,675	692
M	M	60	x			x		3,964	92,780	875
Q	M	66	x			x		1,700	126,143	702
N	M	67	x			x		4,522	110,195	881
K	M	84	x			x		7,247	43,872	879

Sk and S are two samples taken from the same donor at different times (*ie: biological replicates*).
Ck and C two samples taken from the same donor at different times (*ie: biological replicates*).
Samples Sk, S and C were split and used for technical replicates of single cell RNA sequencing.


A
tSNE_2
tSNE_1
B
Component 2
Component 1
1. CD4+ naïve T cells
2. CD4+ memory T cells
3. CD14+ monocytes
4. CD8+ effector T cells (1)
5. CD20+ B cells
6. CD8+ effector T cells (2)
7. CD8+ naïve T cells
8. Late erythroid progenitors
9. Early erythrocytes
10. Natural killer cells
11. Late erythrocytes
12. Early erythroid progenitors
13. HSPCs
14. CD10+ B cells
15. CD16+ monocytes
16. Dendritic cells
17. Monocyte progenitors
18. Plasmacytoid dendritic cells
19. Plasma cells
20. Megakaryocytes
C
HBD
CD3
CD79A
NKG7
D
scRNAseq technical replicate 2
scRNAseq technical replicate 1
population
HSPC
CD4+ CCR7+
CD4+ CCR7−
CD4+
CD8+ CCR7−
CD8+ CCR7+
CD8+
T cells
NK cells
Plasmablasts
Mature B cells
Immature B cells
Total B cells
CD14+ Monocytes
CD16+ Monocytes
Total Monocytes
Monocyte Progenitors
cDC
pDC
donor
C
S
Sk


A
100
75
50
25
0
% of CD45+ cells
scRNAseq
Flow cytometry
HSPC
T cells
NK cells
Plasmablasts
Mature B cells
Immature B cells
Total B cells
CD14+ Monocytes
CD16+ Monocytes
Total Monocytes
Monocyte Progenitors
cDC
pDC
B
HSPC
T cells
NK cells
Mature B cells
Immature B cells
Total B cells
CD14+ Monocytes
CD16+ Monocytes
Total Monocytes
Monocyte Progenitors
cDC
pDC
log Flow Cytometry
log scRNAseq
2
1
0
-1
-2


A
CD3
CD4
CD8
CD197
tSNE2
tSNE1
B
log(Mass Cytometry % of CD3+ cells)
log(Flow Cytometry % of CD3+ cells)
CD4+
CD4+ Naive
CD4+ CM
CD4+ EM
CD4+ TEMRA
CD8+
CD8+ Naive
CD8+ CM
CD8+ EM
CD8+ TE
DNT
DPT
C
% of CD45+ cells
% of CD3+ cells
scRNAseq
Flow
CyTOF
T cells
NK cells
CD4+
CD4+ CCR7+
CD4+ Naive
CD4+ CM
CD4+ C CR7−
CD4+ EM
CD4+ TEM RA
CD4+ TRM
CD4+ TRM CD103+
CD8+
CD8+ C CR7+
C D8+ Naive
CD8+ CM
CD8+ C CR7−
CD8+ EM
CD8+ TE
CD8+ TRM
CD8+ TRM CD103+
DNT
DPT
D
CD4+
CD4+ CCR7+
CD4+ CCR7−
CD8+
CD8+ CCR7+
CD8+ CCR7−
Mass Cytometry
Flow Cytometry
scRNAseq (% of CD3+ cells)