Hematopoietic stem cells undergo bidirectional fate transitions *in vivo*.

Authors: Tsuyoshi Fukushima[1,2,3,4], Trine Ahn Kristiansen[1,2,3], Lai Ping Wong[5,6], Samuel Keyes[1,2,3], Yosuke Tanaka[4], Michael Mazzola[1,2,3], Ting Zhao[1,2,3], Lingli He[1,2,3], Masaki Yagi[1,2,5,7] Konrad Hochedlinger[1,2,5,7], Satoshi Yamazaki[4], Ruslan I. Sadreyev[5,6], David T Scadden[1,2,3]*.

Affiliations:

[1]Center for Regenerative Medicine, Massachusetts General Hospital, Boston, MA 02114, USA;

[2]Harvard Stem Cell Institute, Cambridge, MA 02138, USA;

[3]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA.

[4]Division of Cell Regulation, Institute of Medical Science University of Tokyo, Tokyo, Japan;

[5]Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA;

[6]Department of Pathology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA

[7]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA;

*Corresponding author. Email: david_scadden@harvard.edu

Highlights:

・Sequenctial introduction of DNA barcodes in vivo was developed to assess time dependent changes in cell fate.

・Clonal phylogenetic tracer (CP-tracer) enabled high-resolution phylogenetic analysis of ~100,000 subclones derived from ~500 individual hematopoietic stem cells (HSC).

・Bidirectional fate transitions between myeloid-biased haematopoietic stem cells (My-HSCs) and lineage-balanced haematopoietic stem cells (balanced-HSCs) were observed.

・Hhex was identified as a molecular driver of HSC lymphoid competence.

Summary: Transitions between subsets of differentiating hematopoietic cells are widely regarded as unidirectional *in vivo*. Here, we introduce clonal phylogenetic tracer (CP-tracer) that sequentially introduces genetic barcodes, enabling high-resolution analysis of ~100,000 subclones derived from ~500 individual hematopoietic stem cells (HSC). This revealed previously uncharacterized HSC

functional subsets and identified bidirectional fate transitions between myeloid-biased and lineage-balanced HSC. Contrary to the prevailing view that the more self-renewing My-HSCs unidirectionally transition to balanced-HSCs, phylogenetic tracing revealed durable lineage bidirectionality with the transition favoring My-HSC accumulation over time[1,2]. Further, balanced-HSCs mature through distinct intermediates—My-HSCs and lymphoid-biased-HSCs—with lymphoid competence here shown by CRISPR/Cas9 screening to be dependent on the homeobox gene, *Hhex*. Hhex enables Ly-HSC differentiation, but its expression declines with age. These findings establish HSC plasticity and *Hhex* as a determinant of myeloid-lymphoid balance with each changing over time to favor the age-related myeloid bias of the elderly.

Introduction: Hematopoietic stem cells (HSCs) are a self-renewing, multipotent population comprised of subsets with differences in the balance of production between myeloid and lymphoid lineages[3-5]. We and others developed systems to clonally trace HSCs[1,6-11]. These defined functional subsets such as "balanced-HSCs" producing myeloid and lymphoid cells, "myeloid biased-HSCs (My-HSCs)" predominantly producing myeloid cells and "lymphoid-biased HSCs (Ly-HSCs)" predominantly producing lymphoid cells. These HSC clones with distinct endogenous outcomes or responses to exogenous stimuli[12] are scripted by epigenetic features[11]. Among these, My-HSCs exhibit the highest self-renewal potential and become balanced-HSCs after serial transplantation[1,5,13] leading to the conclusion that they are at the apex of the hematopoietic differentiation cascade. Single cell transplant studies suggest that the cascade flows unidirectionally. Here, we developed a clonal phylogenic tracer, "CP-tracer," that allows sequential genetic labelling of cells to define fate transitions following the initial marking of single clones over time with high resolution. This system revealed unanticipated bidirectional differentiation trajectories among HSC and defined a molecular specifier of cell fate within primitive hematopoietic stem cells. The bidirectionality of HSC favors accumulation of My-HSC over time as is observed in aged mice and humans.

Results:

Time-course tracing by random DNA barcoding identifies novel progenitor types.

DNA barcoding inserts a unique sequence into the genome for each cell that is inherited by daughter cells with cell division enabling lineage tracing from a single parental cell[14,15]. However, non-nucleated Plt and RBC lineage tracing has been difficult to accomplish. We modified and applied previously described methods[16] for recovering expressed barcode information from Plt and RBC mRNA (Figures S1A-S1D and Methods) to achieve more complete definition of clonal myeloid differentiation capability. Very low expression of other lineage-specific genes indicated low mRNA contamination (Figures S1E-S1F). Barcode-transduced immunophenotypic HSCs (CD150+CD48-ckit+Sca1+Lin-) (Figure S1G) were transplanted into irradiated mice and the barcodes analyzed

sequentially in blood and bone marrow (BM) cells (Figures 1A and S1G-S1K). The recovered barcodes were classified by unsupervised k-mean clustering and 6 cell types were identified based on the cell types they produced (Figures 1B and S1L). The functionally classified subgroups are listed on the vertical-axis in the heatmaps throughout the Figures and are based on the presence of the barcode in the immunophenotypically defined hematopoietic cell types indicated on the horizontal-axis; time post-transplant is also indicated on the horizontal-axis. Functional cell grouping of cells bearing the identifying barcodes (represented by a row in the heatmap) included: 1. My-HSCs producing myeloid lineage cells while maintaining immunophenotypic HSCs 20 weeks after transplantation. 2. balanced-HSCs producing all blood cell lineages while maintaining immunophenotypic HSCs. 3. Myeloid Ristricted Progenitors (MyRPs) producing myeloid cells over >8 weeks after transplantation without maintaining immunophenotypic HSCs. 4. Ly-HSCs producing all blood cell lineages but showed no evidence for maintaining immunophenotypic HSCs and decreased myeloid production over time in the primary transplant (Figures 1B, 1C and S1L). These subsets correlate functionally with the 4 cell types reported by others in single cell transplantation studies[1,9,10]. Also observed were two functional groups not previously described. 5. B-biased multipotent progenitors (B-MPPs) that produced myeloid and B cells but not T cells. They did not maintain immunophenotypic HSCs and their myeloid production decreased within the primary transplant (Figures 1B, 1C and 1L). This pattern was similar to Ly-HSCs but differed in T cell potential. 6. Transient myeloid progenitors (TMPs) that produced Plt, RBC and/or GM only through 4 weeks (Figures 1B, 1C and S1L). This pattern differed from MyRPs in the duration of myeloid production. Only My-HSCs and balanced-HSCs could maintain immunophenotypic HSCs 20 weeks after transplantation consistent with a stem cell functional phenotype. Therefore, time-course tracing by random DNA barcoding identifies six functional subsets, including the four previously known subsets[1,9,10] (My-HSCs, balanced-HSCs, Ly-HSCs and MyRPs) and two newly identified subsets (B-MPPs and TMPs). Corroberating My-HSCs, balanced-HSCs, Ly-HSCs, and B-MPPs were also identified in human CD34-positive cells[17] (Figures S2M, S2N).

My-HSCs favor self-renewal and can transition to balanced-HSCs.

My-HSCs are reported to have higher stemness than balanced-HSCs and transition into balanced-HSCs[5,10,13,18]. We assessed these characteristics using random DNA barcode analysis. Consistent with a previous report[10], self-renewing HSC clones could be classified as Plt only (P-HSCs), Plt/RBC/GM (PRG-HSCs), Plt/RBC/GM/B (PRGB-HSCs) and all blood cell lineages (balanced-HSCs) by their lineage output (Figure 1D). P-HSCs, PRG-HSCs and PRGB-HSCs bundled together consitute the My-HSCs group. Differentiated blood cell output per clone was highest in balanced-HSCs (balanced-HSCs > PRGB-HSCs > PRG-HSCs > P-HSCs) (Figure 1E), while the contribution to immunophenotypic HSCs 20 weeks after transplantation was higher in PRGB-HSCs than P or PRG-HSCs and balanced-HSCs (PRGB-HSCs > balanced-HSCs > PRG-HSCs > P-HSCs) (Figure 1F). The

ratio of differentiated cell lineage output (Figure 1E) to immunophenotypic HSCs (Figure 1F)[14] was lower in My-HSCs than balanced-HSCs (P-HSCs < PRG-HSCs < PRGB-HSCs < balanced-HSCs) (Figures 1G); My-HSCs also preferentially biased toward self-renewal over differentiation (Figures 1G). By the criterion of relative self-renewal versus differentiation for 'stemness'[19,20], My-HSCs have greater stemness than balanced-HSCs .

Examining clonal behavior during secondary transplantation (Figure S1O), P-HSCs exhibited PRG or PRGB-HSC patterns and PRG-HSCs exhibited PRGB-HSC patterns. In addition, PRGB-HSCs exhibited balanced-HSC patterns upon secondary transplantation (Figures 1H, 1I and S1P). These data demonstrate that My-HSCs can transition to the less self-renewing balanced-HSC subset. Therefore, the high stemness of My-HSCs and their transition to balanced-HSCs were further validated using random DNA barcodes.

Development of clonal phylogenic tracer: "CP-tracer"

We then explored the capacity to transition between functional subsets by tracing the lineage fate of distinct offspring derived from a parental cell. Using CRISPR/Cas9[21-24] with specialized barcodes (*Scratchpad)* that accumulate mutations over time, we could track distinct progeny from a parental cell after progressive cell divisions (Figure 2A). *Scratchpad* allows for multiple sgRNAs to randomly introduce mutations in their targets by editing at random times[21]. However, a single scratchpad provides ~10^4 genetic diversity, insufficient to robustly trace subclones derived from hundreds to thousands of parental cells[25]. Using multiple loci increases barcode diversity, however it requires single-cell analysis for integrating the information, limiting the number of cells for recovering barcodes and detecting subclones (Figure S2N)[24,26,27]. To overcome these limitations, we designed clonal phylogenic tracer, "CP-tracer," by combining random DNA barcodes and scratchpad in a single locus to identify parental cells and their progeny respectively (Figures 2A and S2A). We modified the original scratchpad[21] using 3 sgRNAs with 2 target sequences for each of the 3 sgRNAs in different orientations (Figures S2A-S2G and Methods). This CP-tracer was transduced into immunophenotypic HSCs from Cas9-EGFP mice and transplanted into lethally irradiated recipients, analyzing blood and BM (more than 4×10^7 hematopoietic cells; Figures S2H-S2M). The DNA barcodes of CP tracer could identify clones such as My-HSCs, balanced-HSCs, Ly-HSCs, and B-MPPs (Figure 2B). The scratchpad codes indicated the fate of descendant subclones. CP tracer detected 92,399 subclones (Figure S2N). In 96 balanced-HSCs clones, 56,301 scratchpad variants were detected; scratchpad variance was 612.1 ± 68.26 subclones per barcode clone (Figure S2O) and 97.3% of scratchpad variants were unique (Figure S2P). Most of the scratchpads were edited in 3-5 out of 6 sites generally without large deletions (Figures S2Q and S2R). Editing percentage at each site 20 weeks after transplantation was very similar to that at 14 days of culture (Figures S2E and S2S), and was stable over time (Figure S2T). Furthermore, CP-tracer labeled around 600 offspring from a single HSC

enabling the tracing of early fate segregation from repopulating HSCs.

Balanced-HSCs can transition into My-HSCs.

Immunophenotypic HSCs transduced with CP tracer were transplanted and the subclones arising were analyzed (Figures 2A, 2B and S2H) at 20 weeks (Figures 2C and S3A). Only 2.08 ± 0.4 % of the subclones derived from balanced-HSCs were self-renewing compared with 5.40 ± 0.8 % of My-HSCs (2.6 fold higher; p=0.0001; Figures 2D and S3A-S3C). Interestingly, the majority of the self-renewing balanced-HSC subclones had an output comparable to My-HSCs in which lymphoid output was suppressed or delayed (Figure 2E). These data are consistent with balanced-HSCs transitioning to My-HSCs despite the higher stemness of My-HSCs. Classification using k-means clustering may incorrectly classify balanced-HSCs as My-HSCs. Assessing whether incorrect classification could account for our findings, we manually confirmed that subclones derived from balanced HSCs generate My-HSCs producing myeloid without lymphoid cell production (Figure 2F) arguing against incorrect classification. To determine the frequency of My-HSCs derived from self-renewing balanced-HSC subclones, the lineage output of each subclone was compared to the average outputs of clusters (P-HSCs, PRG-HSCs, PRGB-HSCs, and balanced-HSCs) in the reference heatmap (Figure 1D). Each subclone was then assigned to the cluster it was closest to. This indicated that 93.4±2.4% of the self-renewing subclones derived from balanced-HSCs were My-HSCs (Figures 2E and 2G); not balanced-HSCs. In addition, My-HSCs derived from balanced-HSCs maintained a My-HSC fate after secondary transplantation (Figure 2H), indicating a durable fate transition.

The evidence for My-HSCs transitioning to functional balanced-HSCs (Figure 1H) and the CP tracer phylogenic tracing showing balanced-HSCs transitioning to My-HSCs (Figures 2E-2G), is consistent with bidirectional fate transitions. Since some My-HSCs derived from balanced-HSCs reverted back to balanced-HSCs upon secondary transplantation (Figure 2I), single clones appear to have bidirectional fate plasticity between balanced-HSCs and My-HSCs.

The majority of self-renewing subclones from balanced-HSCs transitioned to My-HSCs and maintained a My-HSC fate after secondary transplantation (Figure 2H). On the other hand, all self-renewing subclones of My-HSCs performed as My-HSCs in primary transplantation (Figures S3D and S3E) and rarely transitioned to balanced-HSCs after secondary transplantation (Figures S3F and S3G); therefore, the bidirectional transition between balanced-HSCs and My-HSCs was biased in the direction from balanced-HSCs to My-HSCs.

In sum and consistent with previous reports[5,10,13,18], My-HSCs possess greater stemness than balanced-HSCs and can transition to balanced-HSCs (Figures 1G-1I) but unexpectedly, CP tracer showed that balanced-HSCs can reverse their trajectory and transition to My-HSCs (Figures 2E-I).

Balanced-HSCs contribute to mature blood cell lineages through two intermediate

states, My-HSCs or Ly-HSCs

Analyzing the Scratchpad editing pattern (Figures 2J and S4A), balanced-HSCs branched to balanced-HSCs, My-HSCs, and Ly-HSCs (Figures 2J and S4A). The descendent subclones of My-HSCs mainly produced My-HSCs and MyRPs. Ly-HSCs mainly produced B-MPPs, T progenitors, BT progenitors, and TMPs (Figures S3B-S3D and S4B-S4C). Surprisingly, My-HSCs were able to produce B progenitors without evidence of producing Ly-HSCs (Figures S3B and S3C) perhaps indicating a parallel path to B cells. Overall, balanced-HSCs contribute to the spectrum of progenitor cells (Figure S4D) and mature blood cell lineages through two distinct intermediate states, My-HSCs and Ly-HSCs.

Combinaing fate tracing and scRNA-seq identified transcriptional differences between My-HSCs and balanced-HSCs.

Combined DNA barcoding and single-cell RNA-seq analysis was conducted to define the gene expression signatures of immunophenotypic HSC according to their functional states in vivo. Barcoded immunophenotypic HSCs were transplanted and their lineage output in PB and BM analyzed at 4-20 weeks; scRNA-seq of immunophenotypic HSC from BM was then analyzed 20 weeks after transplantation. The expressed DNA barcodes were aligned by scRNA-seq (Figure S1A) and clonal lineage output paired with the transcriptome of the HSC source representing that clone (Figure 3A). Single HSC functional states were annotated according to the differentiation classification shown in Figure 1D. Cell cycle score was calculated by cell cycle gene expression in each cell[28]. A stem cell gene expression score, MolO, was high and cell cycle score was low in cells derived from P-HSCs and PRG-HSCs, indicating higher stemness and dormancy of these cell types (Figures S5A and S5B). Uniform manifold approximation and projection (UMAP)[29] (Figures 3C and S5C-S5F) analysis indicated transcriptional differences between My-HSCs and balanced-HSCs (Figure 3C). Molecules specific for My-HSCs included transcription factors[30] (Mecom and Tcf15) and others (Mllt3, Tgfbr3 and a marker of the My-HSCs, Vwf[13]) (Figures 3D and S5G). Mecom[31] and Mllt3[32] are known self-renewal regulators, and Tcf15 is reported to induce a bias toward self-renewal over differentiation[14], in keeping with the higher stemness observed among My-HSCs. Tgfbr3 ligand TGF is a known critical niche factor for maintaining HSCs[33] and is reported to induce the cell cycle specifically in My-HSCs, but not in Ly-HSCs[12]. Molecules specific for balanced-HSCs included transcription factors (Runx2, Runx3, and critical lymphoid differentiation regulators, Ikzf1 and Ikzf2 (Figures 3D and S5G). Ectopic expression of Runx3 is known to induce epigenetic lymphoid priming in HSCs[34]. Pathway analysis showed evidence of high activity of Mecom and low activity of Stat1 in My-HSCs (Figure 3E) reflecting the higher expression of Mecom itself and the Stat inhibitor, Socs2 (Figure 3D) reported to prevent exhaustion of HSCs after serial transplantation[35]. Pathway analysis also showed evidence of high activity of Runx and Ikzf in balanced-HSCs (Figure 3F) as apparent in gene expression (Figure 3D). Expression of the downstream targets of IRAK4 were lower in My-HSCs and higher in balanced-

HSCs (Figure 3G) suggesting less inflammatory signaling in My-HSCs. The combination of fate tracing and scRNA-seq identified transcriptional differences between these different HSC subsets.

Single cell CRISPR screening identifies genes that control HSC heterogeneity.

To determine the importance of these transcriptional differences, we combined random DNA barcodes with CRISPR/Cas9 based sgRNA library screening (Figure 4A) focusing on 30 genes ,including differentially expressed genes in particular functional HSC subsets (Figure 4B). The screening vector contained both a random DNA barcode library with approximately 10^6 variations and a sgRNA library with 100 sgRNAs containing 3 sgRNAs targeting each candidate gene and 10 NT-sgRNAs (Figure 4A). Lineage tracing of immunophenotypic HSCs was performed and the effect of gene perturbation on My-HSCs/MyRPs, Ly-HSCs/B-MPPs and TMPs was evaluated (Figures S6A and S6E-S6G). The results showed that KO of Junb increased My-HSCs/MyRPs, KO of Bex4 and Junb decreased Ly-HSCs/B-MPPs, and KO of Mllt3 and Hhex increased TMPs (Figure 4B). Junb, Mllt3 and Hhex were reported respectively as a regulator of proliferation[36], self-renewal[32] and lymphoid development[37-39]. However, the effects of these genes at the clonal level had not been reported and doing so here validated the method's utility for functionally defining molecular regulators of individual HSC fate.

Hhex is necessary for Ly-HSCs fate and suppression of My-HSC fate.

Hhex is a homeodomain transcription factor that is thought to be important for embryonic development of definitive HSC[40] and for lymphoid development after adult HSPC commitment[37-39]. It has not been thought to be relevant for the early adult HSPC lymphoid fate decisions because the Hhex KO has reduced B and T cell progenitors without reduction of the lymphoid primed MPP4 (Flt3+ckit+Sca1+Lin-) [41,42] or common lymphoid progenitors (CLP: Lin-cKit^{low}Sca1^{low}Flt3+Il7R+) [37-39]. However, our data indicate that there are multiple paths to MPP4 and CLP that may confound direct linear correlations. That is, both Ly-HSCs and MyRPs produce immunophenotypic MPP4 (Figure S7A) and Ly-HSCs, B-MPPs, MyRPs and platelet progenitors were all able to produce immunophenotypic CLP (Figure S7A). Therefore, immunophenotypic MPP4 and CLP may descend from multiple progenitor populations making them sufficiently heterogeneous to preclude using their population size as indicative of where in differentiation a gene is active. This was validated by Hhex KO in immunophenotypic HSCs and evaluating competitive repopulation capacity (Figure 4C). Hhex KO was not limited to mature cells, rather it significantly increased My-HSCs and MyRPs; furthermore, it decreased Ly-HSCs and B-MPPs (Figures 4D, 4E and S7B-S7E). While no significant Hhex expression changes were detected. chromatin regions[34] containing the Hhex motif were more inaccessible by ATAC-Seq in My-HSCs than balanced-HSCs (Figures S7F and S7G) consistent with lower activity of Hhex in My-HSCs. These data indicate that Hhex plays a role early in adult HSC differentiation, enabling lymphoid fate competence.

Reciprocally, Hhex cDNA was added back to KO cells at the immunophenotypic HSC (CD150+CD48-ckit+Sca1+Lin-) or MPP (CD48+ckit+Sca1+Lin-) level respectively (Figures 4F and S8A-S8C). Hhex overexpression rescued the B cell phenotype in HSC (Figures 4G and S8D). Overexpression in immunophenotypic MPPs was directionally similar but of smaller magnitude (Figures 4H and S8E). Therefore, Hhex appears to be key for commitment to B lineage development early in differentiation likely between immunophenotypic HSCs and MPPs. Apoptosis of prepro-B cells by Hhex KO was reported by others to be rescued by overexpression of anti-apoptotic factor Bcl2[37]. However, when we overexpressed Bcl2 the Hhex KO frequency of My-HSCs/MyRPs was still increased and Ly-HSCs/B-MPPs still decreased (Figure 4I and S8F-S8J) arguing against apoptosis causing Ly-HSC reduction.

These data indicate that Hhex is critical for Ly-HSC/B-MPP fate and lymphoid myeloid balance (Figure 4J) although overexpressing it in wild-type HSC is not sufficient to significantly increase lymphoid competence (Figures S8K-S8L). Notably, Hhex declines in human bone marrow HSC with age[43] (Figure 4K).

Discussion: The CP tracer has the ability to identify a clonal cell of origin and to track early fate decisions of an individual clone labeling approximately 600 progeny per HSC clone within 14 days (Figures S2E, S2N, S2R and S2S). It does so with high barcode variation (at least $10^{10}=2\times10^{5}$ random DNA barcode$\times5\times10^{4}$ scratchpad) in a single locus (Figures 2A and 2O) enabling high-resolution phylogenetic analysis to trace the cell fate of ~100,000 descendant subclones originating from ~500 single parent cells in vivo.

Using CP tracer has provided both new and confirmatory information about hematopoiesis. Single cell transplantation studies previously identified heterogeneity within the immunophenotypic HSC fraction[1,9,10], including My-HSCs transitioning to balanced-HSCs with secondary transplantation[1]. The vWF-GFP+ fraction that enriches My-HSCs, produced the vWF-GFP+ fraction and the vWF-GFP- fraction, while the vWF-GFP- fraction enriched balanced-HSCs[10,13]. In addition, it has been shown that My-HSCs can transition to balanced-HSCs supporting My-HSCs as the upstream population[1]. This is consistent with HSCs biased to myeloid cells having the highest stemness and lose stemness as they gain lymphoid competence[5,18]. However, unexpectedly CP tracer analyses indicate that balanced-HSCs can also reverse their trajectory and transition to My-HSCs that then show a higher stemness. Therefore, unidirectional differentiation that is ascribed to hematopoiesis in vivo is not rigidly maintained, at least in stress conditions like the transplant setting evaluated here.

Further, CP tracer may add mechanistic insight to pathobiology. Myeloid biased hematopoiesis is a well-defined characteristic of aged individuals and My-HSCs accumulate in aged adults, [1,34,44,45] potentially contributing to age-related phenotypes[2]. Yet, the basis for this shift in

hematopoiesis is incompletely understood. CP tracer revealed that balanced-HSCs largely transited toward My-HSCs and that while bidirectional transitions can occur, My-HSCs infrequently transited to balanced-HSCs. These dynamics favor the accumulation of My-HSCs over time and may participate in the altered hematopoiesis observed with aging[1,34,44,45]. Since Hhex, defined here by targeted CRISPR screening, serves as a modulator of My-HSC fate choice; its role in the aging phenotype is also hypothesized, supported by Hhex expression declines observed in aged human HSC. Further testing of whether altering Hhex regulation can affect hematopoietic aging is a topic for future study.

Acknowledgements

We would like to thank all Scadden lab members, HSCI CRM Flow Cytometry Core, Animal facility at Center for Comparative Medicine at MGH and Bauer Core Facility at Harvard University for assistance with next generation sequencing. This research was supported by the Gerald and Darlene Jordan Professor of Medicine Chair, National Institutes of Health HL131477 and HL142494 and the Harvard Stem Cell to D.T.S. T.F. was supported by a JSPS Overseas Research Fellowships (202160002), The Uehara Memorial Foundation Postdoctoral Fellowship for Research Abroad (2020) and The grant of the Mochida Memorial Foundation for Medical and Pharmaceutical Research (2020).

Author contributions

Conceptualization, T.F. and D.T.S.; Methodology, T.F. and D.T.S.; Investigation, T.F., T.A.K., S.K., , M.M., T.Z., L.H., M.Y.,; Data analysis, T.F., L.P.W., R.I.S.,; Writing – Original Draft, T.F., Y.T., D.T.S.; Writing– Review & Editing: T.F., T.A.K., Y.T., M.M., T.Z., L.H., S.Y., D.T.S.; Visualization: T.F.; Project Administration, K.H., S.Y., R.I.S.,; Funding Acquisition, D.T.S.; Supervision, D.T.S.

Competing Interests

D.T.S.: Agios Pharmaceuticals: Membership on an entity's Board of Directors or advisory committees; Editas Medicines: Membership on an entity's Board of Directors or advisory committees; Carisma Therapeutics: Membership on an entity's Board of Directors or advisory committees; Garuda Therapeutics: Current holder of individual stocks in a privately-held company, Membership on an entity's Board of Directors or advisory committees; Sonata Therapeutics: Current holder of individual stocks in a privately-held company, Membership on an entity's Board of Directors or advisory committees; Lightning Biotherapeutics: Current holder of individual stocks in a privately-held company, Membership on an entity's Board of Directors or advisory committees.

Data and Material Availability

Materials are available, subject to material transfer agreement requests submitted to D. T. S. scRNA-sequencing data have been deposited at GEO "GSE255989" and are publicly available as of the date

of publication. All other data available in the manuscript or supplementary materials are available from corresponding author upon reasonable request.

Isolation of Hematopoietic Stem Cells

Bone marrow cells were harvested from C57BL/6J (JAX, 000664) or B6J.129(Cg)-Gt(ROSA)26Sortm1.1(CAG-cas9*,-EGFP)Fezh/J (JAX, 026179) mice by extracting femurs, tibias, pelvic bones, sternums, humeri, and vertebrae, followed by crushing. The bone marrow cells were passed through a 100 μm cell strainer (Corning, 431752), and ACK lysis (Quality Biological, 118-156-721) was performed using 5 mL at room temperature for 5 minutes. After adding 45 mL of PBS, the cells were passed through a 100 μm cell strainer and centrifuged for 5 minutes. The supernatant was discarded. Bone marrow cells were incubated with Lineage antibodies (1:1000) (Anti-mouse CD5-biotin Antibody (clone 53-7.3) (BioLegend, 100604), Anti-mouse/human CD45R/B220-biotin Antibody (clone RA3-6B2) (BioLegend, 103204), Anti-mouse TER-119-biotin Antibody (clone TER-119) (BioLegend, 116204), Anti-mouse Gr-1-biotin Antibody (clone RB6-8C5) (BioLegend, 108404), Anti-mouse/human CD11b-biotin Antibody (clone M1/70) (BioLegend, 101204)) in 1 mL for 10 minutes at 4 °C. After adding 45 mL of PBS and centrifuging for 5 minutes, the supernatant was discarded. This process was repeated twice. The cells were then incubated with 10 mL of PBS and 90 μL of SA microbeads (Miltenyi Biotec, 130-090-858) for 15 minutes at 4 °C. After adding 45 mL of PBS and centrifuging for 5 minutes, the supernatant was discarded. This process was repeated twice. Negative selection was then performed using LS columns (Miltenyi Biotec, 130-042-401). The cells from the negative fraction were centrifuged for 5 minutes, and the supernatant was removed. The cells were incubated with a cocktail of antibodies, including Anti-mouse CD150-PE Antibody (clone TC15-12F12.2) (BioLegend, 115904) (1:200), Anti-mouse cKit-PEcy7 Antibody (clone 2B8) (BioLegend, 105814) (1:400), Anti-mouse Sca1-APC Antibody (clone B7) (BioLegend, 108112) (1:400), Anti-mouse CD48-APCcy7 Antibody (clone HAM48-1) (BioLegend, 103432) (1:400), and SA-BV605 (BioLegend, 405229) (1:400), or Anti-mouse cKit-PEcy7 Antibody (1:400), Anti-mouse CD150-APC Antibody (clone TC15-12F12.2) (BioLegend, 115910) (1:400), Anti-mouse CD48-APCcy7 Antibody (1:400), Anti-mouse Sca1-BV421 Antibody (clone B7) (BioLegend, 108127) (1:400), and SA-BV605 (1:400) for 15 minutes at 4 °C. Following the addition of 1 mL of PBS and centrifugation for 5 minutes, the supernatant was discarded. The cells were resuspended in 6 mL of PBS and sorted for the CD150+CD48-Lin-cKit+Sca1+ fraction using FACS Aria.

Plasmid Preparation and DNA Assembly

The plasmid pMJ114 (Addgene Plasmid #85995) was digested using the restriction enzymes HpaI (NEB, R0105L) and NotI (NEB, R3186L) to linearize the vector. A synthetic DNA fragment containing the hU6 promoter and sgRNA scaffold was obtained from Azenta and used as an insert for

HiFi DNA assembly. The linearized vector (50 ng) and synthetic DNA fragment (5 ng) were assembled using HiFi DNA Assembly Master Mix (5 µl) (NEB, E2621L) in a total reaction volume of 10 µl. The reaction mixture was incubated at 50°C for 1 hour. The plasmid containing scratchpad which was obtained from Azenta was digested using the restriction enzymes Kfl1 (Thermo Scientific, FD2164) to linearize the vector. A synthetic DNA fragment containing the sgRNA cassetes was obtained from Azenta and used as an insert for HiFi DNA assembly. The linearized vector (50 ng) and synthetic DNA fragment (5 ng) were assembled using HiFi DNA Assembly Master Mix (5 µl) (NEB, E2621L) in a total reaction volume of 10 µl. The reaction mixture was incubated at 50°C for 1 hour. Following the assembly, the reaction product was transformed into DH5α competent cells (NEB, C2987H) via heat shock. Colonies obtained from the transformation were screened to identify those containing the backbone plasmid with the desired insert. A second HiFi DNA assembly was performed using the previously generated backbone plasmid and an oligo pool containing randomized sequences. The reaction mixture consisted of 500 ng of the backbone vector, 45 nM of the oligo pool, and 50 µl of HiFi DNA Assembly Master Mix, with a total reaction volume of 100 µl. The reaction was incubated at 50°C for 1 hour. Following the assembly, the reaction mixture was purified using a DNA purification kit, with elution performed in 15 µl of molecular-grade water (MQ). Electroporation was performed by adding 5 µl of the purified assembly reaction to 100 µl of MegaX DH10B competent cells (Invitrogen, C640003). The mixture was subjected to electroporation at 2.0 kV, 200 Ω, and 25 µF. Immediately after electroporation, 4 ml of pre-warmed recovery medium was added to the cells, and the mixture was incubated at 37°C for 1 hour. Following recovery, 4 µl of the transformed cells was diluted to 1 ml with LB medium. A 100 µl aliquot of this dilution was spread on an LB plate containing the appropriate antibiotic. The number of resulting colonies was used to estimate the library diversity, which was calculated to be approximately 10^6 variants. The remaining transformation mixture was spread across four large (15 cm) LB plates containing antibiotics. After overnight incubation, 10 ml of LB liquid medium was added to the surface of each plate, and the bacterial colonies were collected by gently scraping. This process was repeated two to three times, and the collected bacterial suspension was pooled into a single tube. The plasmid library was purified using a Maxi Prep kit, following the manufacturer's protocol.

Virus Production

HEK293T cells were seeded at a density of 4×10^6 cells in a 15 cm dish. After 18 hours, the medium was changed. Opti-MEM (Gibco, 31985-062) (900 µL), VSVG (3.5 µg), PAX2 (7 µg), vector (12 µg), and Fugene6 (67.5 µL) (Promega, E2692) were mixed and incubated at room temperature for 15 minutes, then added to the HEK293T cells. After 24 hours, the medium was changed again. An additional 24 hours later, the supernatant was collected and subjected to ultracentrifugation using an SW32 rotor at 20,000 rpm at 4°C for 2 hours. The supernatant was discarded, and the pellet was

resuspended in PVA medium[46] (200–400 μL).

Virus Infection

RetroNectin (Takara, T100A) was applied to 96-well or 48-well plates at a volume of 100–250 μL per well and incubated at room temperature for 1 hour. The RetroNectin was then removed, and the plate was washed with PBS. Hematopoietic stem cells were seeded at a density of 1–3 × 10^4 cells per well for 96-well plate and 2-6 × 10^4 cells per well for 48-well plate. Virus solution (200–400 μL) was added to each well. After 18 hours, the hematopoietic stem cells were collected.

Bone Marrow Hematopoietic Stem Cell Transplantation

Bone marrow was harvested from the femurs and tibias of C57BL/6J mice aged 8–12 weeks. A cell suspension containing 1 × 10^7 cells in 90 μL was prepared, and 10 μL of anti-Sca1-FITC Antibody (Miltenyi Biotec, 130123-124) was added, followed by a 10-minute incubation at 4°C. After adding 2 mL of PBS and centrifuging for 5 minutes, the supernatant was discarded. The cells were resuspended in 80 μL of PBS, and 20 μL of anti-FITC microbeads (Miltenyi Biotec, 130123-124) was added, followed by a 15-minute incubation at 4°C. After adding 2 mL of PBS and centrifuging for 5 minutes, the supernatant was discarded. Negative selection was then performed using LD columns (Miltenyi Biotec, 130-042-901). C57BL/6J mice aged 8–12 weeks were irradiated with 475 cGy at 4-hour intervals for two sessions. Hematopoietic stem cells (0.5–3 × 10^4) and the Sca1-negative fraction (2 × 10^5) were dissolved in 200 μL of PBS and injected into the tail veins of mice. For secondary transplantation, C57BL/6J mice aged 8–12 weeks were irradiated with 475 cGy at 4-hour intervals for two sessions. Subsequently, 1% or 10% of the primary bone marrow was injected into the tail veins of the mice.

Peripheral Blood Sampling

At 4, 8, 12, and 16 weeks post-transplantation, approximately 120 μL of peripheral blood was collected from the tail vein of the mice into 20 μL of citrate-dextrose solution (Sigma Aldrich, C3821-50ML).

Red Blood Cell Recovery

Of the total 140 μL of peripheral blood, 40 μL was transferred to a separate Eppendorf tube for red blood cell analysis. One microliter each of anti-mouse CD41-biotin Antibody (clone MWReg30) (BioLegend, 133930) and anti-mouse CD45-biotin Antibody (clone 30-F11) (BioLegend, 103104) was added, followed by a 10-minute incubation at room temperature. After adding 1 mL of PBS and centrifuging for 5 minutes, the supernatant was discarded. This process was repeated twice. The pellet was then resuspended in 600 μL of PBS, and 12 μL of SA microbeads (Miltenyi Biotec, 130-090-858) was added, followed by a 15-minute incubation at room temperature. After adding 1 mL of PBS and

centrifuging for 5 minutes, the supernatant was discarded. This process was repeated twice. The sample was resuspended in 1 mL of PBS, loaded onto an LD column (Miltenyi Biotec, 130-042-901) pre-treated with 2 mL of PBS, and immediately loaded with an additional 2 mL of PBS. After the complete passage of liquid through the column, 2 mL of PBS was added to the column. Following the complete passage of the liquid through the column, the negative fraction was centrifuged, and the supernatant was removed.

Platelet, Granulocyte/Monocyte, and lymphoid cells Recovery

One mL of ACK lysis buffer (Quality Biological, 118-156-721) was added to 100 μL of peripheral blood and incubated at room temperature for 15 minutes. After the addition of 1 mL of PBS and centrifugation for 5 minutes, the supernatant was discarded. Another 1 mL of ACK lysis buffer was added and incubated at room temperature for an additional 15 minutes. After adding 1 mL of PBS and centrifuging for 5 minutes, the supernatant was discarded. The cells were incubated at room temperature for 15 minutes with antibodies in a total volume of 100 μL: Anti-mouse CD19-APCcy7 Antibody (clone 6D5) (BioLegend, 115530) (1:400), Anti-mouse CD3ε-BV785 Antibody (BioLegend, 100355) (1:400), Anti-mouse CD41-APC (clone MWReg30) (BioLegend, 133914) (1:400), Anti-mouse TER-119-biotin Antibody (clone TER-119) (BioLegend, 116204) (1:400), Anti-mouse/human CD11b-PE Antibody (clone M1/70) (BioLegend, 101208) (1:1000) or Anti-mouse/human CD11b-BV421 Antibody (clone M1/70) (BioLegend, 101212) (1:400). After the addition of 1 mL of PBS and centrifugation for 5 minutes, the supernatant was discarded. The cells were resuspended in 6 mL of PBS. FACS Aria was used to sort the FSClowSSClowCD41+Ter119-CD11b-CD19- fraction as platelets and collect 1,250,000 cells. Immediately after collection, the cells were centrifuged at 1,200g for 5 minutes and the supernatant was removed. In the BFP+ or RFP+ fractions, cells were sorted based on the following markers: CD11b+CD19- as granulocytes/monocytes, CD11b-CD19b+ as B cells, CD11b-CD19b-CD3e+ as T cells.

Barcode cDNA Synthesis from RNA

To platelet and RBC samples, 350 μL of RLT buffer (RNeasy, Qiagen, 74104) was added. After pipetting, 350 μL of 80% ethanol was added and mixed well. The mixture was loaded onto an RNeasy mini column. Following centrifugation at 12,000g for 1 minute, the liquid was discarded. Subsequently, 500 μL of RW1 buffer was added, and after centrifugation at 12,000g for 1 minute, the flow-through was discarded. Similarly, 500 μL of RPE buffer was added, and after centrifugation at 12,000g for 1 minute, the flow-through was discarded. The column was then placed into a recovery Eppendorf tube. Next, 25 μL of RNase-free water was added. The RNA was eluted by centrifugation at 12,000g for 1 minute. To the RNA, 2 μL each of Oligo(dT) and dNTP were added and incubated at 65°C for 5 minutes. After this, the mixture was placed on ice for 1 minute. Then, 2 μL each of DTT, RNase

inhibitor, and SuperScript IV (Invitrogen/Thermo, 18091200) were added, along with 8 μL of buffer. The mixture was incubated in a thermal cycler with the following protocol: 23°C for 10 minutes, 55°C for 10 minutes, and 80°C for 10 minutes.

Cell Recovery from Bone Marrow

Bone marrow cells were harvested from the femurs, tibias, pelvic bones, sternums, humeri, and spines of mice at 20 weeks post-transplantation, followed by crushing. The bone marrow cells were passed through a 100 μm cell strainer (Corning, 431752), and ACK lysis (Quality Biological, 118-156-721) was performed using 5 mL at room temperature for 5 minutes. After adding 45 mL of PBS, the mixture was passed through a 100 μm cell strainer and centrifuged for 5 minutes. The supernatant was discarded. One-fourth of the bone marrow cells were incubated for 15 minutes at 4°C with 1 mL of the following antibodies: Anti-mouse CD19-APCcy7 Antibody (BioLegend, 115530) (1:400), Anti-mouse CD3ε-BV785 Antibody (BioLegend, 100355) (1:400), Anti-mouse CD41-APC (BioLegend, 133914) (1:400), Anti-mouse TER-119-biotin Antibody (BioLegend, 116204) (1:400), Anti-mouse/human CD11b-PE Antibody (BioLegend, 101208) (1:1000) or Anti-mouse/human CD11b-BV421 Antibody (BioLegend, 101212) (1:400). After washing with 1 mL of PBS and centrifuging, the cell suspension was resuspended in 3 mL of PBS and sorted using Aria. In the BFP+ or RFP+ fractions, cells were sorted based on the following markers: CD11b+CD19- as Granulocyte/Monocyte, CD11b-CD19+ as B cells, CD11b-CD19-CD3e+ as T cells, CD11b-CD19-CD41+ as Megakaryocytes, CD11b-CD19-Ter119+ as Erythrocytes. The remaining three-fourths of the bone marrow cells were incubated for 10 minutes at 4°C with Lineage antibodies (1:1000): Anti-mouse CD41-biotin Antibody, Anti-mouse CD5-biotin Antibody, Anti-mouse/human CD45R/B220-biotin Antibody, Anti-mouse TER-119-biotin Antibody, Anti-mouse Gr-1-biotin Antibody, Anti-mouse/human CD11b-biotin Antibody. After washing with 1 mL of PBS and centrifugation, the cell suspension was further incubated for 15 minutes at 4°C with a mixture of antibodies: Anti-mouse cKit-PEcy7 Antibody (1:400), Anti-mouse Sca1-BV785 Antibody (clone B7) (BioLegend, 108139) (1:400), Anti-mouse CD48-APCcy7 Antibody (1:400), SA-BV605 (1:400), Anti-mouse CD150-PE Antibody (1:200), Anti-mouse Flt3-APC Antibody (clone A2F10) (BioLegend, 135310) (1:50) or Anti-mouse CD150-APC Antibody (1:400), Anti-mouse Flt3-BV421 Antibody (BioLegend, 135313) (1:50). After washing with 1 mL of PBS and centrifugation, the cell suspension was resuspended in 6 mL of PBS and sorted using Aria. In the BFP+ or RFP+ fractions, cells were sorted based on the following markers: CD150+CD48-Flt3-Lin-cKit+Sca1+ as HSC, CD150+CD48+Flt3-Lin-cKit+Sca1+ as MPP2, CD150-CD48+Flt3-Lin-cKit+Sca1+ as MPP3, Flt3+Lin-cKit+Sca1+ as MPP4.

Genomic DNA Recovery from Nucleated Cells

Genomic DNA was extracted from nucleated cells using the DNAeasy Blood Kit (Qiagen, 69504).

Cells were suspended in 200 µl of PBS, and 20 µl of Protein K along with 200 µl of AL buffer were added. The suspension was incubated at 56°C for at least 15 minutes. Subsequently, 200 µl of 100% EtOH (Fisher, BP28184) was added, and the mixture was loaded onto a DNAeasy Blood column. After centrifugation at 12,000 rpm for 1 minute, the flow-through was discarded. Then, 500 µl of AW1 buffer was loaded onto the column. After centrifugation at 12,000 rpm for 1 minute, the flow-through was discarded. Next, 500 µl of AW2 buffer was loaded onto the column. After centrifugation at 12,000 rpm for 1 minute, the flow-through was discarded. The column was then centrifuged at 12,000 rpm for 1 minute to remove any residual wash buffer. Afterward, the column was placed into a recovery Eppendorf tube. Next, 25 µl of AE buffer was added, and the genomic DNA was eluted by centrifugation at 12,000g for 1 minute.

Amplification of Barcodes

For both genomic DNA and cDNA (25 µl), 0.25 µl of 100 µM primers (F: CACGAGGTGGCAGTGGCCAGATAC, R for cDNA: GGCAAACAACAGATGGCTGGCAACTAG, R for genomic DNA: GGGACAGCAGAGATCCAGTTTGGTTAGTAC, F for Scratchpad: tctagacgtttaaactagcctgaggattcc, R for Scratchpad: ttgattcgaagttgagctcgactagctagG) and 25 µl of KOD One (Toyobo KMM-101) were added. The mixture was incubated in a thermal cycler with the following protocol: 98°C for 2 minutes, 22 cycles of (98°C for 10 seconds, 60°C for 30 seconds, and 68°C for 30 seconds), followed by 72°C for 10 minutes. To the PCR product, 90 µl of AMPureXP beads (Beckman Coulter, A63881) were added and incubated at room temperature for 15 minutes. The mixture was placed into a magnetic stand for 5 minutes, and the supernatant was removed. Subsequently, 200 µl of 80% EtOH was added and left for 1 minute, and the supernatant was removed. The beads were resuspended in 50 µl of Nuclease-Free water (Invitrogen, AM9932) and incubated at room temperature for 15 minutes. Finally, the mixture was placed into a magnetic stand for 5 minutes and the supernatant was collected. For secondary PCR, to 10 µl of the PCR products, 0.1 µl each of 100 µM primers (F: CTTTCCCTACACGACGCTCTTCCGATCTCACGAGGTGGCAGTGGCCAGATAC, R: GAGTTCAGACGTGTGCTCTTCCGATCTGGCAAACAACAGATGGCTGGCAACTAG, F for Scratchpad: CTTTCCCTACACGACGCTCTTCCGATCTtctagacgtttaaactagcctgaggattcc, R for Scratchpad: GAGTTCAGACGTGTGCTCTTCCGATCTttgattcgaagttgagctcgactagctagG) and 10 µl of KOD One (Toyobo KMM-101) were added. The mixture was incubated in a thermal cycler with the following protocol: 98°C for 2 minutes, 10 cycles of (98° C for 10 seconds, 60°C for 30 seconds, and 68°C for 30 seconds), followed by 72°C for 10 minutes. To the PCR product, 36 µl of AMPureXP beads (Beckman Coulter, A63881) were added and incubated at room temperature for 15 minutes. The mixture was placed into a magnetic stand for 5 minutes, and the supernatant was removed. Subsequently, 200 µl of 80% EtOH was added and left for 1 minute, and the supernatant was removed.

The beads were resuspended in 50 µl of Nuclease-Free water (Invitrogen, AM9932) and incubated at room temperature for 15 minutes. Finally, the mixture was placed into a magnetic stand for 5 minutes and the supernatant was collected. For indexing PCR, to 1 ng of secondary PCR product, 1 µl each of 10 µM primers (F: AATGATACGGCGACCACCGAGATCTACACNNNNNNNNACACTCTTTCCCTACACGACGC TCTTCCGATC*T, R: NEBNext® Multiplex Oligos for Illumina® (Index Primers Set, NEB E7335S, E7500S, E7710S)) and 10 µl of KOD One (Toyobo KMM-101) were added. The mixture was incubated in a thermal cycler with the following protocol: 98°C for 2 minutes, 5 cycles of (98°C for 10 seconds, 60°C for 30 seconds, and 68°C for 30 seconds), followed by 72°C for 10 minutes. To the indexing PCR product, 36 µl of AMPureXP beads (Beckman Coulter, A63881) were added and incubated at room temperature for 15 minutes. The mixture was placed into a magnetic stand for 5 minutes, and the supernatant was removed. Subsequently, 200 µl of 80% EtOH was added and left for 1 minute, and the supernatant was removed. The beads were resuspended in 50 µl of Nuclease-Free water (Invitrogen, AM9932) and incubated at room temperature for 15 minutes. Finally, the mixture was placed into a magnetic stand for 5 minutes and the supernatant was collected. Individual libraries were diluted to 4nM and pooled for sequencing. Pools were sequenced with 100 cycle run kits (75bp Read1, 8bp Index1, 8bp Index2) on the NextSeq 1000 P1 Sequencing System (Illumina). For Scratchpad, pools were sequenced with 300 cycle run kits (300bp Read1, 8bp Index1, 8bp Index2) on the NextSeq 1000 P1 Sequencing System (Illumina).

The modification of the barcode system for phylogeny analysis

The rate at which mutations accumulate in scratchpad is a critical factor in tracing. If the accumulation of mutations is too rapid, labeling might be completed before a sufficient number of descendants are formed from HSCs. It leads to the labeling of an insufficient number of subclones from the pearent cell. On the other hand, if labeling is too slow, tracing might not refrect early fate segregation from HSCs. Editing efficiency at the target sequences is influenced by mismatches between the target sequence and its sgRNA. It is known that editing efficiency decreases more significantly when mutations occur closer to the PAM sequence[23]. To address this, we created three types of Scratchpads—"V1," "V2," and "V3"—with target sites exhibiting progressively lower editing efficiencies by CRISPR/Cas9 in the order of V3 < V2 < V1 (Extended Data Fig. 2b). These three types were transduced into immunophenotypic HSCs from Cas9-EGFP mice, followed by an analysis of the editing pattern of the scratchpad 14 days after transduction (Extended Data Fig. 2c). Most of the scratchpads produced in V1-V3 were unique to each clone (Extended Data Fig. 2d). A decrease in editing efficiency and large deletions were observed in V2 and V3 (Extended Data Fig. 2e-f). The number of subclones per clone was greatest in V3 (Extended Data Fig. 2g). These findings indicate that V3 was the most suitable for phylogeny analysis. Based on these results, we used V3 in this paper.

Single cell RNA-seq

scRNA-seq libraries were constructed using the Chromium Single Cell 3′v3.1 Reagent Kit (10x Genomics, PN-1000269, PN-1000127, and PN-1000213/2000240) according to the manufacturer's protocol. Briefly, the post-sorting sample volume was reduced. Cells were then loaded into each channel with a target output of approximately 5,000 cells. Single cells were encapsulated into emulsion droplets using the Chromium Controller (10x Genomics). Reverse transcription, fragmentation, and indexing PCR were performed in a thermal cycler. cDNA and final libraries were purified using AMPureXP beads (Beckman Coulter, A63881). A random DNA barcode sequence was amplified using the SI primer (10x Genomics, PN-2000095) and the Reverse primer (GAGTTCAGACGTGTGCTCTTCCGATCTGACCTCCCTAGCAAACTGGGGCACAAG) following this protocol: 98°C for 2 minutes, 10 cycles of (98°C for 10 seconds, 60°C for 30 seconds, and 68°C for 30 seconds), and 72°C for 10 minutes. Indexing PCR was then performed using the SI primer and NEBNext® Multiplex Oligos for Illumina® (Index Primers Set, NEB E7335S) following this protocol: 98°C for 2 minutes, 5 cycles of (98°C for 10 seconds, 60°C for 30 seconds, and 68°C for 30 seconds), and 72°C for 10 minutes. Amplified cDNA and final libraries were evaluated on an Agilent BioAnalyzer using a High Sensitivity DNA Kit (Agilent Technologies, 5067-5584 and 5067-5585). Individual libraries were diluted to 4nM and pooled for sequencing. The pools were sequenced using 150-cycle run kits (26bp Read1, 8bp Index1, and 90bp Read2) on the Nova-seq SP Sequencing System (Illumina).

Screening vector validation

First, 10 μg of vector was incubated with 5 μl of NsiI (NEB R0127S), 5 μl of MscI (NEB R0534S), and 10 μl of rCutsmart, making up to a total volume of 100 μl with water. The reaction was incubated at 37°C overnight. After gel electrophoresis, the vector was purified into 50 μl of water using a column (Qiagen Cat#: 28706). For 50 μl of the extracted vector, 5 μl of T4 polymerase (NEB M0203), 10 μl of rCutsmart, and water were added to make up the total volume of 100 μl. This mixture was incubated at 12°C for 15 minutes. The vector was then purified into 50 μl of water using a column (Qiagen Cat#: 28706). Next, for 50 μl of the extracted vector, 2 μl of Exonuclease V (NEB M0345S), 10 μl of 10 mM ATP, 10 μl of NEBuffer4, and water were added to make up the total volume of 100 μl. The mixture was incubated at 37°C for 30 minutes. The vector was purified into 50 μl of water using a column (Qiagen Cat#: 28706). For the next step, for 50 μl of the extracted vector, 5 μl of T4 ligase (NEB M0202L), 10 μl of T4 ligase buffer, and water were added to make up the total volume of 100 μl. The mixture was incubated at room temperature overnight. The vector was purified into 50 μl of water using a column (Qiagen Cat#: 28706). For 50 μl of the extracted vector, 50 μl of KOD One (Toyobo KMM-101), 1 μM of F primer

(CTTTCCCTACACGACGCTCTTCCGATCTCTTGTGGAAAGGACGAAACACCG), and 1 μM of R primer (GAGTTCAGACGTGTGCTCTTCCGATCTGGCAAACAACAGATGGCTGGCAACTAGC) were added. The mixture was incubated in a thermal cycler with the following protocol: 98°C for 2 minutes, 15 cycles of (98°C for 10 seconds, 60°C for 30 seconds, and 68°C for 30 seconds), and 72°C for 10 minutes. To the PCR product, 180 μl of AMPureXP beads (Beckman Coulter, A63881) was added and incubated at room temperature for 15 minutes. The mixture was placed in a magnetic stand for 5 minutes, and the supernatant was removed. Then, 200 μl of 80% EtOH was added and left for 1 minute, after which the supernatant was removed. The beads were resuspended in 50 μl of Nuclease-Free water (Invitrogen, AM9932) and incubated at room temperature for 15 minutes. Finally, the mixture was placed in a magnetic stand for 5 minutes and the supernatant was collected. For 8 μl of the PCR product, 10 μl of KOD One (Toyobo KMM-101), 1 μl of NEBNext® multiplex primer index 1 primer (NEB E7335S), and 1 μl of 10 μM primer (AATGATACGGCGACCACCGAGATCTACACTATAGCCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T) were added. The mixture was incubated in a thermal cycler with the following protocol: 98°C for 2 minutes, 15 cycles of (98°C for 10 seconds, 60°C for 30 seconds, and 68°C for 30 seconds), and 72°C for 10 minutes. To the PCR product, 36 μl of AMPureXP beads (Beckman Coulter, A63881) was added and incubated at room temperature for 15 minutes. The mixture was placed into a magnetic stand for 5 minutes, and the supernatant was removed. Then, 200 μl of 80% EtOH was added and left for 1 minute, after which the supernatant was removed. The beads were resuspended in 50 μl of Nuclease-Free water (Invitrogen, AM9932) and incubated at room temperature for 15 minutes. Finally, the mixture was placed in a magnetic stand for 5 minutes and the supernatant was collected.

Combined sgRNA-random DNA barcode library validation

To create the random DNA barcode-sgRNA lookup table for the screening vector (Extended Data Fig. 6a), two methods were considered to examine the correspondence between the random DNA barcode and the sgRNA. In Method A, the entire region from the sgRNA to the random DNA barcode was amplified by PCR, followed by ligation of the ends of the PCR product to bring the sgRNA and random DNA barcode information into proximity. Subsequently, the adjacent sgRNA and random DNA barcode information were amplified by PCR. In Method B, the extra sequence between the sgRNA and the random DNA barcode was excised using a restriction enzyme, followed by ligation to bring the sgRNA and random DNA barcode information into proximity. The adjacent sgRNA and random DNA barcode information were then amplified by PCR (Extended Data Fig. 6b). Template switching during PCR is known to cause information from multiple template molecules to be mixed into one molecule of PCR product[47]. To test for template switching, two vectors with known structures were

mixed in equal amounts, and the mixture was processed using both methods. Then, the final products were sequenced (Extended Data Fig. 6c). As a result, many template switches were observed with Method A. In contrast, only a few template switches were observed with Method B (Extended Data Fig. 6b-c). Subsequently, the vector containing the random DNA barcode and the sgRNA library was validated using Method B. Among 97.7% of the random DNA barcodes, over 70% of the sgRNA information originated from a single sgRNA (Extended Data Fig. 6d), indicating the successful integration of the random DNA barcode and sgRNA information.

Barcode analysis

Random DNA barcodes were extracted and clustered using Bartender[48]. The combination of the random DNA barcode and scratchpad was extracted using FASTX-toolkit, Seqkit[49], Cutadapt (DOI: 10.14806/ej.17.1.200), and Python scripts. Sequences were then clustered using Bartender. After filtering, the count data was normalized by $\log_{10}$(count frequency + 0.000001). K-means clustering was performed using ComplexHeatmap[50] (DOI: 10.1002/imt2.43). Clusters were annotated based on their lineage output patterns, and annotated clusters were used for downstream analysis. Barcodes representing different conditions within the same mouse, distinguished by their initial letters (GTG, GAC, GGA, GCT), were separated after k-means clustering. Random DNA barcodes from the single-cell RNA-seq library were extracted using FASTX-toolkit and Python scripts. Sequences were then clustered using Bartender. For vector validation screening, sgRNA and random DNA barcode sequences were extracted using Bartender. The sgRNA and random DNA barcode sequences on the same reads were connected using Python scripts, and sequences were clustered using Bartender. Scratchpad sequences were reconstructed using ClustalW[51].

Statistics and reproducibility

Statistical and graphical data analyses were performed using GraphPad Prism 9 or other specified software. The number of samples and details of the statistical tests used are provided in the figures or figure legends.

References

1. Yamamoto, R., Wilkinson, A.C., Ooehara, J., Lan, X., Lai, C.Y., Nakauchi, Y., Pritchard, J.K., and Nakauchi, H. (2018). Large-Scale Clonal Analysis Resolves Aging of the Mouse Hematopoietic Stem Cell Compartment. Cell Stem Cell *22*, 600-607.e604. 10.1016/j.stem.2018.03.013.
2. Ross, J.B., Myers, L.M., Noh, J.J., Collins, M.M., Carmody, A.B., Messer, R.J., Dhuey, E., Hasenkrug, K.J., and Weissman, I.L. (2024). Depleting myeloid-biased haematopoietic stem cells rejuvenates aged immunity. Nature *628*, 162-170. 10.1038/s41586-024-07238-

x.

3. Busch, K., Klapproth, K., Barile, M., Flossdorf, M., Holland-Letz, T., Schlenner, S.M., Reth, M., Höfer, T., and Rodewald, H.R. (2015). Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. Nature *518*, 542-546. 10.1038/nature14242.
4. Sawai, C.M., Babovic, S., Upadhaya, S., Knapp, D., Lavin, Y., Lau, C.M., Goloborodko, A., Feng, J., Fujisaki, J., Ding, L., et al. (2016). Hematopoietic Stem Cells Are the Major Source of Multilineage Hematopoiesis in Adult Animals. Immunity *45*, 597-609. 10.1016/j.immuni.2016.08.007.
5. Chen, J.Y., Miyanishi, M., Wang, S.K., Yamazaki, S., Sinha, R., Kao, K.S., Seita, J., Sahoo, D., Nakauchi, H., and Weissman, I.L. (2016). Hoxb5 marks long-term haematopoietic stem cells and reveals a homogenous perivascular niche. Nature *530*, 223-227. 10.1038/nature16943.
6. Rodriguez-Fraticelli, A.E., Wolock, S.L., Weinreb, C.S., Panero, R., Patel, S.H., Jankovic, M., Sun, J., Calogero, R.A., Klein, A.M., and Camargo, F.D. (2018). Clonal analysis of lineage fate in native haematopoiesis. Nature *553*, 212-216. 10.1038/nature25168.
7. Pei, W., Shang, F., Wang, X., Fanti, A.K., Greco, A., Busch, K., Klapproth, K., Zhang, Q., Quedenau, C., Sauer, S., et al. (2020). Resolving Fates and Single-Cell Transcriptomes of Hematopoietic Stem Cell Clones by PolyloxExpress Barcoding. Cell Stem Cell *27*, 383-395.e388. 10.1016/j.stem.2020.07.018.
8. Pei, W., Feyerabend, T.B., Rössler, J., Wang, X., Postrach, D., Busch, K., Rode, I., Klapproth, K., Dietlein, N., Quedenau, C., et al. (2017). Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. Nature *548*, 456-460. 10.1038/nature23653.
9. Yamamoto, R., Morita, Y., Ooehara, J., Hamanaka, S., Onodera, M., Rudolph, K.L., Ema, H., and Nakauchi, H. (2013). Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. Cell *154*, 1112-1126. 10.1016/j.cell.2013.08.007.
10. Carrelha, J., Meng, Y., Kettyle, L.M., Luis, T.C., Norfo, R., Alcolea, V., Boukarabila, H., Grasso, F., Gambardella, A., Grover, A., et al. (2018). Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. Nature *554*, 106-111. 10.1038/nature25455.
11. Yu, V.W.C., Yusuf, R.Z., Oki, T., Wu, J., Saez, B., Wang, X., Cook, C., Baryawno, N., Ziller, M.J., Lee, E., et al. (2017). Epigenetic Memory Underlies Cell-Autonomous Heterogeneous Behavior of Hematopoietic Stem Cells. Cell *168*, 944-945. 10.1016/j.cell.2017.02.010.
12. Challen, G.A., Boles, N.C., Chambers, S.M., and Goodell, M.A. (2010). Distinct

hematopoietic stem cell subtypes are differentially regulated by TGF-beta1. Cell Stem Cell *6*, 265-278. 10.1016/j.stem.2010.02.002.

13. Sanjuan-Pla, A., Macaulay, I.C., Jensen, C.T., Woll, P.S., Luis, T.C., Mead, A., Moore, S., Carella, C., Matsuoka, S., Bouriez Jones, T., et al. (2013). Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. Nature *502*, 232-236. 10.1038/nature12495.
14. Rodriguez-Fraticelli, A.E., Weinreb, C., Wang, S.W., Migueles, R.P., Jankovic, M., Usart, M., Klein, A.M., Lowell, S., and Camargo, F.D. (2020). Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis. Nature *583*, 585-589. 10.1038/s41586-020-2503-6.
15. Naik, S.H., Perié, L., Swart, E., Gerlach, C., van Rooij, N., de Boer, R.J., and Schumacher, T.N. (2013). Diverse and heritable lineage imprinting of early haematopoietic progenitors. Nature *496*, 229-232. 10.1038/nature12013.
16. Fan, X., Wu, C., Truitt, L.L., Espinoza, D.A., Sellers, S., Bonifacino, A., Zhou, Y., Cordes, S.F., Krouse, A., Metzger, M., et al. (2020). Clonal tracking of erythropoiesis in rhesus macaques. Haematologica *105*, 1813-1824. 10.3324/haematol.2019.231811.
17. Calabria, A., Spinozzi, G., Cesana, D., Buscaroli, E., Benedicenti, F., Pais, G., Gazzo, F., Scala, S., Lidonnici, M.R., Scaramuzza, S., et al. (2024). Long-term lineage commitment in haematopoietic stem cell gene therapy. Nature *636*, 162-171. 10.1038/s41586-024-08250-x.
18. Morita, Y., Ema, H., and Nakauchi, H. (2010). Heterogeneity and hierarchy within the most primitive hematopoietic stem cell compartment. J Exp Med *207*, 1173-1182. 10.1084/jem.20091318.
19. Ito, K., Turcotte, R., Cui, J., Zimmerman, S.E., Pinho, S., Mizoguchi, T., Arai, F., Runnels, J.M., Alt, C., Teruya-Feldstein, J., et al. (2016). Self-renewal of a purified Tie2+ hematopoietic stem cell population relies on mitochondrial clearance. Science *354*, 1156-1160. 10.1126/science.aaf5530.
20. Loeffler, D., Wehling, A., Schneiter, F., Zhang, Y., Müller-Bötticher, N., Hoppe, P.S., Hilsenbeck, O., Kokkaliaris, K.D., Endele, M., and Schroeder, T. (2019). Asymmetric lysosome inheritance predicts activation of haematopoietic stem cells. Nature *573*, 426-429. 10.1038/s41586-019-1531-6.
21. McKenna, A., Findlay, G.M., Gagnon, J.A., Horwitz, M.S., Schier, A.F., and Shendure, J. (2016). Whole-organism lineage tracing by combinatorial and cumulative genome editing. Science *353*, aaf7907. 10.1126/science.aaf7907.
22. Perli, S.D., Cui, C.H., and Lu, T.K. (2016). Continuous genetic recording with self-targeting CRISPR-Cas in human cells. Science *353*. 10.1126/science.aag0511.
23. Chan, M.M., Smith, Z.D., Grosswendt, S., Kretzmer, H., Norman, T.M., Adamson, B.,

Jost, M., Quinn, J.J., Yang, D., Jones, M.G., et al. (2019). Molecular recording of mammalian embryogenesis. Nature *570*, 77-82. 10.1038/s41586-019-1184-5.

24. Yang, D., Jones, M.G., Naranjo, S., Rideout, W.M., 3rd, Min, K.H.J., Ho, R., Wu, W., Replogle, J.M., Page, J.L., Quinn, J.J., et al. (2022). Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor evolution. Cell *185*, 1905-1923.e1925. 10.1016/j.cell.2022.04.015.
25. Bowling, S., Sritharan, D., Osorio, F.G., Nguyen, M., Cheung, P., Rodriguez-Fraticelli, A., Patel, S., Yuan, W.C., Fujiwara, Y., Li, B.E., et al. (2020). An Engineered CRISPR-Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells. Cell *181*, 1410-1422.e1427. 10.1016/j.cell.2020.04.048.
26. Li, L., Bowling, S., McGeary, S.E., Yu, Q., Lemke, B., Alcedo, K., Jia, Y., Liu, X., Ferreira, M., Klein, A.M., et al. (2023). A mouse model with high clonal barcode diversity for joint lineage, transcriptomic, and epigenomic profiling in single cells. Cell. 10.1016/j.cell.2023.09.019.
27. Weng, C., Yu, F., Yang, D., Poeschla, M., Liggett, L.A., Jones, M.G., Qiu, X., Wahlster, L., Caulier, A., Hussmann, J.A., et al. (2024). Deciphering cell states and genealogies of human haematopoiesis. Nature *627*, 389-398. 10.1038/s41586-024-07066-z.
28. Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., 2nd, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science *352*, 189-196. 10.1126/science.aad0501.
29. Becht, E., McInnes, L., Healy, J., Dutertre, C.A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 10.1038/nbt.4314.
30. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *172*, 650-665. 10.1016/j.cell.2018.01.029.
31. Kataoka, K., Sato, T., Yoshimi, A., Goyama, S., Tsuruta, T., Kobayashi, H., Shimabe, M., Arai, S., Nakagawa, M., Imai, Y., et al. (2011). Evi1 is essential for hematopoietic stem cell self-renewal, and its expression marks hematopoietic cells with long-term multilineage repopulating activity. J Exp Med *208*, 2403-2416. 10.1084/jem.20110447.
32. Calvanese, V., Nguyen, A.T., Bolan, T.J., Vavilina, A., Su, T., Lee, L.K., Wang, Y., Lay, F.D., Magnusson, M., Crooks, G.M., et al. (2019). MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment. Nature *576*, 281-286. 10.1038/s41586-019-1790-2.
33. Yamazaki, S., Ema, H., Karlsson, G., Yamaguchi, T., Miyoshi, H., Shioda, S., Taketo, M.M.,

Karlsson, S., Iwama, A., and Nakauchi, H. (2011). Nonmyelinating Schwann cells maintain hematopoietic stem cell hibernation in the bone marrow niche. Cell *147*, 1146-1158. 10.1016/j.cell.2011.09.053.

34. Meng, Y., Carrelha, J., Drissen, R., Ren, X., Zhang, B., Gambardella, A., Valletta, S., Thongjuea, S., Jacobsen, S.E., and Nerlov, C. (2023). Epigenetic programming defines haematopoietic stem cell fate restriction. Nat Cell Biol *25*, 812-822. 10.1038/s41556-023-01137-5.
35. Vitali, C., Bassani, C., Chiodoni, C., Fellini, E., Guarnotta, C., Miotti, S., Sangaletti, S., Fuligni, F., De Cecco, L., Piccaluga, P.P., et al. (2015). SOCS2 Controls Proliferation and Stemness of Hematopoietic Cells under Stress Conditions and Its Deregulation Marks Unfavorable Acute Leukemias. Cancer Res *75*, 2387-2399. 10.1158/0008-5472.can-14-3625.
36. Santaguida, M., Schepers, K., King, B., Sabnis, A.J., Forsberg, E.C., Attema, J.L., Braun, B.S., and Passegué, E. (2009). JunB protects against myeloid malignancies by limiting hematopoietic stem cell proliferation and differentiation without affecting self-renewal. Cancer Cell *15*, 341-352. 10.1016/j.ccr.2009.02.016.
37. Jackson, J.T., O'Donnell, K., Light, A., Goh, W., Huntington, N.D., Tarlinton, D.M., and McCormack, M.P. (2020). Hhex regulates murine lymphoid progenitor survival independently of Stat5 and Cdkn2a. Eur J Immunol *50*, 959-971. 10.1002/eji.201948371.
38. Goodings, C., Smith, E., Mathias, E., Elliott, N., Cleveland, S.M., Tripathi, R.M., Layer, J.H., Chen, X., Guo, Y., Shyr, Y., et al. (2015). Hhex is Required at Multiple Stages of Adult Hematopoietic Stem and Progenitor Cell Differentiation. Stem Cells *33*, 2628-2641. 10.1002/stem.2049.
39. Jackson, J.T., Nasa, C., Shi, W., Huntington, N.D., Bogue, C.W., Alexander, W.S., and McCormack, M.P. (2015). A crucial role for the homeodomain transcription factor Hhex in lymphopoiesis. Blood *125*, 803-814. 10.1182/blood-2014-06-579813.
40. Migueles, R.P., Shaw, L., Rodrigues, N.P., May, G., Henseleit, K., Anderson, K.G., Goker, H., Jones, C.M., de Bruijn, M.F., Brickman, J.M., and Enver, T. (2017). Transcriptional regulation of Hhex in hematopoiesis and hematopoietic stem cell ontogeny. Dev Biol *424*, 236-245. 10.1016/j.ydbio.2016.12.021.
41. Oguro, H., Ding, L., and Morrison, S.J. (2013). SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. Cell Stem Cell *13*, 102-116. 10.1016/j.stem.2013.05.014.
42. Pietras, E.M., Reynaud, D., Kang, Y.A., Carlin, D., Calero-Nieto, F.J., Leavitt, A.D., Stuart, J.M., Göttgens, B., and Passegué, E. (2015). Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative

Conditions. Cell Stem Cell *17*, 35-46. 10.1016/j.stem.2015.05.003.

43. Aksöz, M., Gafencu, G.A., Stoilova, B., Buono, M., Zhang, Y., Turkalj, S., Meng, Y., Jakobsen, N.A., Metzner, M., Clark, S.A., et al. (2024). Hematopoietic stem cell heterogeneity and age-associated platelet bias are evolutionarily conserved. Sci Immunol *9*, eadk3469. 10.1126/sciimmunol.adk3469.

44. Belander Strålin, K., Carrelha, J., Winroth, A., Ziegenhain, C., Hagemann-Jensen, M., Kettyle, L.M., Hillen, A., Högstrand, K., Markljung, E., Grasso, F., et al. (2023). Platelet and myeloid lineage biases of transplanted single perinatal mouse hematopoietic stem cells. In Cell Res, pp. 883-886. 10.1038/s41422-023-00866-4.

45. Grover, A., Sanjuan-Pla, A., Thongjuea, S., Carrelha, J., Giustacchini, A., Gambardella, A., Macaulay, I., Mancini, E., Luis, T.C., Mead, A., et al. (2016). Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells. Nat Commun *7*, 11075. 10.1038/ncomms11075.

46. Wilkinson, A.C., Ishida, R., Kikuchi, M., Sudo, K., Morita, M., Crisostomo, R.V., Yamamoto, R., Loh, K.M., Nakamura, Y., Watanabe, M., et al. (2019). Long-term ex vivo haematopoietic-stem-cell expansion allows nonconditioned transplantation. Nature *571*, 117-121. 10.1038/s41586-019-1244-x.

47. Davidsson, M., Diaz-Fernandez, P., Schwich, O.D., Torroba, M., Wang, G., and Björklund, T. (2016). A novel process of viral vector barcoding and library preparation enables high-diversity library generation and recombination-free paired-end sequencing. Sci Rep *6*, 37563. 10.1038/srep37563.

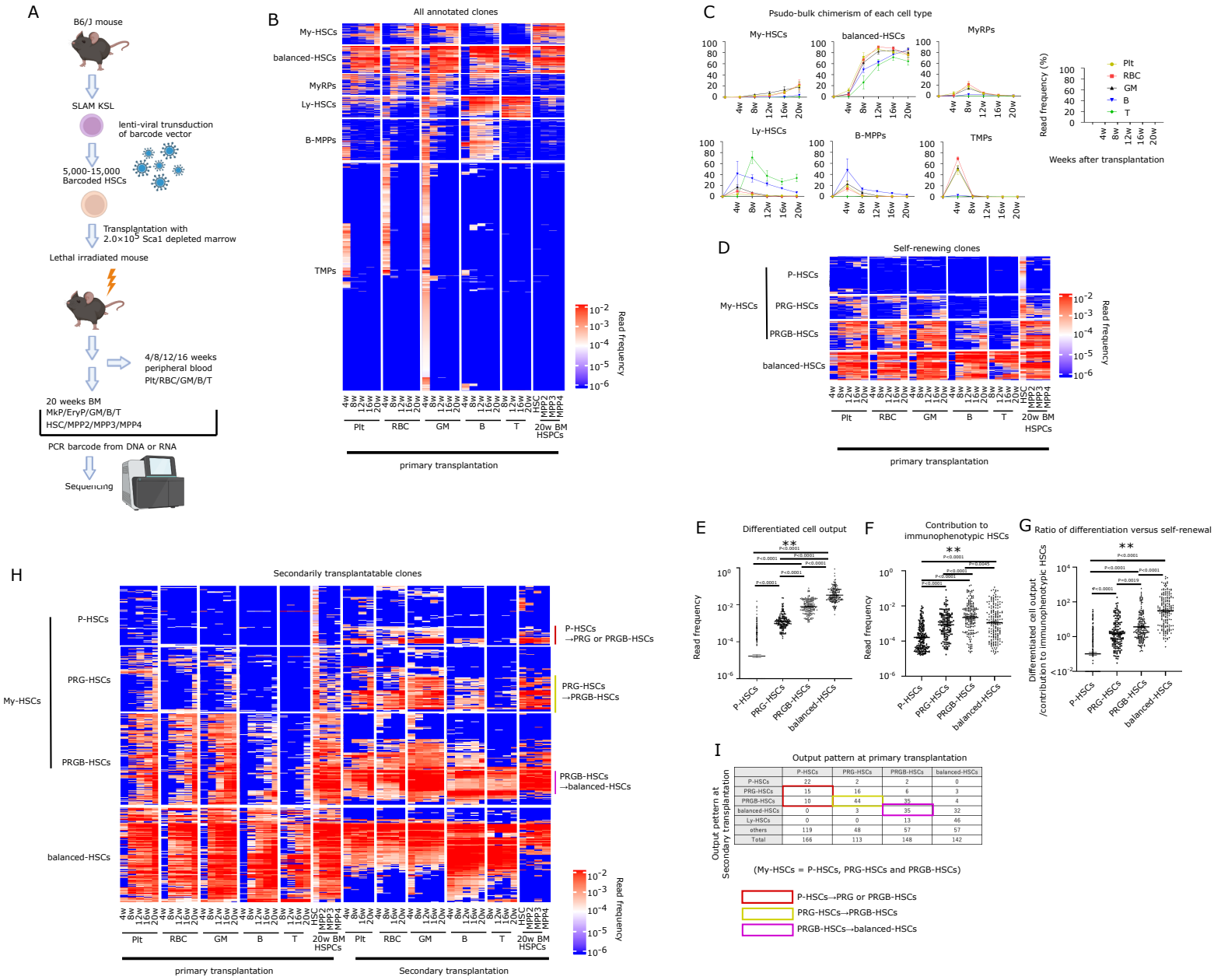
48. Zhao, L., Liu, Z., Levy, S.F., and Wu, S. (2018). Bartender: a fast and accurate clustering algorithm to count barcode reads. Bioinformatics *34*, 739-747. 10.1093/bioinformatics/btx655.

49. Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One *11*, e0163962. 10.1371/journal.pone.0163962.

50. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847-2849. 10.1093/bioinformatics/btw313.

51. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res *22*, 4673-4680. 10.1093/nar/22.22.4673.

Figure 1 My-HSCs favor self-renewal and can transition to balanced-HSCs



	P-HSCs	PRG-HSCs	PRGB-HSCs	balanced-HSCs
P-HSCs	22	2	2	0
PRG-HSCs	15	16	6	3
PRGB-HSCs	10	44	35	4
balanced-HSCs	0	3	35	32
Ly-HSCs	0	0	13	46
others	119	48	57	57
Total	166	113	148	142

Figure 2 Balanced-HSCs can transition into My-HSCs

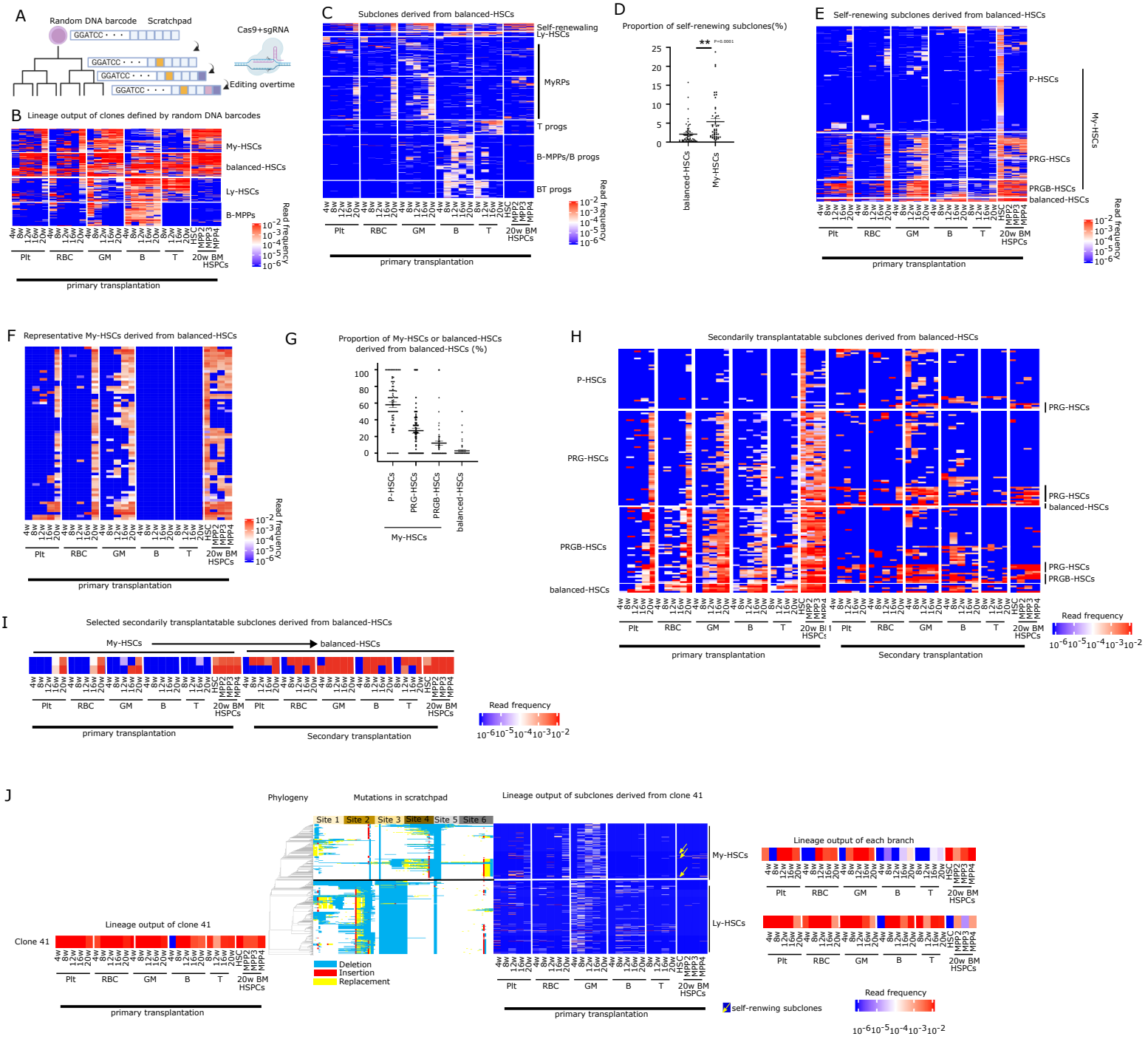


Figure 3 Combinaing fate tracing and scRNA-seq identified transcriptional differences between My-HSCs and balanced-HSCs.

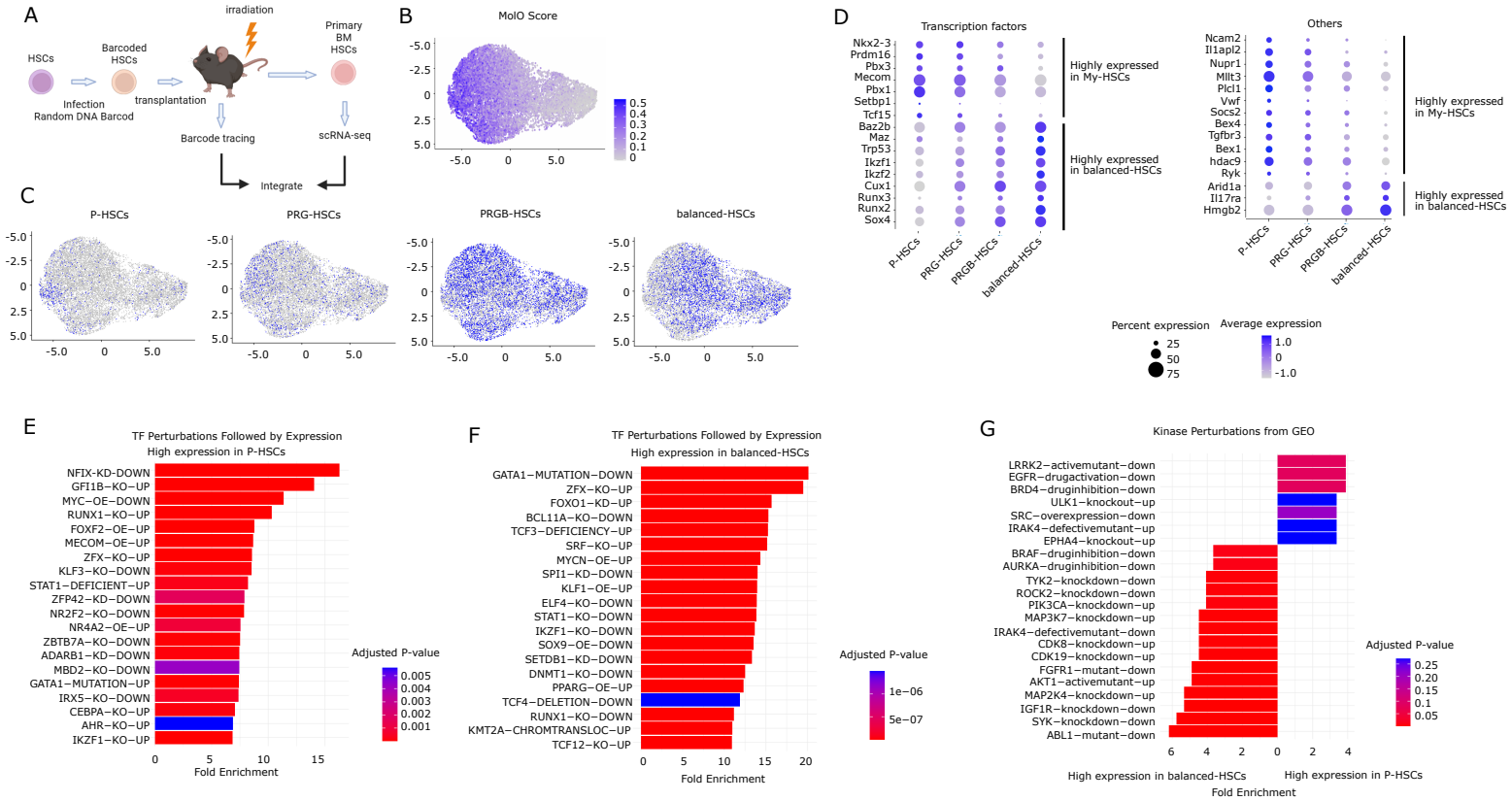


Figure 4 Hhex is necessary for Ly-HSCs fate and suppression of My-HSC fate

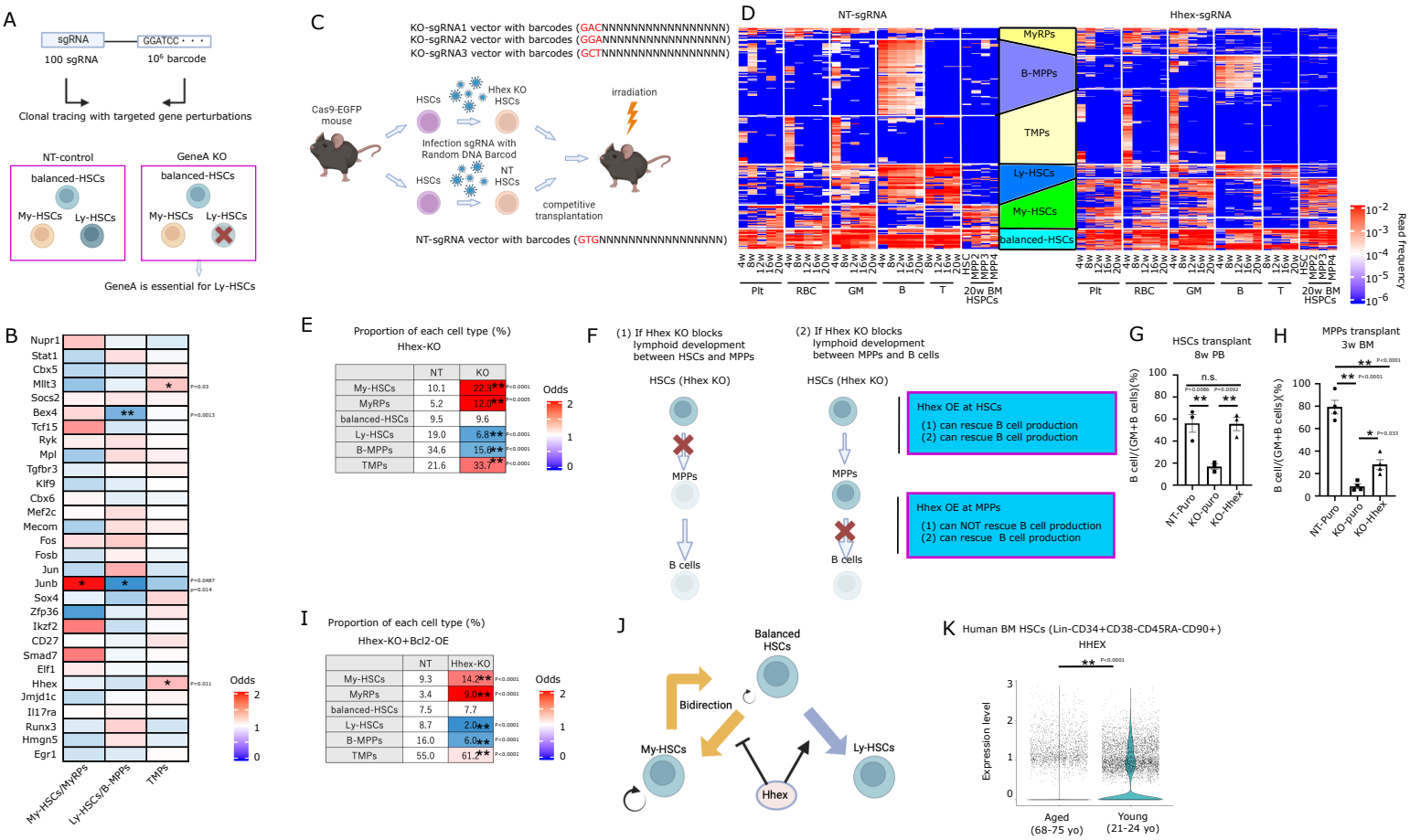


Figure 1 | My-HSCs favor self-renewal and can transition to balanced-HSCs.

(A) Schematic of the experimental setup.

(B) Heatmap displaying clones from the immunophenotypic HSC fraction, with rows representing unique DNA barcodes and columns representing time points (4w, 8w, 12w, 16w, 20w) and cell types (Plt, RBC, GM, B, T, HSPCs). Colors indicate read frequency. Clusters are annotated based on output patterns (My-HSCs, balanced-HSCs, MyRPs, Ly-HSCs, B-MPPs, TMPs).

(C) Pseudo-bulk analysis of the read frequencies of all clones within each cluster (My-HSCs, balanced-HSCs, MyRPs, Ly-HSCs, B-MPPs, TMPs). The x-axis represents weeks after transplantation. Lineage colors are as follows: Orange for Plt, Red for RBC, Black for GM, Blue for B, Green for T (b, c, n=3964 clones).

(D) Heatmap of clones reconstituting the 20w BM immunophenotypic HSC fraction. Clusters are annotated based on output patterns (P-HSCs, PRG-HSCs, PRGB-HSCs, balanced-HSCs) (n=754 clones).

(E) Dot plot showing the highest read frequency of clones across all differentiated cells and all time points for each cluster (P-HSCs, PRG-HSCs, PRGB-HSCs, balanced-HSCs).

(F) Dot plot showing the read frequency of the 20w BM immunophenotypic HSC fraction of clones for each cluster (P-HSCs, PRG-HSCs, PRGB-HSCs, balanced-HSCs).

(G) Dot plot showing the ratio of differentiation versus self-renewal (highest read frequency of clones across all differentiated cells and all time points / read frequency of the 20w BM immunophenotypic HSC fraction) of clones for each cluster (P-HSCs, PRG-HSCs, PRGB-HSCs, balanced-HSCs)

In (E), (F) and (G), P-HSCs n=238, PRG-HSCs: n=155, PRGB-HSCs: n=180, balanced-HSCs: n=181).

(H), Heatmap of clones reconstituting the 20w BM immunophenotypic HSC fraction and secondary transplantation. (P-HSCs n=81, PRG-HSCs: n=80, PRGB-HSCs: n=103, balanced-HSCs: n=133).

(I) Table showing clone numbers based on output patterns at primary (P-HSCs, PRG-HSCs, PRGB-HSCs, balanced-HSCs) and secondary transplantation (P-HSCs, PRG-HSCs, PRGB-HSCs, balanced-HSCs, Ly-HSCs, B-MPPs, other, no output).

In (B)-(G), 4 biological replicates from 2 independent experiments. In (H)-(J), 3 biological replicates from 2 independent experiments). *P<0.05; **P<0.01. P values by one-way ANOVA followed by Tukey's multiple comparison test in (E)-(G).

Figure 2 | balanced-HSCs can transition into My-HSCs

(A) Schematic of the strategy for phylogeny analysis.

(B) Heatmap of clones from the immunophenotypic HSC fraction, with rows representing unique DNA barcodes and columns representing time points (4w, 8w, 12w, 16w, 20w) and cell types (Plt,

RBC, GM, B, T, HSPCs). Colors indicate read frequency. Clusters are annotated by output pattern (My-HSCs, balanced-HSCs, Ly-HSCs, B-MPPs) (n=405 clones).

(C) Heatmap of subclones derived from balanced-HSCs, reconstituting more than three time points or cell types (Plt, RBC, GM, B, T, HSPCs). Clusters are annotated by output pattern (Self-renewing, Ly-HSCs, MyRPs, T progs, B-MPPs/B progs, BT progs) (n=7673 subclones).

(D) Dot plot showing the proportion of subclones reconstituting the 20w BM immunophenotypic HSC fraction of all subclones within each clone in balanced-HSCs (n=73 clones) and My-HSCs (n=46 clones). Average and SEM are shown.

(E) Heatmap of subclones derived from balanced-HSCs, reconstituting the 20w BM immunophenotypic HSC fraction (n=757 subclones). The lineage output of each subclone was compared to the average outputs of clusters (P-HSCs, PRG-HSCs, PRGB-HSCs, and balanced-HSCs) in the reference heatmap (Figure 1D). Each subclone was then assigned to the cluster it was closest to.

(F) Representative heatmap of PRG-HSCs derived from balanced-HSCs (n=66 subclones).

(G) Dot plot showing the proportion of P-HSCs, PRG-HSCs, PRGB-HSCs, and balanced-HSCs of all self-renewing subclones derived from balanced-HSCs (n=73 clones).

(H) Heatmap of subclones derived from balanced-HSCs, reconstituting the 20w BM immunophenotypic HSC fraction and secondary transplantation (n=140 subclones).

(I) Representative heatmap of My-HSCs subclones derived from balanced-HSCs that transitioned to balanced-HSCs during secondary transplantation.

(J) Tree reconstruction of a representative balanced-HSC clone. Heatmap of the clonal output (left), scratchpad editing (middle), and sub-clonal output from the same clone (right), with rows representing unique edit patterns of scratchpad, columns representing target sites (middle), and time points/cell types (Plt, RBC, GM, B, T, HSPCs) (right). Edit patterns are color-coded (Blue: deletion, Yellow: different bp, Red: insertion) (middle); read frequency is color-coded (right).

In (B)-(G), 4 biological replicates from 3 independent experiments. In (H)-(I), 2 biological replicates. *P<0.05; **P<0.01. d, P values calculated by Student t-test.

Figure 3 | Combinaing fate tracing and scRNA-seq identified transcriptional differences between My-HSCs and balanced-HSCs.

(A) Schematic of the strategy for integrating scRNA-seq and barcoding information.

(B) Uniform manifold approximation and projection (UMAP) of cells from primary 20w BM immunophenotypic HSC fraction, with each cell color-coded according to the MolO scores.

(C) UMAP of cells from primary 20w BM immunophenotypic HSC fraction, with each cell color-coded according to the phenotypes of its barcode (P-HSCs, PRG-HSCs, PRGB-HSCs, balanced-HSCs

from Figure1D).

(D) Dot plot of differentially expressed genes between P-HSCs vs. balanced-HSCs or P and PRG-HSCs vs. balanced-HSCs, with dot size indicating percent expression and color code representing average expression. In (B)-(D), n=11,717 cells.

(E) Enrichr pathway analysis (TF perturbations followed by expression) of upregulated pathways in P-HSCs compared to balanced-HSCs. The X-axis represents fold enrichment, and the color code represents the adjusted p-value.

(F) Enrichr pathway analysis (TF perturbations followed by expression) of upregulated pathways in balanced-HSCs compared to P-HSCs. The X-axis represents fold enrichment, and the color code represents the adjusted p-value.

(G) Enrichr pathway analysis (Kinase perturbation from GEO) of differentially expressed pathways between P-HSCs and balanced-HSCs. X-axis represents fold enrichment; color-code represents adjusted p-value.

In (C)-(G), P-HSCs: n=631, PRG-HSCs: n=1282, PRBG-HSCs: n=4958, balanced-HSCs: n=3616 cells. In (B)-(G), 4 biological replicates from 2 independent experiments.

Figure 4 | Hhex is necessary for Ly-HSCs fate and suppression of My-HSC fate

(A) Schematic strategy for combining random DNA barcodes and CRISPR/Cas9 based sgRNA library screening, allowing for tracing individual clones with targeted gene perturbations.

(B) Heatmap displaying perturbation of the proportion of each cell type upon knockout of each gene. Rows represent genes; columns represent cell types (My-HSCs/MyRPs, Ly-HSCs/B-MPPs, TMPs) (n=7190 clones, 4 biological replicates from 2 independent experiments). Odds ratio color-coded.

(C) Experimental setup schematic for clonal tracing with selected gene perturbations.

(D) Heatmap showing clones from NT-sgRNA (left) and Hhex-sgRNA (right) groups, with rows as unique DNA barcodes and columns representing time points (4w, 8w, 12w, 16w, 20w) and cell types (Plat, RBC, GM, B, T, HSPCs). Read frequency is color-coded, and clusters are annotated by output pattern (My-HSCs, balanced-HSCs, MyRPs, Ly-HSCs, B-MPPs, TMPs).

(E) Table showing the proportion of each cell type in NT-sgRNA and Hhex-sgRNA groups, with odds ratio color-coded. (d-e, NT-sgRNA: n=306 clones, Hhex-sgRNA: n=924 clones, 4 biological replicates from 2 independent experiments).

(F) Schematic diagram of experiments adding back Hhex to Hhex-KO HSCs and MPPs derived from Hhex-KO HSCs.

(G) The ratio of peripheral blood B cells/(GM + B cells) 8 weeks after the transplantation of immunophenotypic HSCs.

(H) The ratio of BM B cells/(GM+B cells) 3 weeks after the transplantation of immunophenotypic MPPs.

(I) Table showing the proportion of each cell type in NT-sgRNA-Bcl2-OE and Hhex-sgRNA-Bcl2-OE groups (NT-sgRNA-Bcl2-OE: n=3100; Hhex-sgRNA-Bcl2-OE: n=4743, from 3 biological replicates).

(J) Schematic diagram illustrating how Hhex regulates the lymphoid versus myeloid fate within primitive HSC subsets.

(K) Violin plots showing the expression levels of HHEX in human BM HSCs ($Lin^{-}CD34^{+}CD38^{-}CD45RA^{-}CD90^{+}$) in aged (68–75 years old; n = 3) and young (21–24 years old; n = 3) groups with adjusted p-value displayed, based on data from Aksöz et al., 2024. *P<0.05; **P<0.01.

In (B), (E) and (I), P values calculated by the Fisher's exact test. In (G)-(H), P values calculated by one-way ANOVA followed by Tukey's multiple comparison test.