Hematopoiesis at single cell resolution spanning human development and maturation

Hojun Li^{1,2,3,4,*}, Jidefor Ezike^{5,6}, Anton Afanassiev⁷, Laura Greenstreet⁷, Stephen Zhang⁷, Jennifer Whangbo^{1,2,3}, Vincent L. Butty⁸, Enrico Moiso⁴, Guinevere G. Connelly⁴, Vivian Morris², Dahai Wang², George Q. Daley^{2,3}, Salil Garg^{4,9}, Stella T. Chou^{10,11}, Aviv Regev^{6,12}, Edroaldo Lummertz da Rocha¹³, Geoffrey Schiebinger⁷, and R. Grant Rowe^{1,2,3,*}

¹Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA

²Division of Hematology/Oncology, Boston Children's Hospital, Boston, MA

³Harvard Medical School, Boston, MA

⁴Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA

⁵Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, MA

⁶Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA

⁷Department of Mathematics, University of British Columbia, Vancouver, BC

⁸Barbara K. Ostrom Bioinformatics Facility, Integrated Genomics and Bioinformatics Core of the Koch Institute

⁹Department of Pathology, Massachusetts General Hospital, Boston, MA

¹⁰Division of Hematology, Children's Hospital of Philadelphia, Philadelphia, PA

¹¹Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA

¹²Genentech, South San Francisco, CA

¹³Departments of Microbiology, Immunology and Parasitology, Federal University of Santa Catarina, Florianopolis, Brazil

*co-corresponding authors, contact: hojunli@mit.edu (H.L.) and grant_rowe@dfci.harvard.edu (R.G.R.)

Summary: Single cell RNA sequencing reveals that the mechanisms of human hematopoietic stem and progenitor cell (HSPC) fate commitment change over a lifetime from gestation to mature adulthood.

Abstract:

Hematopoiesis is a process of constitutive regeneration whereby hematopoietic stem and progenitor cells (HSPCs) replenish mature blood cells. During maturation and aging, HSPCs shift their output to support the demands of prenatal development and postnatal maturation both at homeostasis and in response to stress. How HSPC ontogeny changes throughout life is unknown; studies to date have largely focused on specific individual ages, particularly at single cell resolution. Here, we performed single cell RNA-seq of human HSPCs from early prenatal development into mature adulthood. We observed shifts in HSPC transcriptional states and differentiation trajectories over time. We identified age-specific gene expression patterns throughout human maturation and developed methods for identifying, prospectively purifying, and functionally validating age-specific HSC states. Together, our findings define the temporal maturation of human HSPCs and uncover principles applicable to age-biased blood diseases.

Main Text:

Over a lifetime, stem cells must continuously adjust their prioritization of self-renewal and differentiation to execute tightly choreographed programs of development and maturation (1). Although stem cell ontogenies have been finely mapped at specific stages of human life, how stem and progenitor cell ontogeny changes over the course of a lifetime remains a fundamental, but unanswered question (2-9). To meet age-specific physiologic demands, the human hematopoietic system adjusts the output of mature cells capable of oxygen transport, hemostasis, innate and adaptive immunity, and tissue repair; however, the molecular regulation of these age-dependent alterations is not well understood.

Since lineage destiny is transcriptionally encoded within HSPCs, we performed single cell RNA-seq (scRNA-seq) of human HSPCs to reconstruct hematopoiesis over the course of human maturation. We procured CD34⁺ HSPCs at 12 time points across human development, from 10 weeks post-gestation to 45 years of age (1-2 donors per time point) and used the InDrop platform to profile individual transcriptomes (**Fig. 1A**)(10). After filtering of doublets, non-hematopoietic cells, and cells with poor coverage (see methods), we retained 38,873 HSPCs with high quality profiles. We performed dimensionality reduction with principal component analysis (PCA), followed by embedding via Universal Manifold Approximation and Projection (UMAP) and partition-based graph abstraction force directed layout (FDL) to visualize individual ages and the overall dataset (**Fig. 1B,I, fig. S1**)(11, 12). Expression of lineage specific marker genes revealed clear transcriptional priming of lymphoid and myeloerythroid lineages (**Fig. 1B-H, fig. S2**)(8, 13, 14).

We identified 21 distinct HSPC states by unsupervised clustering (see methods) that vary in abundance over time but are detectable at all time points (**Fig. 2A,C,F, fig. S3**). We used the singleCellNet algorithm to score each cluster against an existing human adult bone marrow scRNA-seq dataset (**Fig. 2B**)(14, 15); examining them in the context of the FDL and UMAP, we hierarchically ordered and annotated HSPC states spanning from the most undifferentiated HSCs

and multipotent progenitors (MPPs) to unilineage-restricted progenitors (**Fig. 2A**)(12). Cells in HSC and MPP states were located centrally in both UMAP and FDL space with branches terminating in lineage-restricted progenitors (**Fig. 2A, fig. S3**).

We found that the abundance of HSPC states and usage of differentiation trajectories changes over time. Early gestation fetal liver (FL, 10-13 weeks) prioritized trajectories toward myeloid output; later in gestation prioritization shifted toward HSC expansion (**Fig. 2F, fig. S4**). Erythroid production was maintained throughout gestation, supportive of growth within the hypoxic uterus (**Fig. 2C,D,F, fig. S4**). In childhood, lymphoid progenitors are relatively abundant, but diminish with age; myeloid and erythroid output increases over time as recapitulated in clonogenesis assays (**Fig. 2C-F, fig. S4**). The pediatric wave of lymphopoiesis occurs via differentiation through the Ly-MPP-2 state, which diminishes with age, while Ly-MPP-1 maintain steady-state lymphoid output throughout life (**Fig. 2C,F, fig. S4**). These overall patterns of lineage prioritization are supported by changes that occur in peripheral blood leukocyte populations across human life and the myeloid-biased hematopoiesis of mature adulthood (16, 17).

We next investigated the processes of age-specific lineage commitment using both Population Balance Analysis and an adaptation of Optimal Transport to systems in equilibrium (18-20). Both approaches model cell trajectories with the same diffusion-drift equation, which represents Waddington's landscape, and both produce the same output: an estimate of the lineage commitment of each cell, for each lineage (i.e. lineage fate probabilities). While these two methods differ in their underlying mathematical approaches, they produced reassuringly similar results. In both approaches we found high myeloid differentiation probability for early gestation fetal liver HSPCs, with a mid-gestation shift toward more balanced hematopoiesis; higher lymphoid probability predominated in childhood, shifting back to higher myeloid probability in adulthood (**Fig. 3A, fig. S5**). These analyses also distinguished age-biased fate probabilities within the HSC state subsets that appeared uncommitted in UMAP and FDL space (**fig. S6**).

To identify the mechanisms underlying changes in lineage commitment during human development and maturation, we searched for transcripts whose levels are associated with age-specific differentiation. Elastic Net semi-sparse regression of cells undergoing lineage commitment identified transcripts for each lineage whose levels either consistently predicted fate probabilities, or only predicted fate probabilities at certain ages (21). For each lineage, we identified the top 200 transcripts showing either consistent or age-specific prediction of fate commitment (**Table S1**). For example, during erythroid differentiation we observed expected expression patterns in temporally consistent (*GATA1*, *KLF1*) and variable (*HBG1*, *HBG2*, *HBB*) erythropoietic genes, as well as several novel age-associated factors (**Fig. 3B**). These data indicate that while certain factors control differentiation throughout life, differentiation programs are also modified with age.

In addition to age-dependent programming of normal hematopoietic development, many types of leukemia are biased toward certain ages (14, 22, 23). We used singleCellNet to classify blasts of childhood B-acute lymphoblastic leukemia (B-ALL) and adult acute myeloid leukemia (AML) profiled by scRNA-seq (**fig. S7**)(14, 22). In childhood B-ALL, we observed shifting classification from diagnosis to relapse, with several cases adopting more differentiated states, and notably one case recruiting the fetal-biased HSC-3 state (**fig. S7**). In poorly differentiated adult AMLs, activation of differentiation programs occurred with treatment (**fig. S7**). These data demonstrate potential application of our atlas of human HSPC maturation to pinpoint leukemic differentiation state, uncover activation of oncofetal programs, and novel potential mechanisms of therapeutic evasion via differentiation or dedifferentiation.

As age-dependent changes in stem cell states have yet to be described at single cell resolution in humans, we sought to determine whether such a phenomenon occurs in HSCs by querying for age-specific HSC states that have yet to be detected in prior seminal studies of narrow developmental time windows (24-26). We observed that the most undifferentiated HSCs exist in several states closely related in UMAP and FDL space, but with distinct transcriptomes

(**Fig. 2A, fig. S3, fig. S8A**). The subsets HSC-1, HSC-2, and HSC-5 were similar, enriched for ribosomal proteins and the important atypical homeodomain protein HOPX. Three subsets were particularly distinguishable: HSC-6 enriched for cell cycle genes, HSC-4 enriched for the histone acetyltransferase NAA40 and the D-type IL-17 receptor, and HSC-3. While most HSC states were present across all ages, HSC-3 was largely specific to prenatal development, increasing prenatally and becoming the dominant HSC state at 20 weeks' gestation before diminishing postnatally; consistent with overall heterogeneity of fetal HSCs (**Fig. 4A**)(27). HSC-3 expressed several transcripts suggestive of either a uniquely activated state *in vivo*, or the effect of tissue processing (28, 29). We therefore sought to identify a surface marker for prospective isolation of HSC-3. Differential expression analysis of transcripts encoding cell surface proteins highlighted CD69 as specifically expressed in HSC-3 (**Fig. 4B**). Since CD69 is known primarily as a marker of activated T-cells, we corroborated the existence of CD69⁺ HSC populations in three additional mid-gestation fetal liver samples and independently published human FL datasets (**fig. S9**)(30-33).

We next analyzed CD69 protein expression in fetal HSPCs. In freshly thawed human FL HSPCs, we did not detect appreciable CD69 expression in the undifferentiated Lineage⁻CD34⁺CD38⁻ compartment, possibly indicating loss of cell surface CD69 during processing (**fig. S8B**). However, upon mitogenic stimulation, we observed heterogeneity in expression of cell surface CD69 specifically in mid-gestation FL HSCs/MPPs, consistent with the observed heterogeneity in *CD69* transcript expression (**Fig. 4C-D, fig. S8B**). To determine whether CD69 expression distinguishes functionally distinct HSC/MPP subpopulations, following mitogenic stimulation we sorted CD69⁺ and CD69⁻ gestational week 20 or later FL Lineage⁻CD34⁺CD38⁻ HSPCs (**fig. S8C**) onto MS5 stroma cells cultured with cytokines as well as in the presence or absence of erythropoietin at limiting dilution to quantify clonogenic potential and single cell lineage outcomes. Across multiple donors, mid-gestation CD69⁺ HSCs/MPPs were consistently more clonogenic with higher multilineage capacity (**fig. S10A-D**).

These findings indicate the CD69⁺ fraction of stimulated Lineage⁻CD34⁺CD38⁻ week 20 FL is enriched for HSCs compared to CD69⁻ cells. To test this, we performed *in vivo* limiting dilution xenotransplantation in NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wj}/SzJ (NSG) mice. Upon measuring human chimerism in recipient mouse bone marrow at twelve weeks following transplantation, we found CD69⁺ cells contained a significantly higher proportion of long-term engrafting cells compared to CD69⁻ cells (**Fig. 4E, fig. S10E**). Mice bearing CD69⁺ grafts showed a higher frequency of multilineage and particularly erythroid engraftment, compared to predominantly unilineage B-lymphoid production of CD69⁻ grafts (**Fig. 4F**). Lineage⁻CD34⁺CD38⁻CD69⁺ cells also reconstituted the Lineage⁻CD34⁺CD38⁻ compartment in recipient mice (**fig. S10F**). These findings uncover a mid-gestation human HSC state with potential for efficient multipotent engraftment.

Our findings show that during human development and aging, HSPCs temporally shift their transcriptional states to tailor mechanisms of lineage commitment and differentiation trajectories to support age-appropriate physiology. Although human HSPC ontogeny changes over time, HSCs and MPPs appear heterogeneous with skewed HSC fate probabilities and lineage biased MPPs at all ages. Within mid-gestation fetal liver, a transient HSC state emerges; engagement of this state may ensure balanced hematopoiesis prior to birth or efficiently populate the newly-formed bone marrow niche. Age-specific HSPC states can be retained or pathologically coopted in leukemia, and these states can shift from diagnosis to relapse as an apparent mechanism of therapeutic evasion. Our reconstruction of developmental hematopoiesis will serve as a foundation for future studies of human hematopoietic and immune development, aging, and age-specific blood disorders (7).

References and Notes:

1. Y. Li *et al.*, Single-Cell Analysis of Neonatal HSC Ontogeny Reveals Gradual and Uncoordinated Transcriptional Reprogramming that Begins before Birth. *Cell Stem Cell* **27**, 732-747 e737 (2020).
2. Z. Bian *et al.*, Deciphering human macrophage development at single-cell resolution. *Nature* **582**, 571-576 (2020).
3. J. D. Buenrostro *et al.*, Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell* **173**, 1535-1548 e1516 (2018).
4. J. Cao *et al.*, A human cell atlas of fetal gene expression. *Science* **370**, (2020).
5. X. Han *et al.*, Construction of a human cell landscape at single-cell level. *Nature* **581**, 303-309 (2020).
6. Y. Lu *et al.*, Single-Cell Analysis of Human Retina Identifies Evolutionarily Conserved and Species-Specific Mechanisms Controlling Development. *Dev Cell* **53**, 473-491 e479 (2020).
7. J. E. Park, L. Jardine, B. Gottgens, S. A. Teichmann, M. Haniffa, Prenatal development of human immunity. *Science* **368**, 600-603 (2020).
8. D. Pellin *et al.*, A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. *Nat Commun* **10**, 2395 (2019).
9. A. M. Ranzoni *et al.*, Integrative Single-Cell RNA-Seq and ATAC-Seq Analysis of Human Developmental Hematopoiesis. *Cell Stem Cell*, (2020).
10. A. M. Klein *et al.*, Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187-1201 (2015).
11. E. Becht *et al.*, Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*, (2018).
12. F. A. Wolf *et al.*, PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol* **20**, 59 (2019).
13. L. Velten *et al.*, Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol* **19**, 271-281 (2017).
14. P. van Galen *et al.*, Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell* **176**, 1265-1281 e1224 (2019).
15. Y. Tan, P. Cahan, SingleCellNet: A Computational Tool to Classify Single Cell RNA-Seq Data Across Platforms and Across Species. *Cell Syst* **9**, 207-213 e202 (2019).
16. W. W. Pang *et al.*, Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. *Proc Natl Acad Sci U S A* **108**, 20012-20017 (2011).
17. A. A. MacKinney, Jr., Effect of aging on the peripheral blood lymphocyte count. *J Gerontol* **33**, 213-216 (1978).
18. G. Schiebinger *et al.*, Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. *Cell* **176**, 1517 (2019).
19. C. Weinreb, S. Wolock, B. K. Tusi, M. Socolovsky, A. M. Klein, Fundamental limits on dynamic inference from single-cell snapshots. *Proc Natl Acad Sci U S A* **115**, E2467-E2476 (2018).

20. S. Zhang, A. Afanassiev, L. Greenstreet, T. Matsumoto, G. Schiebinger, Optimal transport as a framework for dynamic inference from stationary snapshot data. *BioRxiv*, (2021).
21. H. Zou, T. Hastie, Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society, Series B* **67**, 301-320 (2005).
22. M. T. Witkowski *et al.*, Extensive Remodeling of the Immune Microenvironment in B Cell Acute Lymphoblastic Leukemia. *Cancer Cell* **37**, 867-882 e812 (2020).
23. R. G. Rowe *et al.*, The developmental stage of the hematopoietic niche regulates lineage in MLL-rearranged leukemia. *J Exp Med* **216**, 527-538 (2019).
24. F. Notta *et al.*, Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science* **333**, 218-221 (2011).
25. R. Majeti, C. Y. Park, I. L. Weissman, Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood. *Cell Stem Cell* **1**, 635-645 (2007).
26. S. L. Prasad *et al.*, GPI-80 defines self-renewal ability in hematopoietic stem cells during human development. *Cell Stem Cell* **16**, 80-87 (2015).
27. A. E. Beaudin *et al.*, A Transient Developmental Hematopoietic Stem Cell Gives Rise to Innate-like B and T Cells. *Cell Stem Cell* **19**, 768-783 (2016).
28. S. C. van den Brink *et al.*, Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat Methods* **14**, 935-936 (2017).
29. M. Slyper *et al.*, A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat Med* **26**, 792-802 (2020).
30. P. Lauzurica *et al.*, Phenotypic and functional characteristics of hematopoietic cell lineages in CD69-deficient mice. *Blood* **95**, 2312-2320 (2000).
31. D. M. Popescu *et al.*, Decoding human fetal liver haematopoiesis. *Nature* **574**, 365-371 (2019).
32. S. F. Ziegler, F. Ramsdell, M. R. Alderson, The activation antigen CD69. *Stem Cells* **12**, 456-465 (1994).
33. N. Bujanover *et al.*, Identification of immune-activated hematopoietic stem cells. *Leukemia* **32**, 2016-2020 (2018).

Acknowledgments: We thank Harvey Lodish, Stuart Orkin, Caleb Weinreb, and Omer Yilmaz for helpful discussion. **Funding:** This work was directly funded by Koch Institute for Integrative Cancer Research institutional funds (to H.L.) and Boston Children's Hospital institutional funds (to G.Q.D. and R.G.R.). H.L. is supported by a National Institute of Diabetes and Digestive and Kidney Diseases K08 DK123414, an American Society of Hematology Scholar Award, and Charles W.(1955) and Jennifer C. Johnson Clinical Investigator Award. V.B. and the Koch Institute for Integrative Cancer Research are supported by a National Cancer Institute P30 CA14051. S.T.C. is supported by a National Institute of Diabetes and Digestive and Kidney Diseases R01 DK100854 and National Heart, Lung, Blood Institute U01 HL134696. G.S. is supported by a Career Award at the Scientific Interface, New Frontiers in Research Exploration Grant, and Discovery Grant from the Natural Sciences and Engineering Research Council of Canada. R.G.R. is supported by a National Institute of Diabetes and Digestive and Kidney Diseases K08 DK114527. **Author contributions:** H.L. and R.G.R. designed the project and the experiments. H.L., V.M., D.W., and R.G.R. performed the experiments. J.E., A.A., L.G., S.Z., V.L.B., E.M., S.G., A.R., E.L.R., and G.S. developed analytical pipelines. H.L., J.E., A.A., L.G., S.Z., V.L.B., E.M., G.C., S.G., E.L.R., G.S., and R.G.R. analyzed the data. J.W., G.Q.D., and S.T.C. procured samples, funding, and reagents. All authors contributed to the writing and editing of the manuscript. **Competing interests:** A.R. is a co-founder and equity holder of Celsius Therapeutics, an equity holder in Immunitas, and was an SAB member of ThermoFisher Scientific, Syros Pharmaceuticals, Neogene Therapeutics and Asimov. From August 1, 2020, A.R. is an employee of Genentech. All other authors declare no competing interests. **Data and materials availability:** Sequencing data from this study will be made available through the National Institutes of Health database of Genotypes and Phenotypes (dbGAP), for which a submission has been initiated. All other data is available in the main text or the supplementary materials.

Supplementary Materials:

Materials and Methods

Figure S1-10

Table S1

References (34-44)

Figure 1

Figure 1. Single cell transcriptome profiling of human HSPCs from fetal life through adulthood. (A) Schematic of HSPC sampling. **(B)** UMAP-based embedding of all cells profiled, highlighted based on lineage scores for each individual cell. **(C-H)** Relative expression of lineage marker genes. **(I)** Sample distribution across the UMAP embedding for specific time points (red) relative to all other samples (gray).

Figure 2

Figure 2. Unbiased clustering identifies lineage biases of human HSPCs. (A) HSPC subtypes from Louvain clustering displayed on UMAP and FDL. **(B)** Classification of HSPC clusters based on similarity to bone marrow cell types using SingleCellNet. **(C)** Barplot of distribution of individual HSPC clusters over human life. **(D)** Barplot of distribution of general progenitor cell types over human life. HSC clusters (HSC1-6), granulocyte clusters (My-MPP, Ba-prog, My-prog-1-4), lymphoid clusters (Ly-MPP-1-2, B-prog-1-2), Meg/Ery clusters (Meg/Ery-MPP, Ery-prog-1-2), mono/DC clusters (Ly/DC-prog, Mono/DC-prog). **(E)** Colony formation of CD34⁺ cells from human fetal liver and bone marrow. **(F)** Distribution of HSPC clusters over human life displayed on UMAP and FDL.

Figure 3

Figure 3. Single cell lineage fate and lineage gene marking from fetal life through adulthood. (A) Individual cell fate probability of erythroid, myeloid, and lymphoid differentiation based on population balance analysis (PBA). Proximity of each cell dot to each vertex indicates likelihood of differentiation toward that lineage. Centroid indicates average fate of all cells in that specimen. **(B)** Relative strength of select individual genes predicting erythroid commitment at different timepoints.

Figure 4

Figure 4. A transient HSC state in human fetal hematopoiesis. (A) HSC populations 1-6 at each age as a proportion of all HSC clusters. **(B)** Expression of transcripts encoding CD34, CD38, and CD69 in HSC1-6. **(C-D)** Following 48 hours of mitogenic stimulation, surface expression of CD69 was analyzed in the Lineage⁻CD34⁺CD38⁻ HSC/MPP population by flow cytometry, and the range of signal quantified. **(E-F)** CD69⁺ or CD69⁻ Lineage⁻CD34⁺CD38⁻ cells were transplanted into NSG recipients, and human chimerism and lineage outcome analyzed at 12 weeks ($X^2 = 5.17$, 1 DF; B=B-cell, M=myeloid, E=erythroid, MK=megakaryocyte).