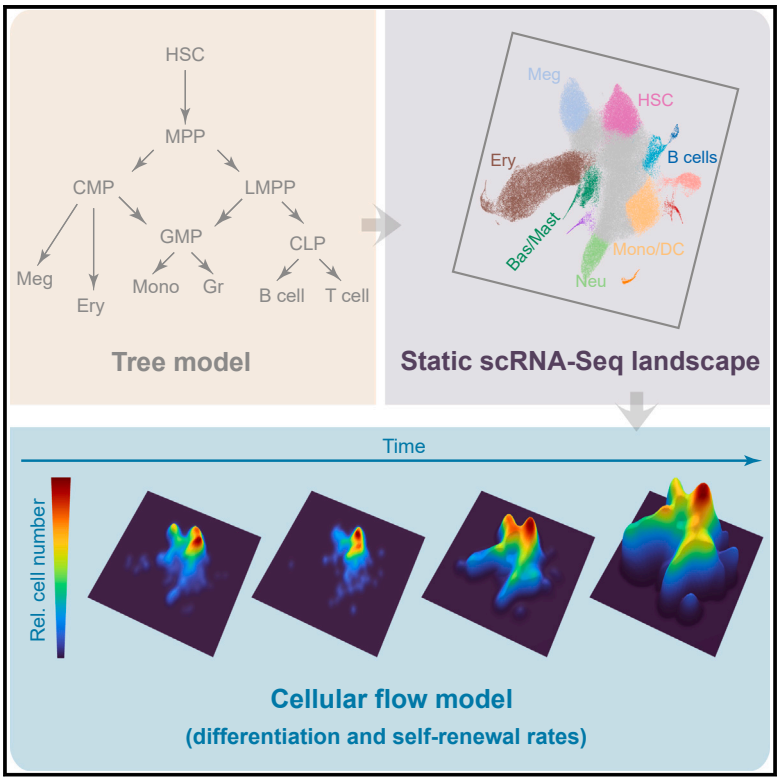
Cell Stem Cell

Resource

A time- and single-cell-resolved model of murine bone marrow hematopoiesis

Graphical abstract



Highlights

- Labeling and time-series scRNA-seq reveal kinetics of HSC multilineage differentiation
- Model captures *in vivo* real-time cell behavior connected with gene expression patterns
- Progenitor subpopulations display diverse self-renewal and differentiation properties
- Transplanted HSCs display accelerated stage- and lineage-specific differentiation

Authors

Iwo Kucinski, Joana Campos, Melania Barile, ..., Dónal O'Carroll, Kamil R. Kranc, Berthold Göttgens

Correspondence

donal.ocarroll@ed.ac.uk (D.O.), kamil.kranc@icr.ac.uk (K.R.K.), bg200@cam.ac.uk (B.G.)

In brief

Kucinski and colleagues construct a quantitative and real-time model of mouse bone marrow hematopoiesis by combining scRNA-seq and persistent HSC labeling technologies. The model reveals lineage- and stage-specific self-renewal and differentiation properties and explains how these are altered in a transplantation setting.

Kucinski et al., 2024, Cell Stem Cell *31*, 244–259
February 1, 2024
https://doi.org/10.1016/j.stem.2023.12.001

CellPress

CellPress
OPEN ACCESS

Cell Stem Cell

Resource

A time- and single-cell-resolved model of murine bone marrow hematopoiesis

Iwo Kucinski,[1,7] Joana Campos,[2,6,7] Melania Barile,[1,3,7] Francesco Severi,[4,5,7] Natacha Bohin,[2] Pedro N. Moreira,[4,5] Lewis Allen,[2,6] Hannah Lawson,[2,6] Myriam L.R. Haltalli,[1] Sarah J. Kinston,[1] Dónal O'Carroll,[4,5,*] Kamil R. Kranc,[2,6,*] and Berthold Göttgens[1,8,*]

[1]Wellcome–MRC Cambridge Stem Cell Institute, Department of Haematology, Jeffrey Cheah Biomedical Centre, University of Cambridge, Cambridge, UK
[2]Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London, London EC1M 6BQ, UK
[3]Centre for Translational Stem Cell Biology, Hong Kong SAR, China
[4]Centre for Regenerative Medicine, University of Edinburgh, Edinburgh EH16 4UU, UK
[5]Wellcome Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3BF, UK
[6]Institute of Cancer Research, London SM2 5NG, UK
[7]These authors contributed equally
[8]Lead contact
*Correspondence: donal.ocarroll@ed.ac.uk (D.O.), kamil.kranc@icr.ac.uk (K.R.K.), bg200@cam.ac.uk (B.G.)
https://doi.org/10.1016/j.stem.2023.12.001

SUMMARY

The paradigmatic hematopoietic tree model is increasingly recognized to be limited, as it is based on heterogeneous populations largely defined by non-homeostatic assays testing cell fate potentials. Here, we combine persistent labeling with time-series single-cell RNA sequencing to build a real-time, quantitative model of *in vivo* tissue dynamics for murine bone marrow hematopoiesis. We couple cascading single-cell expression patterns with dynamic changes in differentiation and growth speeds. The resulting explicit linkage between molecular states and cellular behavior reveals widely varying self-renewal and differentiation properties across distinct lineages. Transplanted stem cells show strong acceleration of differentiation at specific stages of erythroid and neutrophil production, illustrating how the model can quantify the impact of perturbations. Our reconstruction of dynamic behavior from snapshot measurements is akin to how a kinetoscope allows sequential images to merge into a movie. We posit that this approach is generally applicable to understanding tissue-scale dynamics at high resolution.

INTRODUCTION

A continuous flow of cells replenishes blood throughout life to maintain hematopoietic homeostasis. This flow originates from hematopoietic stem cells (HSCs) and progresses through a complex hierarchy of progenitors, collectively called hematopoietic stem and progenitor cells (HSPCs). Decades of research have revealed immunophenotypically defined HSPCs and their fate potentials, thus positioning them within the hematopoietic hierarchy and establishing the hematopoietic tree model.[1,2] Although scRNA-seq introduced high-resolution and resolved HSPC heterogeneity, scRNA-seq typically provides snapshot measurements with limited temporal information. Thus, the hematopoietic tree model, even complemented by scRNA-seq data, remains static and qualitative and does not capture the highly dynamic HSPC biology in real time.

To facilitate real-time modeling of HSPC dynamics, a previous study[3] induced a persistent fluorescent reporter within the HSC compartment and assessed label propagation into progeny by flow cytometry. However, immunophenotyping has limited resolution, and flow-cytometry-defined HSPCs are functionally heterogeneous. For instance, common myeloid progenitors (CMPs)[4,5] and lymphoid-primed multipotent progenitors (LMPPs)[6,7] are heterogeneous at functional and RNA level. Further scRNA-seq studies suggested gradual molecular transitions from HSCs toward 8 distinct lineages,[8–10] including specific stages of erythroid differentiation.[10] Nonetheless, although molecular states captured by scRNA-seq can be predictive of progenitor fate potential when assessed *in vitro*,[11–13] gaining insights into single-cell fates *in vivo* during homeostasis has remained more challenging.[14]

Lineage tracing in non-hematopoietic tissue combined with scRNA-seq has provided insights into progenitor cell differentiation to the airway epithelial lineage.[15] Nevertheless, such an approach has never been applied to a complex multilineage differentiation process, such as hematopoiesis. Furthermore, it remains unclear whether predictive tissue-scale computational models of steady-state tissue homeostasis at single-cell resolution can be constructed based on such approaches. Here, we reveal high-resolution HSPC kinetics of multilineage bone marrow (BM) hematopoiesis *in vivo*. We combined inducible HSC-labeling to track label propagation to downstream progeny during steady-state hematopoiesis with scRNA-seq at different

time points after label induction. This enabled us to reveal real-time dynamics and build quantitative cellular flow models of BM hematopoiesis. These models describe numbers of cells produced and transported across the HSPC compartment, properties that have so far only been measured for a selected few subpopulations. Notably, the ample molecular information allowed us to construct continuous models to associate gene expression changes with cell behaviors, such as increased proliferation or accelerated differentiation, thus directly connecting tissue and cellular behavior with the underpinning layer of molecular processes. Finally, we demonstrate that our dynamic HSPC model is transferable and able to predict HSPC fate outcomes based on published datasets.

RESULTS

Hoxb5-Cre^{ERT2}-Tomato reporter tracks HSC differentiation over time

To analyze HSPC dynamics, we aimed to employ a labeling approach (based on principles from Busch et al.[3]), in which an inducible HSC-specific CRE excises a STOP cassette in the *Rosa26-LoxP-STOP-LoxP-tdTomato* (*R26$^{LSL\text{-}tdTomato}$*) reporter to permanently label HSCs and their subsequent progeny. We hypothesized that *Hoxb5*, which is specifically expressed in HSCs,[16] is a suitable driver locus. To validate the specificity of *Hoxb5* expression at the protein level, first we generated *Hoxb5^{mKO2}* mice, where HOXB5 and mKO2 fluorescent reporter expression is driven by the endogenous *Hoxb5* locus (Figure S1A). mKO2 expression was selectively confined to the BM Lin$^-$Sca-1$^+$c-Kit$^+$ (LSK) HSPC compartment (Figures S1B–S1D, extended data Figure E1A, extended data figures "E" are available in Mendeley Data, see key resources table). Although high mKO2 expression was exclusive to the LSKCD48$^-$CD150$^+$ HSC fraction and enriched for this population (Figures S1B–S1D), low-level expression was also detected in LSKCD48$^-$CD150$^-$ multipotent progenitors (MPPs) (Figures S1B–S1D). At the functional level, we observed robust long-term multilineage repopulation activity of mKO2$^+$ HSCs upon serial transplantation. Notably, chimerism in the HSC compartment of primary recipients was significantly lower in the mKO2$^-$ cohort, and mKO2$^-$ HSCs failed to efficiently sustain all lineages in secondary recipients (Figures S1E and E1B–E1D). Furthermore, scRNA-seq demonstrated that mKO2$^+$ cells express canonical HSC-affiliated genes, display the highest HSC-score (Figures E2A–E2C),[17] and tightly occupy the region of the most immature stem cells on high-resolution HSPC landscape[7] (Figures E2D–E2F). Altogether, HOXB5 selectively marks HSCs with the long-term multilineage reconstitution potential and stem cell signature.

Having validated *Hoxb5* as a suitable locus, we generated *Hoxb5^{CreERT2}* mice[16] and crossed them with *R26$^{LSL\text{-}tdTomato}$* reporter[18] to establish the *Hoxb5^{CreERT2};R26$^{LSL\text{-}tdTomato}$* mice (referred to as Hoxb5-Tom, Figure 1A), which allow for inducible labeling of HSCs *in situ* by tamoxifen administration and subsequent tracking of HSC progeny over time (Figure 1B). To validate this system, we used flow cytometry to track label propagation across BM HSPC subpopulations and differentiated peripheral blood (PB) cells at indicated intervals (Figures 1B–1F, S1F, S1G, and E3). Upon tamoxifen administration, we observed specific labeling of 1.8% of HSCs, with the label gradually accumulating in downstream cell compartments over time (Figures 1D–1F). Internal controls (i.e., vehicle-treated Hoxb5-Tom mice or those lacking *Hoxb5^{CreERT2}*) showed no labeling (Figure E3A). Labeled differentiated cells were detectable in PB within 1–2 months post-treatment, with particularly fast contribution to the platelet lineage, followed by erythrocytes and myeloid cells, and T and B cells appearing later (Figures 1E, 1F, and E3B–E3D). We observed non-decreasing labeling for at least 9 months post-treatment (Figures 1D, 1E, S1F, and S1G), indicating that the label is persistent and inert.

Computational inference of population dynamics relies on a simple principle (Figure 2A): as heritable label propagates down from the label-rich upstream compartment, the speed of differentiation is proportional to label equilibration (Figure 2B, see STAR Methods). To benchmark our experimental model, we compared flow cytometry data obtained from tamoxifen-treated Hoxb5-Tom mice with previously published results of analogous label propagation obtained with Tie2-YFP mice.[3] Our data are highly consistent for both MPP/HSC and HPC-1/HSC relative abundances across the entire time range (Figure 2C), thus validating our transgenic models and unlocking our next goal—modeling of population dynamics.

A unified reference HSPC landscape with time-resolved differentiation

To capture scRNA-seq profiles of cells traversing the HSPC landscape over time (Figure 3A), we harvested BM from tamoxifen-treated mice at 9 time points ranging between 3 days (providing just enough time for Tom protein expression) and 269 days, when the label is mostly equilibrated. Next, we sorted and pooled together cells from two overlapping Lin$^-$cKit$^+$ and Lin$^-$Sca1$^+$ populations, which contain all HSPCs[9] (Figure E3E). To ensure accuracy and reproducibility, we profiled multiple independent biological replicates for each time point (36 animals in total). Although our focus was labeled Tom$^+$ cells, we also profiled Tom$^-$ cells at each time point to obtain accurate background cell density in case it changes over time. We generated a common reference landscape by integrating all data followed by clustering, UMAP embedding, and manual annotation (Figures 3B, 3C, and S2A–S2E). Clusters disjointed from the main landscape body (mostly mature cell types) and those representing technical artifacts (e.g., doublets or dying cells) were excluded (unfiltered data in Figures S2F and S2G). The refined landscape (>115,000 cells) served as the basis for our analysis. To place our data within the broader scope of hematopoiesis research and extend its interpretability, we provide multiple layers of annotation. Manual annotation[9,12] used lineage marker expression, cell-cycle phases, HSC-score (molecular signature of long-term repopulating HSCs—LT-HSCs[17]), and pseudotime (Figures 3A–3D; Table S1) to highlight the upstream cluster containing HSCs (Figure S2C) (cluster 0) and 8 terminal clusters (Figure 3C), where clear expression of definitive markers is observed. Please note that we refer to the populations as terminal within the constraints of our stem and progenitor landscape, but most of them are not mature cells, and cells differentiate beyond our landscape. To add functional information, we mapped external scRNA-seq datasets using our Cellproject package. First, we overlaid canonical immunophenotypic subpopulations with our scRNA-seq landscape (Figures 3D, 3E, S3A, and S3B) (data from Nestorowa et al.[7]) comprising highly

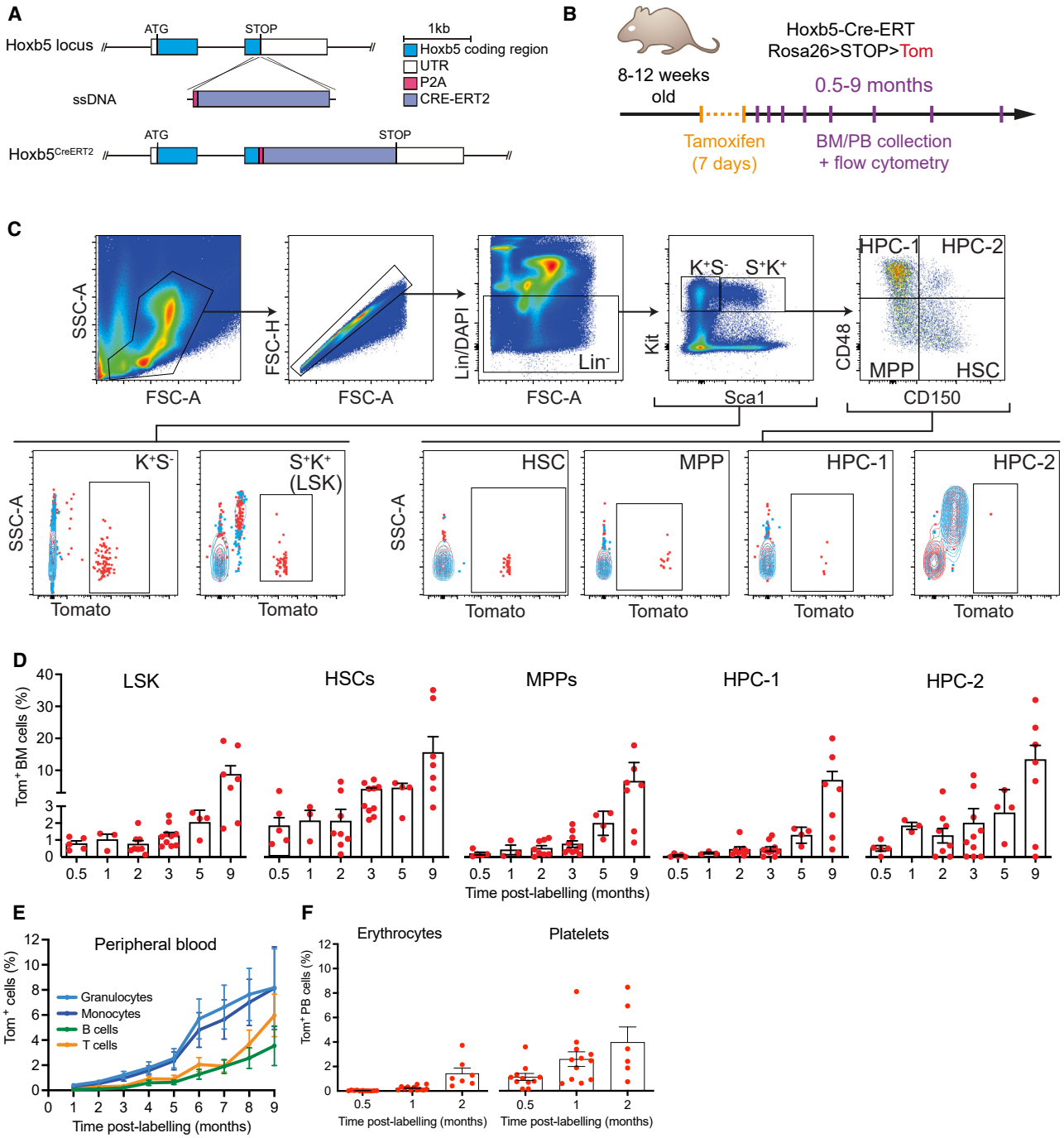


Figure 1. Hoxb5-Tom persistent labeling enables time-resolved tracking of HSCs and their progeny

(A) Targeting strategy to generate the *Hoxb5*^{CreERT2} allele.

(B) Hoxb5-Tom mice were treated with tamoxifen and label frequency was analyzed in BM and PB cells within 0.5–9 months post-treatment.

(C) Representative flow cytometry gates used for isolation of HSPC subpopulations and Tom$^+$ cells from mouse BM. Tom labeling (red) is shown in each population compared with control cells (blue). FACS plots correspond to mouse analyzed 3 months after label induction.

(D) Percentage of Tom + cells in the BM HSPC subpopulations at 0.5 (n = 5), 1 (n = 3), 2 (n = 8), 3 (n = 10), 5 (n = 4), and 9 (n = 7) months after label induction. Dots represent individual mice, bars indicate mean ± SEM.

(E and F) Percentage of Tom + cells in PB of lymphoid/myeloid cell compartments (E) and erythrocytes/platelets (F). Data represent mean ± SEM (n = 4–32 animals). LSK, Lin$^-$Sca1$^+$cKit$^+$; HSCs, LSKCD150$^+$CD48$^-$; MPP, LSKCD150$^-$CD48$^-$; HPC-1, LSKCD150$^-$CD48$^+$; HPC-2, LSKCD150$^+$CD48$^+$ cells.

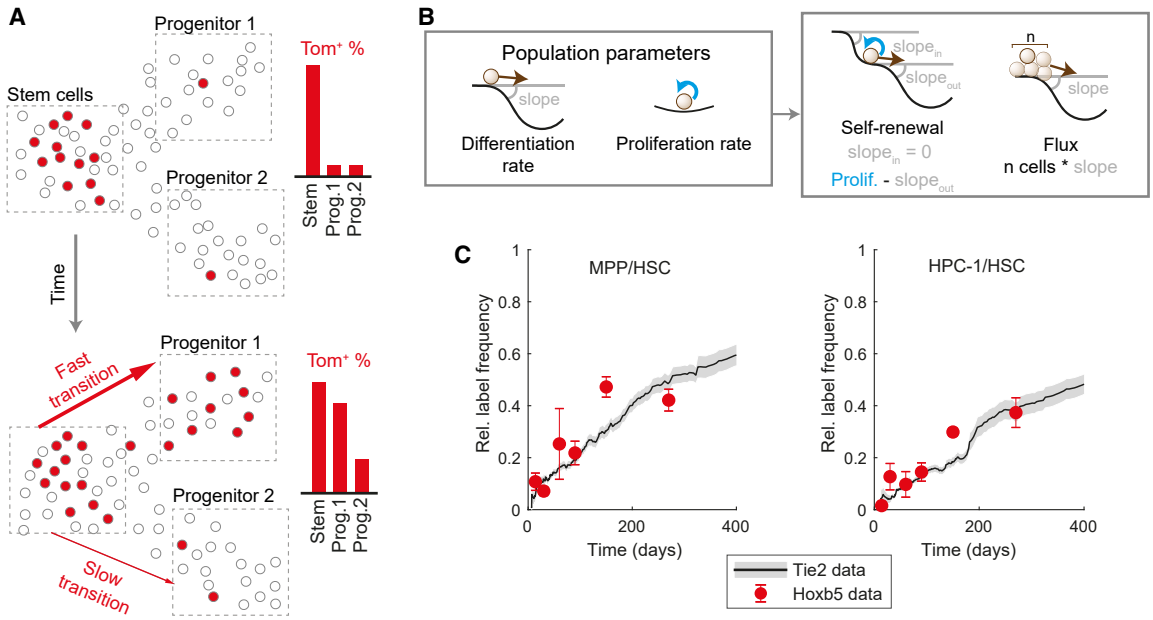


Figure 2. Label propagation contains information about population dynamics

(A) Diagram portraying the concept of inferring population dynamics from heritable label propagation. The rate of label accumulation in the downstream compartments is proportional to the differentiation rate between the compartments.

(B) Diagrams providing analogy between the shape of the Waddington landscape and the key population parameters estimated: differentiation rate is akin to the slope of the landscape; self-renewal (and related residence time or half-life) depend on the input, output, and proliferation; flux the number of cells multiplied by the slope.

(C) Comparison of Tie2-YFP and Hoxb5-Tom label progression displayed as relative labeling frequency between MPP or HPC-1 and HSC compartments. Red dots, Hoxb5-Tom data points (see Figure 3) with SEM error bars; black line and gray shades, rolling average and rolling SEM for matching Tie2-YFP data.[19]

purified LT-HSCs, MPPs 1 and 3, ST-HSCs, granulocyte-monocyte progenitors (GMPs), LMPPs, and megakaryocyte-erythroid progenitors (MEPs). Second, we highlighted cell states associated with specific cell fate outcomes based on *in vitro* lineage tracing experiments[12] (Figures 3F and S3C). Importantly, the *in vitro* cell potency is broadly aligned with the manual cluster annotation. Finally, we included information about the active/inactive HSC status under proliferative challenge based on lineage tracing data from Bowling et al.[20] (Figure 3G). Together, these annotations place cell clusters into a functional framework facilitating interpretation of the population dynamics models discussed below.

The HSPC landscape split by time point shows clear propagation of labeled cells (Figure 3H), a full quantification of labeled/unlabeled cell ratios for all time points is provided in Figure S4A and follows the expected behavior[3] (Figure 2A). Certain clusters (e.g., 8 and 7) very quickly accumulate labeled cells, others are slower (clusters 11 or 10) and some very slow (clusters 13 or 14) (Figures 3H and S4A). Eventually, the label largely equilibrated, as compared with the Tom$^-$ population (Figures S4A and S4B). Importantly, scRNA-seq clustering resolves heterogeneity within cell populations defined by conventional flow cytometry gates (Figures S3A and S3B)[4,7,21] and is predictive of cell fate.[12] To provide a quantitative description of population dynamics, we employed two types of models: discrete and continuous, each built for specific purposes. The former captures dynamics across the entire compartment and intuitively combines hierarchical tree models of hematopoiesis with a quantitative view based on more precisely defined cell types. It also serves as a necessary reference for the latter, a more advanced continuous modeling approach, which focuses on specific trajectories, but provides cellular flux parameter estimates for each single cell and thus directly connects single-cell transcriptomic profiles with tissue-scale cellular behavior.

Discrete model reveals HSPCs with lineage-specific patterns of self-renewal and differentiation

To capture the flow of cells through the HSPC compartment in real time, we utilized the concepts from previous label propagation studies[3,22] to build a discrete model consisting of multiple, interconnected cell clusters (Figures 4A–4C). We explain two variables changing over time: number of labeled cells (Tom$^+$ cells, Figures 4D and E4; Table S2) and size (Tom$^-$ cells, Figure E5; Table S2) for each cluster (labeling frequency in Figure E6). Each cluster has two basic properties: *net proliferation* (number of divisions reduced by the number of cells lost, e.g., by cell death) and *differentiation rates* (number of ingoing and outgoing cells between clusters per day, scaled to a single cell). Thus, our model simultaneously estimates (net) proliferation balancing it with the influx, efflux, and time-dependent cluster size. Importantly a common set of parameters fits both labeled and unlabeled cells (except cluster 0, see the next section) indicating similar dynamics. Additionally, we introduce two derived parameters useful for interpreting cell behavior (Figure 2B). *Residence time*, which corresponds to a half-time of one cell in a cluster, is the time required for the cluster to shrink by 63% (to 1/e of original size, where e is the Euler's number) in absence

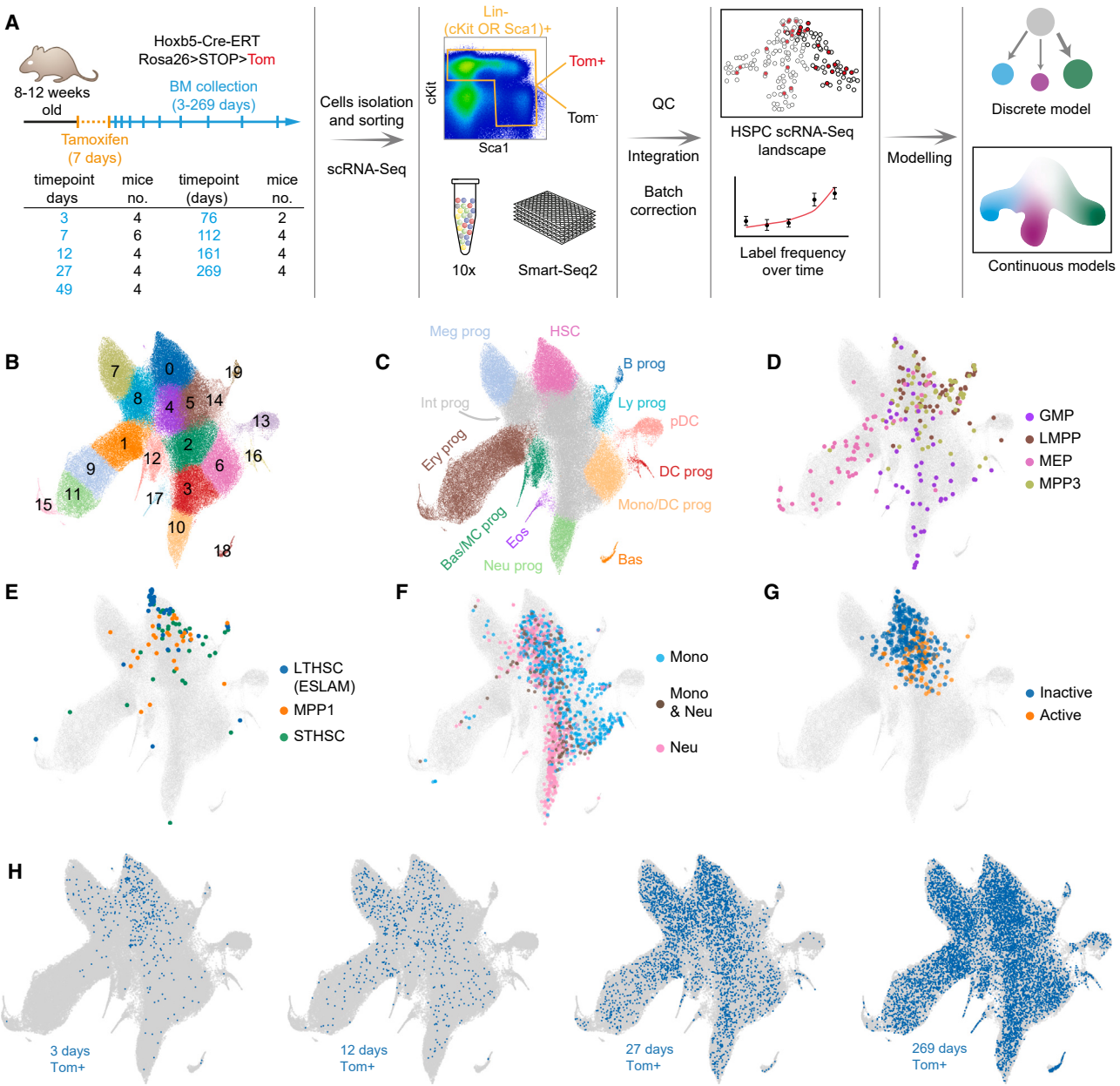


Figure 3. Time-resolved reference HSPC landscape at single-cell level

(A) Hoxb5-Tom mice were administered with tamoxifen. Table indicates time points and mouse numbers used for Tom$^+$ HSPC scRNA-seq. Two mice per time point were used for the Tom$^-$ scRNA-seq.

(B) UMAP projection of the integrated HSPC scRNA-seq landscape (all Tom$^+$ and Tom$^-$ cells combined) with color-coded clusters. Outlier or aberrant clusters were removed for clarity (see Figures S2F and S2G).

(C) Manual annotation of the landscape in (B). Most differentiated clusters with clearly defined lineage markers are color-coded, intermediate undifferentiated states are shown in gray (Int prog), cluster containing HSCs is shown in pink.

(D and E) Projection from (B) in gray, with embedded and color-coded immunophenotypic populations from Nestorowa et al. data.[7] Up to 60 randomly selected cells in each category are plotted. All cells are plotted in Figure S3A.

(F) Projection from (B) in gray, with embedded and color-coded cKit$^+$ progenitors, based on their output in lineage tracing *in vitro* cultures. Data from Weinreb et al.[12]

(G) Projection from (B) in gray, with embedded and color-coded HSCs with no detected cellular output (inactive) or contributing to hematopoiesis (active) following 5-FU challenge in mice (data from Bowling et al.[20]).

(H) Projection from (B), with Hoxb5-Tom$^+$ cells at indicated time points shown in blue. Nestorowa et al.[7] population definitions: LT-HSC, lin-cKit$^+$Sca1$^+$CD34$^-$Flt3$^-$; MPP1, lin-cKit$^+$Sca1$^+$Flt3$^-$CD34$^+$CD150$^+$CD48$^-$; ST-HSC, Lin$^-$cKit$^+$Sca1$^+$Flt3$^-$CD34$^+$CD150$^-$CD48$^-$; GMP, Lin$^-$cKit$^+$Sca1$^+$CD16/32$^+$CD34$^+$; LMPP, Lin$^-$cKit$^+$Sca1$^+$Flt3$^+$CD34$^+$; MEP, Lin$^-$cKit$^+$Sca1$^+$CD16/32$^-$CD34$^-$; MPP3, Lin$^-$cKit$^+$Flt3$^-$CD34$^+$CD150$^-$CD48$^+$; CMP, Lin$^-$cKit$^+$Sca1$^+$CD16/32$^-$CD34$^+$ cells. Prog, progenitors; B, B cell; Bas, basophils; Bas/MC, basophil and mast cell; DC, dendritic cell progenitors; Eos, eosinophils; Ery, erythroid; Int, intermediate; Ly, lymphoid; Meg, megakaryocyte; Mono/DC, monocyte and dendritic cell; Neu, neutrophil; pDC, plasmacytoid dendritic cells.

CellPress
OPEN ACCESS

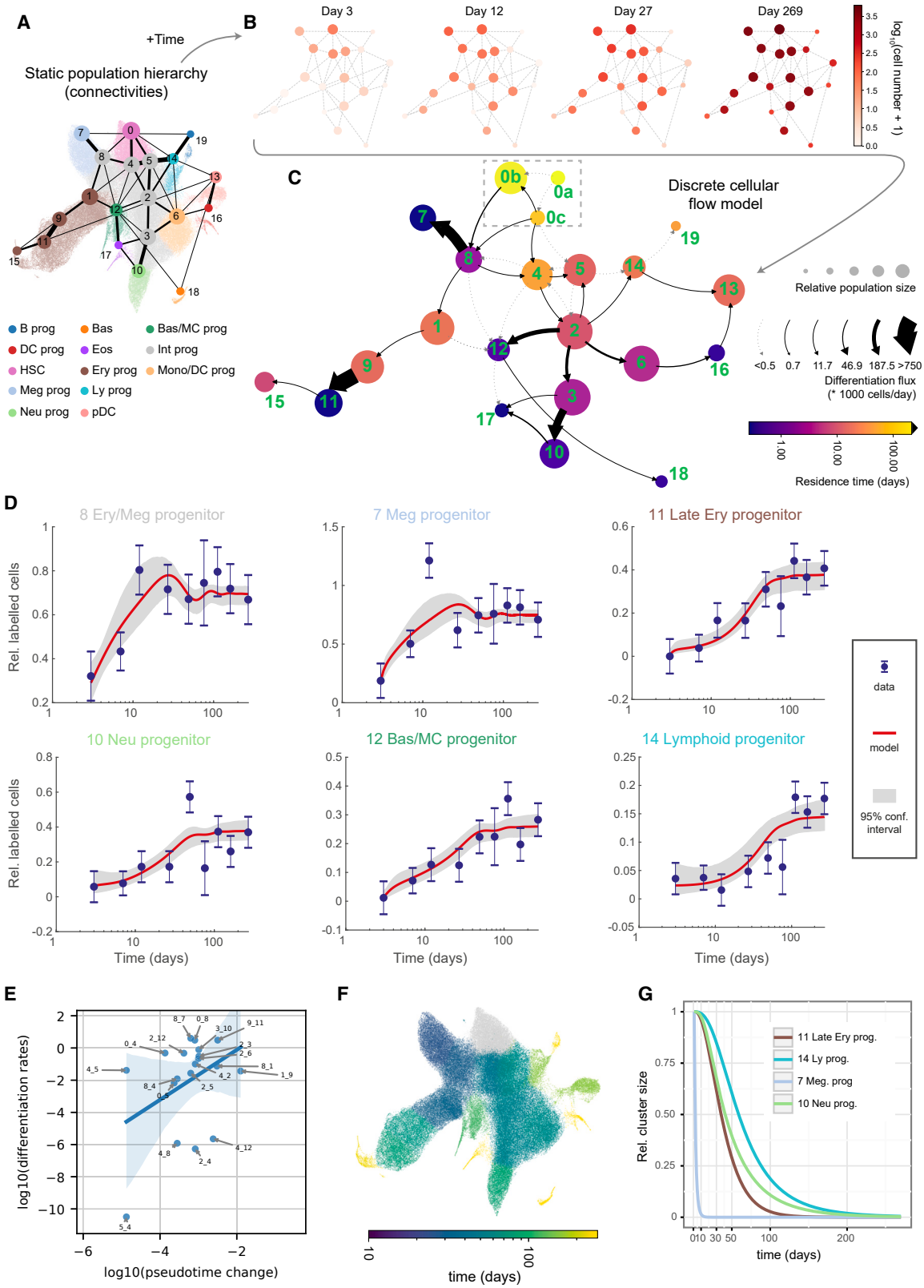


Figure 4. Quantitative discrete model of the HSPCs highlights progenitor-specific self-renewal and differentiation properties

(A) Annotated UMAP projection overlaid with PAGA graph abstraction view of the HSPC landscape. The graph shows putative transitions between clusters (related to Figure 3B).

(legend continued on next page)

of any incoming cells. Residence time is defined as the inverse of ([death + differentiation] − proliferation) and thus residence time increases as proliferation rate rises, and death/differentiation rates decrease and vice versa. Finally, *flux* depicts the total number of cells transported between clusters in a unit of time (i.e., differentiation rate multiplied by cluster size). We limited the number of differentiation parameters by assuming that cells travel only between adjacent clusters (i.e., with highest PAGA[23] connectivities—Figure 4A). Although PAGA is a robust method with relatively few assumptions, there is currently no consensus in trajectory inference methodology. Thus, we also provide tools to explore alternative topologies (see STAR Methods) and apply a cluster-independent, continuous model (see later).

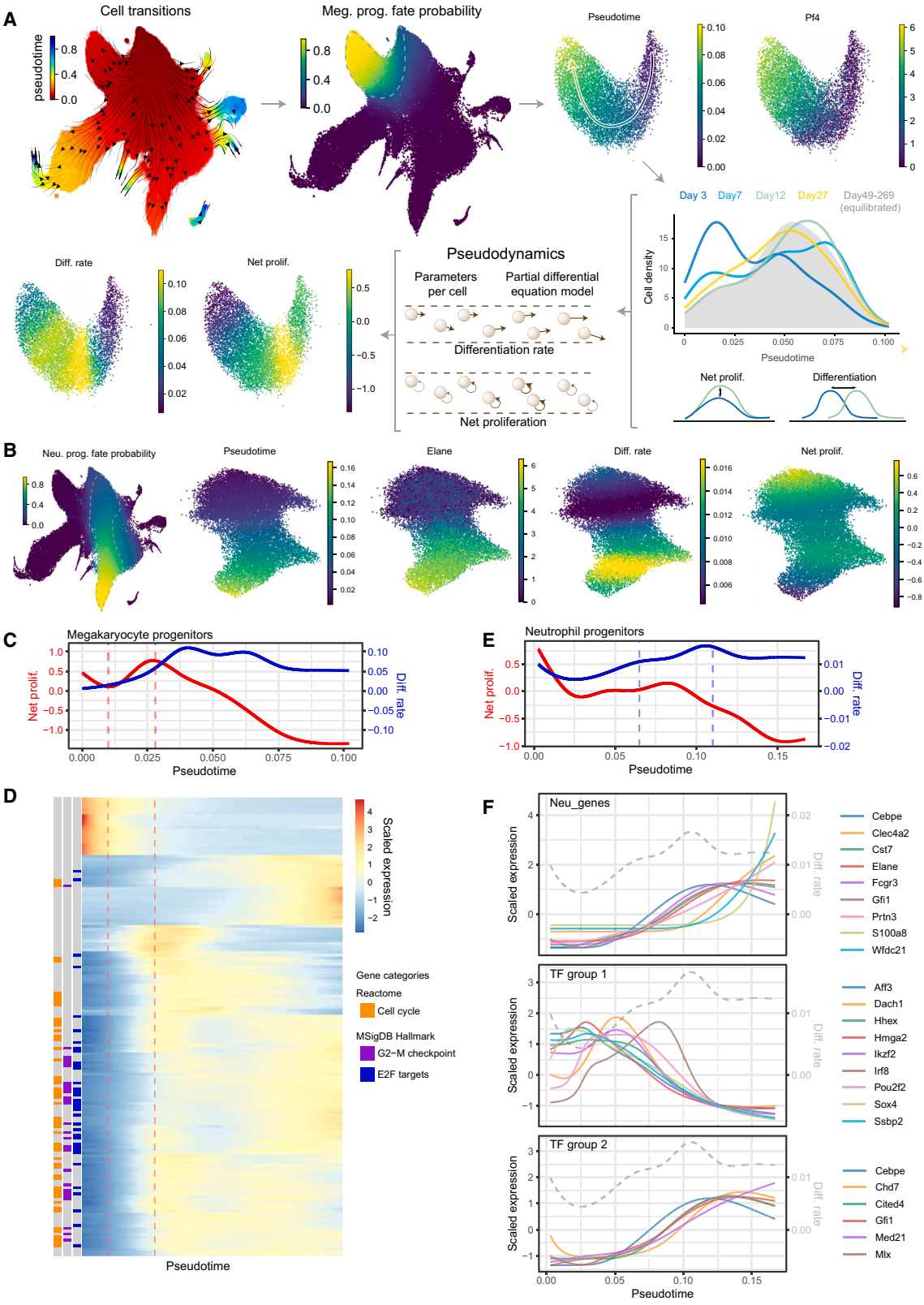
Of note, we observed changes in relative cluster size over time (i.e., the background unlabeled cells), in particular a quick increase in relative abundance (compared with cluster 0) of clusters 7 and 8 (>50% in <20 days) and a coordinated relative decrease in other major clusters (Figures E5 and S5A–S5C). Cluster 0 size also modestly increases size in the same time frame. Previous tamoxifen-based label propagation studies also observed a quick rise in ST-HSC, MPP2, and MPP3 total numbers (Figure S5D), but no explanations were provided.[19] It had previously been suggested that application of tamoxifen interferes with JAK-STAT signaling.[24] Consistent with recovery from cell depletion caused by tamoxifen interference with JAK/STAT, this pathway was most active in the depleted clusters 7, 8 in addition to cluster 0 (Figure S6A). To assess how recovery from short-term cell depletion may influence model parameters, we compared our main model with a bi-phasic fit, which permits a switch in differentiation/proliferation rates between the recovery and homeostasis phases, albeit at some cost of increased parameter uncertainty (Figures S6B and S6C). We observed changes in 14 out of 58 rates between the two phases (Figures S6D and S6E; Table S3). Of note, all bar one of the homeostasis rates in the bi-phasic model are essentially the same as the rates in the main model. We thus explain and account for a previously overlooked side-effect of using tamoxifen for label induction.

We formulated our main model into a graph in Figures 4C and S6F, where node sizes are proportional to the average cluster size, node color indicates residence time (or net proliferation in Figure S6F) and arrows indicate cell flux (differentiation rate in Figure S6F). Please note that some transitions occur infrequently (transition rates and their confidence intervals are provided in Table S3), and we cannot exclude that some may be redundant (for the discussion on the minimal model, please see the methods section "Model selection"). Interestingly, differentiation rates poorly correlate with similarities between gene expression states (Figures 4E and S6G), indicating that discovery of real-time dynamics requires temporal information. Moreover, the compartment-wide view clearly shows lineage-specific dynamics (Figure 4C). Megakaryocyte progenitors emerge through a rapid transition via the fast-proliferating cluster 8, which also generates erythroid cells, albeit more slowly (cluster 1). Substantial erythroid output is achieved via sequential cell states with considerable self-renewal (clusters 1 and 9) and proliferation (cluster 9), followed by fast differentiation between clusters 9 and 11. Furthermore, myeloid progenitors transition from cluster 0 either into cluster 4 or via a shared route with the erythroid and megakaryocytic progenitors into cluster 8, with gradually increasing differentiation rates from cluster 2 onward. The myeloid branch therefore employs additional progenitor populations analogously to the erythroid trajectory, albeit with lower proliferation rates (Figure S6F).

The lymphoid trajectory is altogether different showing exclusively slow transitions via clusters 5 and 2 into cluster 14 (which overlaps mostly with a subset of MPP4 cells). Cluster 5, compared with the more myeloid-biased cluster 4, proliferates and differentiates more slowly, while expressing higher levels of key lymphoid factors, including Flt3, Satb1, Pou2f2 (and to some extent the monocytic factor Irf8, discussed later) (Figure E7A). The lymphoid program therefore displays restricted proliferation and differentiation rates already from its immature stages. Plasmacytoid dendritic cell (cluster 13, pDCs) differentiation through the lymphoid cluster 14 and myeloid clusters 6 and 16 is similarly slow. The emergence of mast cell, basophil, and eosinophil progenitors in the adult BM is unclear.[25,26] Our results are consistent with a model whereby basophil and mast cell progenitors (cluster 12) are continuously generated and originate at least by a transition from the early myeloid cluster 2 but may also have some contributions from other clusters (dashed lines). Furthermore, despite limited cell numbers, we observed some label accumulation in eosinophil progenitors (cluster 17), most likely originating from neutrophil progenitors (cluster 10).

Interestingly, residence time (self-renewal) varies widely across the HSPC landscape, with lineage-specific patterns (Figure 4C; Table S3). As expected, cluster 0 contains the only perfectly self-sustaining population; intermediate populations show a range of residence times, from just 2.5 days for erythroid/megakaryocytic progenitor (cluster 8), 11 days for monocyte/granulocyte progenitors (cluster 2) and up to 53 days for the medial cluster 4. The latter falls close to the residence time previously estimated for MPPs (70 days)[3] and highlights that progenitors can also show considerable self-renewal. Importantly, cells in clusters 8, 2, and 4 fall within the immunophenotypic CMP and MPP definitions (Figures 3D, 3E, S3A, and S3B), illustrating how historically used

(B) Graph from (A) color-coded by the absolute number of labeled cells observed in each cluster. four out of 9 time points are shown for clarity.
(C) Graph abstraction view of the discrete cellular flow model. Size of the nodes is proportional to square roots of relative cluster size, node color is proportional to the residence time (log-scale), arrows indicate differentiation directions, arrow stem thickness is proportional to cell flux. Note: cluster 0a is fully self-renewing and thus exhibits infinite residence time.
(D) Best discrete model fit (with 95% confidence intervals) for Tom^+ cell number in chosen clusters relative to cluster 0. Error bars indicate pooled SEM.
(E) Scatter plot showing relation of pseudotime distance to differentiation rates, each point corresponding to a transition between clusters. Only transitions among clusters 0–12 and differentiation rates greater than 10^{-12} are shown. Please note that in the case of the transitions between clusters 4 and 8 two differentiation rates are plotted (each direction). Blue line indicates linear model fit with shaded 95% confidence interval.
(F) UMAP projection of the HSPC landscape, with cells color-coded by simulated time required for 1 cell to accumulate in the corresponding cluster starting from cluster 0. Please mind that the color is logarithm scaled.
(G) Simulated relative cluster size of chosen clusters following complete ablation of cluster 0.



(legend on next page)

flow cytometry gates capture populations with vastly different dynamics. We also note that among some intermediate clusters, our model permits a degree of forward and backward differentiation suggesting that some states may exist in an equilibrium, with each cluster having distinct differentiation properties. Thus, diverse hematopoietic progenitors exhibit widely different, lineage-specific dynamics consistent with distinct mechanisms maintaining cell output.

Composition of the top HSPC compartment changes over time

Based on immunophenotype annotations (Figures 3C–3E), the top cluster 0 contains virtually all LT-HSC and a large subset of ST-HSC and MPP1 cells. The overall cluster size increases over time (Figures S5B and S5C), reminiscent of previous reports noting the expansion of ST-HSCs and MPP3s as mice age (Figure S5D).[19] Of note, the Hoxb5-Tom-labeled cells within cluster 0 grow almost exponentially (Figure S5A), which mirrors the previously reported behavior of Tie2-YFP labeled LT-HSCs[19] and is consistent with the observation of dramatic expansion of Hoxb5-, Tie2-, or Fgd5-labeled cells in aging animals.[27] This suggests that the Hoxb5 and Tie2 systems mark, in addition to the canonically quiescent LT-HSCs, a subset of immature cells with high self-renewal or proliferation capacity.

To investigate the apparent heterogeneity within cluster 0, we tested multiple models and put forward a potential explanation, which assumes a logistic growth for cluster 0 and three subclusters within in it: a top, perfectly self-renewing cluster 0a, the megakaryocyte & myeloid-biased cluster 0b, and the multipotent cluster 0c (Figure 4C, dashed box). We constrained cluster 0a size and differentiation rate to match previously reported LT-HSC numbers but left clusters 0b and 0c sizes unconstrained. We defined the tip cluster by finely subclustering cluster 0 and picking as cluster 0a the subcluster with the highest HSC-score (subcluster 8, Figures S5E and S5F). Reassuringly, this cluster size is compatible with our model prediction, is enriched for HSC markers *Procr* and *Ly6a*, and, most importantly, has a non-growing labeling frequency, as one would expect from the candidate tip cluster (Figures S5G–S5I). Cluster 0c remains stable over time but it proliferates quickly and feeds both downstream progenitors and cluster 0b, which in turn grows over time (Figures S5B and S5C). Hence, the flux between clusters 0b and 8 increases with mouse age. This is in line with the increased myeloid output[28,29] and relative proportion of megakaryocyte-biased and myeloid-biased HSCs in aged animals.[30] Of note, cluster 0b shows high self-renewal (residence time of 180 days), consistent with high repopulation potential of lineage-biased HSCs.[30] Altogether, our discrete model faithfully recapitulates cell flux through the HSPC compartment and provides a possible explanation of aging-associated changes in HSC behavior.

Continuous model of hematopoiesis connects dynamics of gene expression with cell behavior

Although the discrete model provides compartment-wide dynamics, a complementary model is required to associate gene expression changes at the single-cell level with cell behavior, such as increased proliferation or accelerated differentiation. For this purpose, we employed a continuous model based on the Pseudodynamics framework.[31] For tractability, we considered one lineage at a time, based on cells with highest fate probabilities toward each lineage[32,33](Figures 5A, 5B, E8, and E9). The continuous model assigns differentiation and net proliferation rates to each cell (Figure 5A) by solving partial differential equations describing cell densities along pseudotime over real time. Hence, model parameters and gene expression share a common pseudotime (and real-time) axis, enabling direct comparison. Of particular interest are states (pseudotime ranges) with changes in proliferation or differentiation rates. An increase in proliferation rates indicates an expansion stage, whereas a rise in differentiation rates marks a potentially irreversible molecular transition.

We set out to analyze gene expression dynamics occurring at such changes in cell behavior. For instance, correlating the first derivative of the differentiation rates and gene expression highlights complex matching patterns and shortlists potential regulators driving cell differentiation in an unbiased manner (Figure E11, extended data Table E1, extended data tables "E" available in Mendeley Data, see key resources table). A more targeted approach tests for differential expression around specific stages of differentiation (matching changes in cell behavior). For brevity, we showcase the megakaryocyte and neutrophil trajectories (Figures 5, E9, 10) but also provide analogous analyses for the erythroid and monocytic/dendritic lineages (Figure E9; Tables S4, S5, extended data Table E2). As shown in Figure 5A, megakaryocyte progenitors display characteristic changes in growth and differentiation rates. Cells rapidly increase their net proliferation early on, ahead of the peak in differentiation and around the stage where *Pf4* (megakaryocyte marker) mRNA becomes detectable. In this growth phase, we identified 170

Figure 5. Continuous models capture single-cell growth and differentiation rates alongside their molecular state

(A) Diagram of megakaryocyte trajectory analysis. Following the arrows: putative cell transitions (pseudotime kernel) were used to estimate cell fate, from which trajectory was isolated (dashed line). Along the pseudotime cell densities were computed for each time point (color-coded lines) and analyzed using the pseudodynamics framework providing differentiation and net proliferation rate estimates for each cell.

(B) (Left) UMAP projection of the HSPC landscape color-coded by cell fate probability of neutrophil lineage (estimated with pseudotime kernel, see A). Panels on the right show UMAP projections of isolated neutrophil trajectory color-coded by indicated parameters or gene expression.

(C) Pseudodynamics fitted net proliferation parameter (red) and differentiation rate parameters (blue) along pseudotime for megakaryocyte trajectory. Vertical lines indicate the region of interest with increasing proliferation.

(D) Heatmap of genes differentially expressed around the region of interest shown in (C). Left columns indicate genes belonging to enriched categories: E2F target (FDR $<10^{-38}$), G2-M checkpoint (FDR $<10^{-24}$), and cell cycle (FDR $<10^{-38}$).

(E) Pseudodynamics fitted net proliferation (red) and differentiation rate (blue) parameters along pseudotime for neutrophil trajectory. Vertical lines indicate the region of interest with increasing differentiation.

(F) Fitted gene expression values along pseudotime for neutrophil markers and two TF groups shown in (full analysis in Figure E10). Gray, dashed line indicated differentiation rates shown in (E). Gene expression was scaled around the mean.

dynamically expressed genes with distinct patterns along pseudotime (Figures 5C and 5D, similar analysis of the differentiation phase is showed in Figures E9C and E9D). These genes are strongly enriched for cell growth and proliferation genes with almost all of them showing an upward trend in the relevant pseudotime range. This serves as a proof of principle, as the model based solely on total cell numbers, predicts the growth stage matching the respective gene signature.

While following the neutrophil differentiation kinetics (Figures 5B and 5E), we found gradually increasing differentiation rates accompanied by a complex pattern of gene expression. Indeed, we observed two phases of neutrophil-affiliated gene expression (Figure 5F), with *Cebpe*, *Cst7*, *Elane*, *Fcgr3*, and *Gfi1* appearing almost simultaneously at the onset of differentiation, while *Clec4a2*, *Wfdc21*, and *S100a8* increasing at different intervals later. To gain insight into potential mechanisms regulating the process, we scrutinized transcription factors with dynamic expression along the trajectory (Figure E10A) and classified them into 4 groups based on expression patterns. Group 2 (Figure 5F) largely mirrored the expression of early neutrophil markers described above and reassuringly contained *Gfi1*, a key determinant of the neutrophil fate, which indeed suppresses Irf8 expression,[34] a member of the downregulated group 1 TFs. Group 3 (Figure E10B) contained factors with the highest expression in the most immature HSPCs (e.g., *Gata2*, *Hlf*, and *Meis1*) and showed early and nearly synchronous decay in expression, suggesting involvement in self-renewal. Finally, Group 1 (Figure 5F) TFs exhibit unique patterns of expression with peaks at different stages, all of which ultimately decaying as late neutrophil markers appear. These contain multiple TFs associated with specific lineages such as: *Irf8* (Monocyte/DC fate[34]), *Aff3* (lymphoid/B cells[35]), *Dach1* (myeloid[36]), *Hmga2* (myeloid, erythroid, megakaryocytic,[37] *Pou2f2* (lymphoid/B cells[38]) or are important for HSPC self-renewal, including *Ikzf2*[39] or *Ssbp2*.[40] Thus, our analysis indicates that progenitors exhibit transient expression of major lineage determinants at specific differentiation stages on their way to becoming neutrophils (see *Gfi1*, *Flt3*, *Irf8* in Figures E10D and E10E). Early accumulation of these factors is correlated with increased differentiation rate, but eventually, a single program takes over and accelerates the differentiation even further. Thus, the continuous model unlocks access to full single-cell transcriptome data and thus enables integrated analysis of cellular and molecular dynamics, revealing mechanistic insights into cell behavior during differentiation.

HSPC models simulate cell journeys in real-time consistent with basic properties of hematopoiesis

Mathematical models combined with our new datasets offer unique prediction capabilities allowing us to unravel fundamental facets of hematopoiesis. Specifically, we focused on computing cell journeys in real time and consequences of cluster ablation. First, we estimate the 'average journey times' with the discrete model. We placed a single cell in cluster 0 and computed the average time required to accumulate one cell for each target cluster. The required time depends on the specific influx/efflux and proliferation rates, including the loss of cells out of the terminal populations (via differentiation/death). Highly transient populations can therefore take longer to be populated stably. As shown in Figures 4F and S6H, average journey time widely varies between terminals states of different lineages (Table S3). For instance, accumulating a cell in Meg progenitors (cluster 7) requires 27 days, neutrophil progenitors (cluster 10) or late erythroid progenitors (cluster 11) >80 days and finally producing pDCs takes about 150 days. Second, we predict what would happen if, under normal conditions, the self-renewing cluster 0 was ablated. As expected, without cluster 0 input, downstream cluster sizes would gradually decline over time (Figure 4G), due to limited self-renewal of intermediate progenitors. As we described above, progenitor self-renewal is lineage specific, hence corresponding clusters wane at different rates, with megakaryocyte progenitors depleted to 50% after 2–3 days, whereas lymphoid progenitors are maintained for >50 days. Of note, the substantial effect of the depletion in some compartments is due to the fact that we are simulating ablation of all cells in cluster 0, which includes progenitors immediately downstream of HSCs. For comparison, we also simulated the effect of the depletion of just cluster 0a and ascertained that the effect on the downstream populations is barely noticeable (Figure E7B).

Predictions revealed by our model agree with the order of lineage emergence inferred from transplantation[12,30,41–43] or cell culture[12,41] experiments. The time frame of the process is expectedly much longer but is compatible with previous studies of HSPC dynamics *in vivo*.[3] Our approach is therefore anchored firmly in the long tradition of hematopoiesis research and opens the opportunity to serve as a predictive framework for *in vivo* experiments.

Integrative model is predictive and resolves the effects of transplantation on HSPC dynamics

To demonstrate the predictive capabilities of our models, we utilized data from an independent study (Upadhaya et al.[44]). In this setting, HSCs and their descendants were labeled using the *Pdzk1ip1-CreER*;*tdTomato* system (analogous to Hoxb5-Tom but using a different HSC-specific driver) and analyzed after 3, 7, and 14 days. Upadhaya et al.[44] profiled cells by scRNA-seq; thus, we were able to integrate them into our HSPC landscape (Table S6). As the limited number of cells and replicates was insufficient to build a standalone model, we used the Hoxb5 model parameters to predict expected cell numbers using the day 3 time point as initial condition and compared our predictions with the observed data. As shown in Figures S7A and E12, both the discrete model and continuous models faithfully predict the evolution of the system over time for most of the large clusters and trajectories. Curiously, our model indicates faster differentiation toward megakaryocytes (see clusters 7 and 8) at the expense of erythroid (clusters 9 and 1). We noted that Upadhaya et al.[44] used a milder tamoxifen treatment than our study, hence consulted the Hoxb5 bi-phasic model (Figure S6E) for potential explanation. Reassuringly, the bi-phasic parameters show that shortly after our tamoxifen treatment megakaryocytic differentiation occurs faster while erythroid slower, thus suggesting that the discrepancy is associated with the difference in tamoxifen dosage. Thus, our model, with some uncertainty, is able to quantitatively predict dynamics of adult *in vivo* hematopoiesis. Furthermore, our approach paves the way for future studies, which, avoiding the transient tamoxifen effect, will provide even more accurate models.

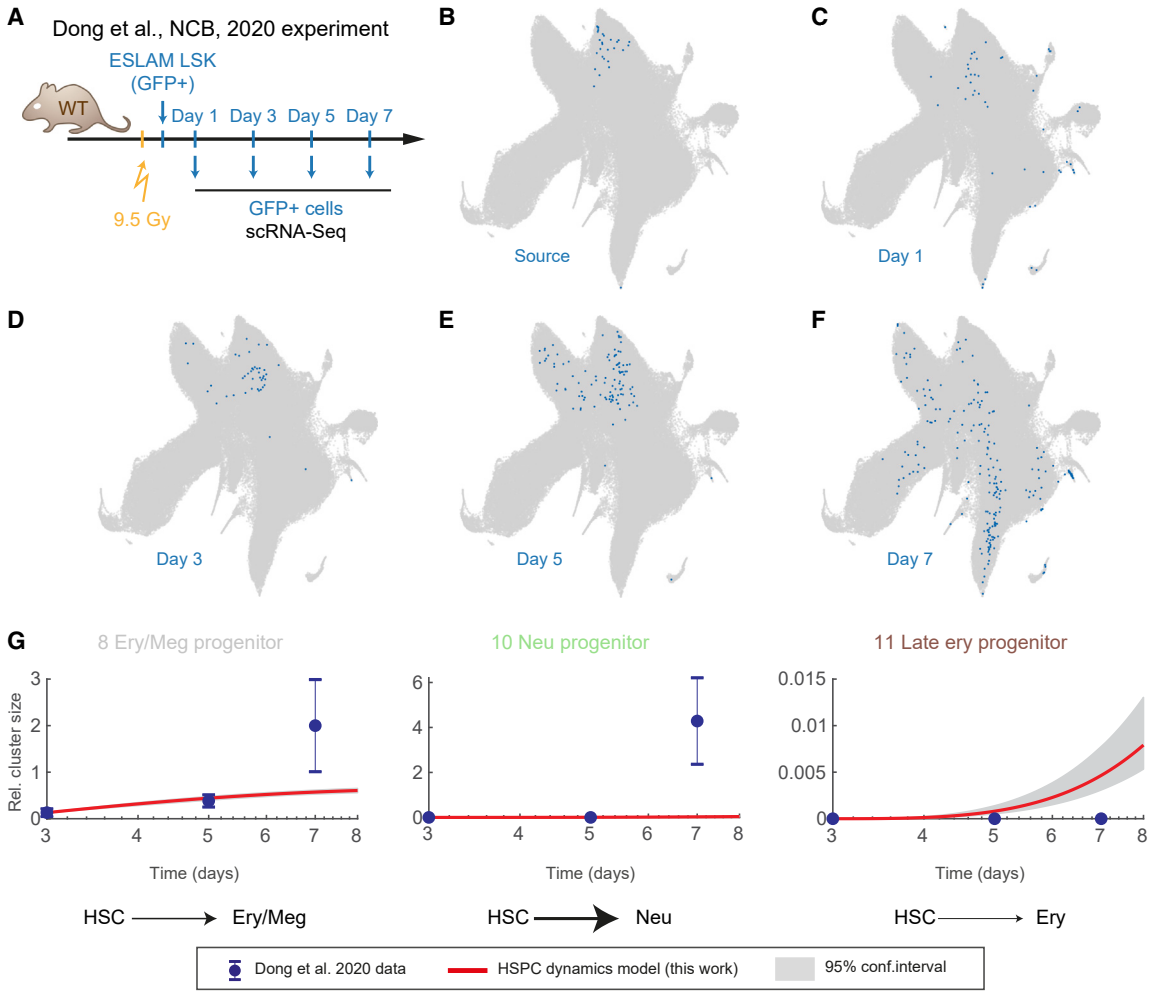


Figure 6. Growth and differentiation rates of HSPCs adapt to cellular stress conditions

(A) Diagram of the experiment performed by Dong et al.,[45] with HSC transplanted into an irradiated animal and followed over time with scRNA-seq.

(B–F) UMAP projections of the HSPC landscape (gray) with embedded cells from Dong et al.[45] in blue.

(G) Relative cluster size, points indicate observed data from Dong et al.[45] Red line indicates our discrete model prediction (shaded area: 95% confidence interval) starting from the day 3 time point. Error bars indicate propagated SEM.

We next employed the same approach to predict multi-lineage differentiation trajectories *in vitro* (Figures E13) using previously published data.[12] We found that almost all clusters and trajectories accumulate differentiating cells much faster *in vitro* than *in vivo*, though interestingly megakaryocytic differentiation occurs at roughly the same speed as *in vivo*.

We analyzed a previous study (Dong et al.[45]), which used scRNA-seq to track the progeny of highly purified HSCs in transplanted animals over time (Figure 6A). After integrating the scRNA-seq profiles into our reference landscape (Figures 6B–6F), we derived cell frequencies per cluster at day 3 and used the discrete model to predict the cell abundance expected under non-transplantation conditions (Figures 6G and S7B). Although some general features match normal hematopoiesis, for instance, megakaryocyte progenitors being the first emerging lineage, cells under transplantation conditions differentiate much faster in most directions, particularly toward the neutrophil fate (Figure 6G, cluster 10). The erythroid lineage behaves differently, whereas early megakaryocyte and erythrocyte differentiation is accelerated upon transplantation (Figure 6G, cluster 8), late erythroid progenitor cell emergence is delayed, compared with the steady-state counterparts (Figure 6G, cluster 11). To go beyond qualitative interpretation, we performed combinatorial model re-fit of the transplantation data to pinpoint the changes in differentiation rates and proliferation rates in each cluster/transition most likely to be responsible for altered transplantation landscape dynamics (Figure E14A). This procedure highlighted stage and lineage-specific effects. For instance, the erythroid lineage differentiates around 10 times faster between clusters 1 and 9, whereas myeloid progenitor cluster 2 exhibits 2-fold higher net proliferation and 7-fold faster differentiation toward neutrophil progenitors and 3-fold higher toward monocyte/DC progenitor (Figure E14B). In conclusion, we demonstrated that our model can be easily applied to other datasets and provide quantitative predictions and interpretation, which would not be available from static measurements alone.

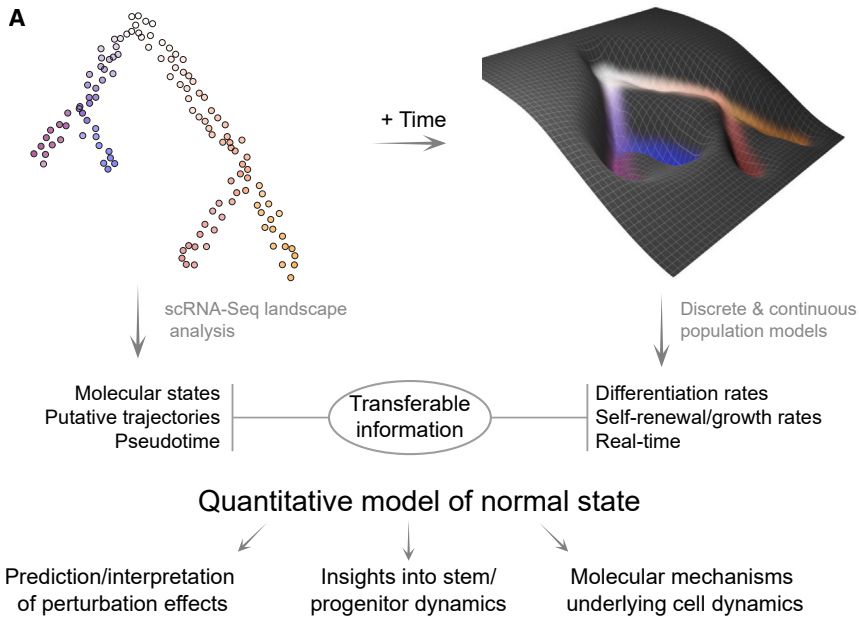


Figure 7. The quantitative model of HSPC dynamics in the mouse BM

Diagram highlighting the transferable information and the model utility.

DISCUSSION

Quantitative models describing cell differentiation (e.g., Waddington landscape) were conceptualized decades ago.[46] However, the generation of dynamic and quantitative abstractions of native hematopoiesis has been hampered by a lack of suitable experimental approaches, particularly reaching single-cell resolution. Here, we report a major effort, combining persistent HSC labeling, time-series scRNA-seq analyses, and mathematical modeling to build a predictive model of *in vivo* hematopoiesis dynamics. Analogously to the moving images in a kinetoscope, our approach employs multiple high-resolution snapshots of differentiation to reconstruct the real-time cellular flow between single-cell states within the BM multilineage hematopoiesis. Our model describes cell behavior with self-renewal and differentiation rates, which intuitively can be represented as the shape of a Waddington-like landscape (Figure 7). Using this analogy, the discrete model is a set of fixed platforms connected with slides, whereas the continuous model follows the curvature for all observed states (here: single cells). Differentiation rate indicates the slope between two states, with steeper slopes indicating faster transition. In turn, stable states, the flat areas, have little or no downward slope and combined with proliferation, constitute areas of high self-renewal (Figure 2B).

Differentiation rate and cell fate are naturally connected, but, crucially, exist in specific experimental contexts. CMPs have been originally proposed as a multipotent population with combined erythroid, megakaryocytic, neutrophilic, and monocytic potential.[47] However, later studies reported that most CMPs are transcriptionally and epigenetically primed toward specific lineages,[4] exhibit lineage bias, and are primarily unipotent[5] in transplantation cell fate assays. Importantly, transplantation, as we show in this work, is associated with greatly increased differentiation rates, most likely due to high cellular demand, as other means of ablating cells, similar to 5-FU treatment, also cause accelerated differentiation.[3] *In vitro* assays, performed under cytokine-rich conditions also drive rapid differentiation, and again CMPs also rarely show combined megakaryocyte, erythroid, granulocyte, and monocyte output.[12,47] However, if the differentiation is slowed down and cells given the opportunity to expand (for approx. 3 divisions) under cytokine-restricted conditions (SCF, IL-11, TPO only), >50% CMP clones generate multipotent output after switching to a cytokine-rich secondary culture.[47] Similarly, LMPPs have been described as largely unipotent cells in transplantation assays[48] but can produce multipotent output in two-phase cultures analogous to the CMPs,[49] i.e., given the opportunity to grow first under slower differentiation conditions. Our model, suggests that intermediate clusters 8, 4, 5, which largely overlap with CMPs, are able to slowly transition among each other. In particular, cells can shift from 8 to 4 between the transient megakaryocyte/erythroid-biased cluster 8 and the long-lived myeloid-biased cluster 4, but potential bidirectional transitions are also permitted by our model. This prediction is consistent with cell fates estimated from the static data (using cellrank), where only a small subset of cells is assigned to a single lineage (e.g., ~5% to neutrophil fate, Figure E14C), thus suggesting that at least a subset of CMP cells are balanced and produce multi-linage output. This is also consistent with the *in vivo* observation of progenitors with combined myeloid and megakaryocytic/erythroid outputs.[14,21,50] Importantly, we find that transitions between clusters 4 and 8 are slow. Under strong differentiation conditions (e.g., transplantation or differentiation-promoting media), progenitor cells therefore simply do not have time to 'explore' the multipotent states, thus emphasizing the obvious, but at time underappreciated notion, that if a molecularly multipotent progenitor cell does not divide before being channeled down a particular lineage, alternative fates can never be realized (as illustrated in Figure E14D).

Although tamoxifen has broadly been used to activate CRE in multiple studies,[3,19,51,52] we found that hematopoiesis upon tamoxifen treatment perturbs the steady state in the short term (i.e., first two weeks). Indeed, we observed changes in cluster sizes and differentiation rates associated with tamoxifen treatment, which we teased apart using a bi-phasic model (Figures S6B–S6E). Development of tamoxifen-independent models will help avoid such confounding effects. In the long-term, as mice age, we observed only modest differences of most cluster sizes but observed striking differences in cluster 0 composition. Although further work will be required to better resolve the HSC subpopulations (in cluster 0) and their age-related dynamics, we consider the tentative sub-structure provided here as a critical first step in this

endeavor, as it fits both our data and experimental evidence of HSC behavior in aging mice.[3,19,27,30]

We fully leverage the scRNA-seq approach to extend our model's applicability. To ensure broad accessibility and interpretability, we integrated published annotation from multiple sources.[7,12,20] This places our unified landscape (and its subpopulations) in the biological context of previous immunophenotyping and lineage tracing experiments. Moreover, static cell properties (cluster, pseudotime) and model parameters (differentiation rates, self-renewal) are transferable. Crucially, new scRNA-seq data can be readily incorporated into our landscape and our model is capable of predicting differentiation outcomes for chosen time points given initial conditions, as we demonstrated using an independent time course data.[44] Finally, our model can be used to simulate putative explanations for changes in cell abundance, e.g., between healthy and disease tissues, even if only few snapshot measurements are available. We showcased this capability by shedding light on changes cell dynamics after HSC transplantation, which displays stage and lineage-specific acceleration of differentiation in the erythroid and neutrophilic/monocytic-DC lineages (see transitions 1–9 and 2-3/2-6 respectively).

Differentiation and growth involve coordinated up- and down-regulation of thousands of genes, where it remains unknown for the vast majority of those genes whether and, if so, how they play a role in controlling cell behavior. To access the relevant molecular states with high precision, we introduce the continuous model of near-native hematopoiesis, which includes per-cell growth and differentiation rates, thus providing a direct comparison between cellular behavior and underlying gene expression. We observed complex, sequential gene expression patterns, some of which overlap with increasing differentiation rates, implying irreversible molecular changes. For example, we show that neutrophil differentiation is coupled with expression of multiple lineage determinants (*Irf8*, *Flt3*, *Pou2f2*, and *Gfi1*) followed by a single program taking over and a further increase in differentiation.

The current and predominant view of hematopoiesis has been constructed through the identification of progenitor populations by FACS and definition of their potential by transplantation.[1] This approach not only lacks resolution, but more importantly, researchers end up describing homeostatic hematopoiesis within a framework derived from assays that measure potential in a non-homeostatic context[3]; transplantation defines potential in a non-homeostatic assay and therefore does not reveal the actual contribution of any given population to steady-state hematopoiesis. The revolution of single-cell transcriptomics has provided evidence for additional progenitor populations,[4,6,7,21] but so far had been severely limited by having to place those putative populations on a static transplantation-defined map of hematopoiesis. Here we have overcome these shortcomings by observing near-native hematopoiesis *in situ* and over time.

The combination of lineage tracing with a single cell transcriptomics chase delivered a truly quantitative and dynamic model of hematopoiesis including previously unknown dynamic relationships between precisely defined stem and progenitor cells. The model also reveals fundamental quantitative system properties from cell trajectories, cell division rates, and number of cell divisions to individual lineage-specific differentiation rates.

Unshackling the field from the static transplantation-defined view of hematopoiesis shifts the paradigm from qualitative models with limited predictive capabilities to integrative, quantitative, and predictive models. The latter are highly transferable and thus key to providing insight into human hematopoiesis, where experimental options are limited. As recently demonstrated scRNA-seq can be integrated across species,[53–55] thus potentially enable mapping HSPC dynamics onto human counterparts. Self-renewal and differentiation capacities are particularly relevant to leukemia research because they are the precise cellular behaviors whose dysregulation causes the malignant phenotype. As we show here and supported by previous studies,[3,22] progenitors can also operate close to self-renewal and a small proliferative advantage may be sufficient to immortalize them. Finally, population dynamic models are universally applicable across biological fields, as adult tissues are commonly replenished from their own stem cell pools.[56] To inspire such future endeavors, we showcase how to build a model connecting high-resolution molecular information with tissue-scale cell behavior.

Limitations of the study

Despite vastly improved resolution over immunophenotyping, scRNA-seq does not capture cellular states in full. Additional variables such as chromatin state, protein levels, metabolism, and environmental factors also affect cell behavior and may manifest in unappreciated heterogeneity and dynamic properties. These characteristics may be heritable in which case they may be tractable with lineage tracing approaches. In addition, the discrete model relies on hard clustering, which averages any finer cell heterogeneity. Although most of the early cell fate decisions will occur within the landscape presented in this work, with increased throughput a BM-wide landscape could be generated, thus providing better insight into the entire lymphoid and myeloid differentiation trajectories. More work will also be required to better understand hematopoiesis dynamics in a wide-range of non-homeostatic settings such as inflammation or chemotherapy.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Animals
- METHOD DETAILS
  - Hoxb5^{CreERT2} and Hoxb5^{mKO2} mouse lines
  - Transplantation assays and hematopoietic reconstitution analysis
  - Induction of reporter gene expression by tamoxifen
  - Flow cytometry
  - Cell isolation for the scRNA-Seq experiments
  - scRNA-seq data generation

- 10X genomics
- scRNA-Seq data analysis
- mKO2 cell analysis
- Subclustering of cluster 0
- Embedding external datasets into the integrated HSPC landscape
- Trajectory inference and selection
- Differential expression analysis

- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Flow cytometry data analysis
  - Discrete population model analysis
  - Generalized model for testing alternative topologies
  - Model selection for perturbed systems
  - Continuous population model analysis
  - Transplantation data analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.stem.2023.12.001.

ACKNOWLEDGMENTS

The authors thank Reiner Schulte, Chiara Cossetti, and Gabriela Grondys-Kotarba from the Cambridge Institute for Medical Research Flow Cytometry Core facility for their assistance with cell sorting. We would also like to thank Katarzyna Kania and others at the Cancer Research UK Cambridge Institute Genomics Core Facility for generating the 10x Genomics libraries and performing high-throughput sequencing. The authors are also grateful to all staff of the Biological Services Unit at Queen Mary University of London for their technical support. Work in the Kranc Laboratory is supported by Cancer Research UK (C29967/A14633 and C29967/A26787) and Barts Charity and Blood Cancer UK. The O'Carroll laboratory is supported by the Wellcome Trust Investigator Award (106144), the Wellcome Centre for Cell Biology (203149), and a Wellcome multi-user equipment grant (108504). Work in the Göttgens laboratory is supported by Wellcome (206328/Z/17/Z and 203151/Z/16/Z), Blood Cancer UK(18002), Cancer Research UK (C1163/A21762), and UKRI Medical Research Council (MC_PC_17230). For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

AUTHOR CONTRIBUTIONS

Part 1 (Hoxb5-mKO2 and Hoxb5-Tom model: conceptualization, generation, and characterization): Conceptualization, K.R.K. and D.O.; methodology, J.C., F.S., N.B., P.N.M., K.R.K., and D.O.; software, M.B.; validation, J.C., F.S., N.B., P.N.M., L.A., H.L., K.R.K., and D.O.; formal analysis, I.K., J.C., M.B., F.S., N.B., K.R.K., D.O., and B.G.; investigation, J.C., F.S., N.B., P.N.M., L.A., and H.L.; resources, J.C., F.S., N.B., P.N.M., L.A., H.L., K.R.K., and D.O.; data curation, J.C., F.S., N.B., L.A., H.L., K.R.K., and D.O.; writing – original draft, I.K., writing – review & editing, I.K., J.C., M.B., K.R.K., D.O., and B.G.; visualization, I.K., J.C., M.B., F.S., N.B., K.R.K., D.O., and B.G.; supervision, J.C., H.L., K.R.K., D.O., and B.G.; project administration, J.C., F.S., N.B., P.N.M., L.A., H.L., K.R.K., D.O., and B.G.; funding acquisition, K.R.K. and D.O.

Part 2 (scRNA-seq and dynamics modeling conceptualization, data generation, and analysis): Conceptualization, I.K., M.B., and B.G.; methodology, I.K., M.B., and B.G.; software, I.K. and M.B.; validation, I.K., M.B., and B.G.; formal analysis, I.K., M.B., and B.G.; investigation, I.K., J.C., N.B., M.L.R.H., and S.J.K; resources, I.K., J.C., M.B., K.R.K., D.O., and B.G.; data curation, I.K., M.B., and B.G.; writing – original draft, I.K. and M.B.; writing – review & editing, I.K., M.B., K.R.K., D.O., and B.G.; visualization, I.K., M.B., and B.G.; supervision, I.K. and B.G.; project administration, I.K., J.C., M.B., K.R.K., and B.G.; funding acquisition, K.R.K. and B.G.

DECLARATION OF INTERESTS

N.B. is now an employee of AstraZeneca. I.K. is now an employee of Xap Therapeutics.

Received: September 7, 2023
Revised: September 25, 2023
Accepted: December 4, 2023
Published: January 5, 2024

REFERENCES

1. Seita, J., and Weissman, I.L. (2010). Hematopoietic stem cell: self-renewal versus differentiation. Wiley Interdiscip. Rev. Syst. Biol. Med. *2*, 640–653.
2. Reya, T., Morrison, S.J., Clarke, M.F., and Weissman, I.L. (2001). Stem cells, cancer, and cancer stem cells. Nature *414*, 105–111.
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.
4. Paul, F., Arkin, Y., Giladi, A., Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Winter, D., Lara-Astiaso, D., Gury, M., Weiner, A., et al. (2015). Transcriptional heterogeneity and lineage commitment in myeloid progenitors. Cell *163*, 1663–1677.
5. Perié, L., Duffy, K.R., Kok, L., de Boer, R.J., and Schumacher, T.N. (2015). The branching point in erythro-myeloid differentiation. Cell *163*, 1655–1662.
6. Klein, F., Roux, J., Cvijetic, G., Rodrigues, P.F., von Muenchow, L., Lubin, R., Pelczar, P., Yona, S., Tsapogas, P., and Tussiwand, R. (2022). Dntt expression reveals developmental hierarchy and lineage specification of hematopoietic progenitors. Nat. Immunol. *23*, 505–517.
7. Nestorowa, S., Hamey, F.K., Pijuan Sala, B., Diamanti, E., Shepherd, M., Laurenti, E., Wilson, N.K., Kent, D.G., and Göttgens, B. (2016). A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. Blood *128*, e20–e31.
8. Velten, L., Haas, S.F., Raffel, S., Blaszkiewicz, S., Islam, S., Hennig, B.P., Hirche, C., Lutz, C., Buss, E.C., Nowak, D., et al. (2017). Human haematopoietic stem cell lineage commitment is a continuous process. Nat. Cell Biol. *19*, 271–281.
9. Dahlin, J.S., Hamey, F.K., Pijuan-Sala, B., Shepherd, M., Lau, W.W.Y., Nestorowa, S., Weinreb, C., Wolock, S., Hannah, R., Diamanti, E., et al. (2018). A single-cell hematopoietic landscape resolves 8 lineage trajectories and defects in Kit mutant mice. Blood *131*, e1–e11.
10. Tusi, B.K., Wolock, S.L., Weinreb, C., Hwang, Y., Hidalgo, D., Zilionis, R., Waisman, A., Huh, J.R., Klein, A.M., and Socolovsky, M. (2018). Population snapshots predict early haematopoietic and erythroid hierarchies. Nature *555*, 54–60.
11. Wang, S.W., Herriges, M.J., Hurley, K., Kotton, D.N., and Klein, A.M. (2022). CoSpar identifies early cell fate biases from single-cell transcriptomic and lineage information. Nat. Biotechnol. *40*, 1066–1074.
12. Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F.D., and Klein, A.M. (2020). Lineage tracing on transcriptional landscapes links state to fate during differentiation. Science *367*, eaaw3381.
13. Yeo, G.H.T., Saksena, S.D., and Gifford, D.K. (2021). Generative modeling of single-cell time series with prescient enables prediction of cell trajectories with interventions. Nat. Commun. *12*, 3222.
14. 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.e8.
15. Montoro, D.T., Haber, A.L., Biton, M., Vinarsky, V., Lin, B., Birket, S.E., Yuan, F., Chen, S., Leung, H.M., Villoria, J., et al. (2018). A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. Nature *560*, 319–324.

16. 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.

17. Hamey, F.K., and Göttgens, B. (2019). Machine learning predicts putative hematopoietic stem cells within large single-cell transcriptomics data sets. Exp. Hematol. *78*, 11–20.

18. Madisen, L., Zwingman, T.A., Sunkin, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L., Palmiter, R.D., Hawrylycz, M.J., Jones, A.R., et al. (2010). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. Nat. Neurosci. *13*, 133–140.

19. Barile, M., Busch, K., Fanti, A.-K., Greco, A., Wang, X., Oguro, H., Zhang, Q., Morrison, S.J., Rodewald, H.-R., and Höfer, T. (2020). Hematopoietic stem cells self-renew symmetrically or gradually proceed to differentiation. preprint at bioRxiv *1234*. https://doi.org/10.1101/2020.08.06.239186.

20. 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.e27.

21. 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.

22. Takahashi, M., Barile, M., Chapple, R.H., Tseng, Y.J., Nakada, D., Busch, K., Fanti, A.K., Säwén, P., Bryder, D., Höfer, T., et al. (2021). Reconciling flux experiments for quantitative modeling of normal and malignant hematopoietic stem/progenitor dynamics. Stem Cell Rep. *16*, 741–753.

23. Wolf, F.A., Hamey, F.K., Plass, M., Solana, J., Dahlin, J.S., Göttgens, B., Rajewsky, N., Simon, L., and Theis, F.J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol. *20*, 59.

24. Sánchez-Aguilera, A., Arranz, L., Martín-Pérez, D., García-García, A., Stavropoulou, V., Kubovcakova, L., Isern, J., Martín-Salamanca, S., Langa, X., Skoda, R.C., et al. (2014). Estrogen signaling selectively induces apoptosis of hematopoietic progenitors and myeloid neoplasms without harming steady-state hematopoiesis. Cell Stem Cell *15*, 791–804.

25. Hamey, F.K., Lau, W.W.Y., Kucinski, I., Wang, X., Diamanti, E., Wilson, N.K., Göttgens, B., and Dahlin, J.S. (2021). Single-cell molecular profiling provides a high-resolution map of basophil and mast cell development. Allergy *76*, 1731–1742.

26. Wu, C., Boey, D., Bril, O., Grootens, J., Vijayabaskar, M.S., Sorini, C., Ekoff, M., Wilson, N.K., Ungerstedt, J.S., Nilsson, G., and Dahlin, J.S. (2022). Single-cell transcriptomics reveals the identity and regulators of human mast cell progenitors. Blood Adv. *6*, 4439–4449.

27. Zhang, L., Mack, R., Breslin, P., and Zhang, J. (2020). Molecular and cellular mechanisms of aging in hematopoietic stem cells and their niches. J. Hematol. Oncol. *13*, 157.

28. Benz, C., Copley, M.R., Kent, D.G., Wohrer, S., Cortes, A., Aghaeepour, N., Ma, E., Mader, H., Rowe, K., Day, C., et al. (2012). Hematopoietic stem cell subtypes expand differentially during development and display distinct lymphopoietic programs. Cell Stem Cell *10*, 273–283.

29. Muller-Sieburg, C.E., Cho, R.H., Karlsson, L., Huang, J.F., and Sieburg, H.B. (2004). Myeloid-biased hematopoietic stem cells have extensive self-renewal capacity but generate diminished lymphoid progeny with impaired IL-7 responsiveness. Blood *103*, 4111–4118.

30. 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.e4.

31. Fischer, D.S., Fiedler, A.K., Kernfeld, E.M., Genga, R.M.J., Bastidas-Ponce, A., Bakhti, M., Lickert, H., Hasenauer, J., Maehr, R., and Theis, F.J. (2019). Inferring population dynamics from single-cell RNA-sequencing time series data. Nat. Biotechnol. *37*, 461–468.

32. Lange, M., Bergen, V., Klein, M., Setty, M., Reuter, B., Bakhti, M., Lickert, H., Ansari, M., Schniering, J., Schiller, H.B., et al. (2022). CellRank for directed single-cell fate mapping. Nat. Methods *19*, 159–170.

33. Setty, M., Kiseliovas, V., Levine, J., Gayoso, A., Mazutis, L., and Pe'er, D. (2019). Characterization of cell fate probabilities in single-cell data with Palantir. Nat. Biotechnol. *37*, 451–460.

34. Olsson, A., Venkatasubramanian, M., Chaudhri, V.K., Aronow, B.J., Salomonis, N., Singh, H., and Grimes, H.L. (2016). Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. Nature *537*, 698–702.

35. Ma, C., and Staudt, L.M. (1996). LAF-4 Encodes a Lymphoid Nuclear Protein with Transactivation Potential That Is Homologous to AF-4, the Gene Fused to MLL in t(4;11) Leukemias. Blood *87*, 734–745.

36. Amann-Zalcenstein, D., Tian, L., Schreuder, J., Tomei, S., Lin, D.S., Fairfax, K.A., Bolden, J.E., McKenzie, M.D., Jarratt, A., Hilton, A., et al. (2020). A new lymphoid-primed progenitor marked by Dach1 downregulation identified with single cell multi-omics. Nat. Immunol. *21*, 1574–1584.

37. Kumar, P., Beck, D., Galeev, R., Thoms, J.A.I., Talkhoncheh, M.S., de Jong, I., Unnikrishnan, A., Baudet, A., Subramaniam, A., Pimanda, J.E., and Larsson, J. (2019). HMGA2 promotes long-term engraftment and myeloerythroid differentiation of human hematopoietic stem and progenitor cells. Blood Adv. *3*, 681–691.

38. Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T., et al. (2011). Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell *144*, 296–309.

39. Park, S.M., Cho, H., Thornton, A.M., Barlowe, T.S., Chou, T., Chhangawala, S., Fairchild, L., Taggart, J., Chow, A., Schurer, A., et al. (2019). IKZF2 drives leukemia stem cell self-renewal and inhibits myeloid differentiation. Cell Stem Cell *24*, 153–165.e7.

40. Li, J., Kurasawa, Y., Wang, Y., Clise-Dwyer, K., Klumpp, S.A., Liang, H., Tailor, R.C., Raymond, A.C., Estrov, Z., Brandt, S.J., et al. (2014). Requirement for ssbp2 in hematopoietic stem cell maintenance and stress response. J. Immunol. *193*, 4654–4662.

41. Notta, F., Zandi, S., Takayama, N., Dobson, S., Gan, O.I., Wilson, G., Kaufmann, K.B., McLeod, J., Laurenti, E., Dunant, C.F., et al. (2016). Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. Science *351*, aab2116.

42. 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.

43. Boyer, S.W., Rajendiran, S., Beaudin, A.E., Smith-Berdan, S., Muthuswamy, P.K., Perez-Cunningham, J., Martin, E.W., Cheung, C., Tsang, H., Landon, M., and Forsberg, E.C. (2019). Clonal and quantitative in vivo assessment of hematopoietic stem cell differentiation reveals strong erythroid potential of multipotent cells. Stem Cell Rep. *12*, 801–815.

44. Upadhaya, S., Sawai, C.M., Papalexi, E., Rashidfarrokhi, A., Jang, G., Chattopadhyay, P., Satija, R., and Reizis, B. (2018). Kinetics of adult hematopoietic stem cell differentiation in vivo. J. Exp. Med. *215*, 2815–2832.

45. Dong, F., Hao, S., Zhang, S., Zhu, C., Cheng, H., Yang, Z., Hamey, F.K., Wang, X., Gao, A., Wang, F., et al. (2020). Differentiation of transplanted haematopoietic stem cells tracked by single-cell transcriptomic analysis. Nat. Cell Biol. *22*, 630–639.

46. Waddington, C.H. (1957). The Strategy of the Genes (George Allen & Unwin).

47. Akashi, K., Traver, D., Miyamoto, T., and Weissman, I.L. (2000). A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature *404*, 193–197.

48. 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.

49. Adolfsson, J., Månsson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C.T., Bryder, D., Yang, L., Borge, O.J., Thoren, L.A.M., et al.

(2005). Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. Cell *121*, 295–306.

50. 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.
51. Göthert, J.R., Gustin, S.E., Hall, M.A., Green, A.R., Göttgens, B., Izon, D.J., and Begley, C.G. (2005). In vivo fate-tracing studies using the Scl stem cell enhancer: embryonic hematopoietic stem cells significantly contribute to adult hematopoiesis. Blood *105*, 2724–2732.
52. Joseph, C., Quach, J.M., Walkley, C.R., Lane, S.W., Lo Celso, C., and Purton, L.E. (2013). Deciphering hematopoietic stem cells in their niches: a critical appraisal of genetic models, lineage tracing, and imaging strategies. Cell Stem Cell *13*, 520–533.
53. Lotfollahi, M., Wolf, F.A., and Theis, F.J. (2019). scGen predicts single-cell perturbation responses. Nat. Methods *16*, 715–721.
54. Lotfollahi, M., Naghipourfar, M., Luecken, M.D., Khajavi, M., Büttner, M., Wagenstetter, M., Avsec, Ž., Gayoso, A., Yosef, N., Interlandi, M., et al. (2022). Mapping single-cell data to reference atlases by transfer learning. Nat. Biotechnol. *40*, 121–130.
55. Welch, J.D., Kozareva, V., Ferreira, A., Vanderburg, C., Martin, C., and Macosko, E.Z. (2019). Single-cell multi-omic integration compares and contrasts features of brain cell identity. Cell *177*, 1873–1887.e17.
56. Goodell, M.A., Nguyen, H., and Shroyer, N. (2015). Somatic stem cell heterogeneity: diversity in the blood, skin and intestinal stem cell compartments. Nat. Rev. Mol. Cell Biol. *16*, 299–309.
57. Lawson, H., Sepulveda, C., van de Lagemaat, L.N., Durko, J., Barile, M., Tavosanis, A., Georges, E., Shmakova, A., Timms, P., Carter, R.N., et al. (2021). JMJD6 promotes self-renewal and regenerative capacity of hematopoietic stem cells. Blood Adv. *5*, 889–899.
58. Mapperley, C., van de Lagemaat, L.N., Lawson, H., Tavosanis, A., Paris, J., Campos, J., Wotherspoon, D., Durko, J., Sarapuu, A., Choe, J., et al. (2021). The mRNA m6A reader YTHDF2 suppresses proinflammatory pathways and sustains hematopoietic stem cell function. J. Exp. Med. *218*, e20200829.
59. Bagnoli, J.W., Ziegenhain, C., Janjic, A., Wange, L.E., Vieth, B., Parekh, S., Geuder, J., Hellmann, I., and Enard, W. (2018). Sensitive and powerful single-cell RNA sequencing using mcSCRB-seq. Nat. Commun. *9*, 2937.
60. Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181.
61. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15.
62. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell *177*, 1888–1902.e21.
63. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods *16*, 1289–1296.
64. Hie, B., Bryson, B., and Berger, B. (2019). Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. Nat. Biotechnol. *37*, 685–691.
65. Polański, K., Young, M.D., Miao, Z., Meyer, K.B., Teichmann, S.A., and Park, J.E. (2020). BBKNN: fast batch alignment of single cell transcriptomes. Bioinformatics *36*, 964–965.
66. Haghverdi, L., Lun, A.T.L., Morgan, M.D., and Marioni, J.C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat. Biotechnol. *36*, 421–427.
67. McInnes, L., Healy, J., and Melville, J. (2020). UMAP: uniform manifold approximation and projection for dimension reduction. Preprint at arXiv *1234*. https://doi.org/10.48550/arXiv.1802.03426.
68. Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. *9*, 5233.
69. Ravasi, T., Suzuki, H., Cannistraci, C.V., Katayama, S., Bajic, V.B., Tan, K., Akalin, A., Schmeier, S., Kanamori-Katayama, M., Bertin, N., et al. (2010). An atlas of combinatorial transcriptional regulation in mouse and Man. Cell *140*, 744–752.
70. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90–W97.
71. Kent, D.G., Copley, M.R., Benz, C., Wöhrer, S., Dykstra, B.J., Ma, E., Cheyne, J., Zhao, Y., Bowie, M.B., Zhao, Y., et al. (2009). Prospective isolation and molecular characterization of hematopoietic stem cells with durable self-renewal potential. Blood *113*, 6342–6350.
72. Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., and Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics *25*, 1923–1929.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-Mouse CD45.1 (BV605 conjugated, clone A20)	Biolegend	Cat#110738; RRID: AB_11204076
Anti-Mouse CD45.2 (PerCP conjugated, clone 104)	Biolegend	Cat#109826; RRID: AB_893349
Anti-Mouse CD4 (Biotin conjugated, clone H129.19)	BD Biosciences	Cat#553649; RRID: AB_394969
Anti-Mouse CD5 (Biotin conjugated, clone 53-7.3)	BD Biosciences	Cat#553019; RRID: AB_394557
Anti-Mouse CD8a (Biotin conjugated, clone 53-6.7)	BD Biosciences	Cat#553029; RRID: AB_394567
Anti-Mouse CD11b (Biotin conjugated, clone M1/70)	BD Biosciences	Cat#553309; RRID: AB_394773
Anti-Mouse CD45R/B220 (Biotin conjugated, clone RA3-6B2)	BD Biosciences	Cat#553086; RRID: AB_394616
Anti-Mouse Gr-1/Ly-6G/C (Biotin conjugated, clone RB6-8C5)	BD Biosciences	Cat#553125; RRID: AB_394641
Anti-Mouse Ter119 (Biotin conjugated, clone TER-119)	BD Biosciences	Cat#553672; RRID: AB_394985
Anti-Mouse c-Kit/CD117 (BV711 conjugated, clone 2B8)	Biolegend	Cat#105835; RRID: AB_2565956
Anti-Mouse Sca-1 (APC-Cy7 conjugated, clone D7)	Biolegend	Cat#108126; RRID: AB_10645327
Anti-Mouse CD48 (APC conjugated, clone HM48-1)	Biolegend	Cat#103411; RRID: AB_571996
Anti-Mouse CD150 (PE-Cy7 conjugated, clone 12F12.2)	Biolegend	Cat#115914; RRID: AB_439797
Anti-Mouse CD19 (Biotin conjugated, clone 1D3)	Biolegend	Cat#1; 553784; RRID: AB_395048
Anti-Mouse Sca-1 (PB conjugated, clone D7)	Biolegend	Cat#108120; RRID: AB_493273
Anti-Mouse CD16/32 (APC-CY7 conjugated, clone 93)	Biolegend	Cat#101328; RRID: AB_2104158
Anti-Mouse CD41 (BV605 conjugated, clone MWReg30)	Biolegend	Cat#133921; RRID: AB_2563933
Anti-Mouse CD105 (APC conjugated, clone MJ7/18)	Biolegend	Cat#120413; RRID: AB_2277915
Anti-Mouse Ter119 (FITC conjugated, clone TER-119)	Biolegend	Cat#116206; RRID: AB_313707
Anti-Mouse CD45R/B220 (APC conjugated, clone RA3-6B2)	Biolegend	Cat#103212; RRID: AB_312997
Anti-Mouse CD19 (APC-Cy7 conjugated, clone 6D5)	Biolegend	Cat#115529; RRID: AB_830707
Anti-Mouse CD11b (PB conjugated, clone M1/70)	Biolegend	Cat#101224; RRID: AB_755986
Anti-Mouse Gr-1/Ly-6G/C (PE-Cy7 conjugated, clone RB6-8C5)	Biolegend	Cat#108416; RRID: AB_313381
Anti-Mouse Ter119 (PE-Cy5 conjugated, clone TER-119)	Biolegend	Cat#116210; RRID: AB_313711
Anti-Mouse CD8a (APC conjugated, clone 53-6.7)	Biolegend	Cat#100712; RRID: AB_312751
Anti-Mouse CD4 (APC conjugated, clone GK1.5)	Biolegend	Cat#100411; RRID: AB_312696
Anti-Mouse CD8a (APC-CY7 conjugated, clone YTS156.7.7)	Biolegend	Cat#126620; RRID: AB_2563951
Anti-Mouse CD25 (PB conjugated, clone PC61)	Biolegend	Cat#102022; RRID: AB_493643
Anti-Mouse CD44 (PE-CY7 conjugated, clone IM7)	Biolegend	Cat#103030; RRID: AB_830787
Fc Block (anti-mouse CD16/32, clone 93)	Biolegend	Cat#101320; RRID: AB_1574975
Streptavidin (PerCP conjugated)	Biolegend	Cat#405213
Streptavidin (Pacific Blue conjugated)	ThermoFisher Scientific	Cat# S11222
DAPI	BD Biosciences	Cat#564907; RRID: AB_2869624
Mouse hematopoietic progenitor cell isolation cocktail	Stem Cell Technologies	19856

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CD48-APC	ThermoFischer	17-0481-82, RRID:AB_469408
c-Kit-APC/Cy7	Biolegend	105826, RRID:AB_1626278
Sca1-BV421	Biolegend	108133, RRID:AB_2650926
CD150-PE/Cy7	Biolegend	115914, RRID:AB_439797
Streptavidin-BV510	Biolegend	405234
Chemicals, peptides, and recombinant proteins		
Tamoxifen	Sigma	T5648; CAS: 10540-29-1
Corn Oil	Sigma	C8267; CAS: 8001-30-7
DAPI	BD Pharmigen	564907
Ammonium Chloride	Stem Cell Technologies	07800
SUPERase-In RNase Inhibitor	ThermoFisher	AM2694
dNTP mix	ThermoFisher	10319879
ERCC RNA Spike-In Mix	ThermoFisher	4456740
Maxima H minus Reverse Transcriptase	ThermoFisher	EP0753
Terra PCR Direct Polymerase Mix	Takara	639270
Agencourt AMPure XP beads	Beckman Coulter	A63881
Nextera XT DNA sample preparation kit 96 samples	Illumina	FC-131-1096
Triton X-100 solution	Sigma	93443
PEG 8000 solution	Sigma	P1458
Phusion polymerase	ThermoFischer	F530L
Critical commercial assays		
10x Genomics Single Cell 3' v3	10X Genomics	PN-1000268
Deposited data		
Sequencing data	This paper	GEO: GSE207412
Pre-processed input data	This paper	Mendeley Data: https://doi.org/10.17632/vwg6xzmrf9.1
Extended data figures	This paper	Mendeley Data: https://doi.org/10.17632/vwg6xzmrf9.1
Extended data tables	This paper	Mendeley Data: https://doi.org/10.17632/vwg6xzmrf9.1
scRNA-Seq data from Bowling et al.	Bowling et al.[20]	GEO: GSE146972 See pre-processed input data
scRNA-Seq data from Nestorowa et al.	Nestorowa et al.[7]	GEO: GSE81682 See pre-processed input data
scRNA-Seq data from Weinreb et al.	Weinreb et al.[11]	GEO: GSE140802 See pre-processed input data
scRNA-Seq data from Dong et al.	Dong et al. data[45]	GEO: GSE116530 See pre-processed input data
scRNA-Seq data from Upadhaya et al.	Upadhaya et al.[44]	GEO: GSE120239 See pre-processed input data
Interactive visualization of the HSPC landscape and model parameters	This paper	https://gottgens-lab.stemcells.cam.ac.uk/bgweb2/HSPC_dyn2022/
Experimental models: Organisms/strains		
Mouse: Hoxb5^{mKO2}	This paper	N/A
Mouse: Hoxb5^{CreERT2}	This paper	N/A
Oligonucleotides		
See Table S7 for list of oligonucleotides and ssDNA	This paper	N/A
TSO 5′-AAGCAGTGGTATCAACGCAGA GTACATrGrG+G-3′	IDT	NA
Oligo-dT30VN 5′–AAGCAGTGGTATCAA CGCAGAGTAC(T30)VN-3'	IDT	NA

(*Continued on next page*)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
ISPCR oligo 5′-AAGCAGTGGTATCAAC GCAGAGT-3′	IDT	NA
Nextera XT 96-Index kit, 384 samples	Illumina	FC-131-1002
Recombinant DNA		
Topo blunt HA-P2A-CRE-ERT2-HA plasmid	This paper	NA
Topo blunt HA-P2A-Mko2-p2A-Mko2-Caax-HA plasmid	This paper	NA
Software and algorithms		
FlowJo v10	FlowJo,Tree Star Inc.	N/A
GraphPad Prism 6 software	GraphPad Software, Inc.	N/A
Analysis code	This paper	Mendeley Data: https://doi.org/10.17632/vwg6xzmrf9.1 and Github: https://github.com/Iwo-K/HSPCdynamics
Singularity container (containing all scRNA-Seq analysis software)	This paper	Mendeley Data: https://doi.org/10.17632/vwg6xzmrf9.1
Cellproject	This paper	https://github.com/Iwo-K/cellproject
Cellranger v3.1.0	10X genomics	NA

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Berthold Göttgens (bg200@cam.ac.uk)

Materials availability

Plasmids and mouse lines generated in this work are available upon request.

Data and code availability

Single-cell RNA-seq data have been deposited at GEO and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.

Original extended data figures, tables, pre-processed input data and software environment (Singularity container) and analysis code have been deposited at Mendeley and are publicly available as of the date of publication. The DOI is listed in the key resources table.

This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table.

An interactive visualization of the landscape is available on a dedicated website and is publicly available as of the date of publication. The URL is listed in the key resources table.

All original code has been deposited at Mendeley Data and is publicly available as of the date of publication. DOIs are listed in the key resources table.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

External online data will be maintained long-term using: the GEO repository (sequencing data), Mendeley Data repository (code, pre-processed input data and software environment) and a dedicated server maintained by the University of Cambridge Stem Cell Institute (interactive visualization).

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Animals

All experiments on animals were in accordance with the guidelines of the Queen Mary University of London and performed under UK Home Office authorisation (Project License number: PP4153210, Project Licence holder: Kamil R Kranc). Animals were maintained under specific pathogen-free conditions. All mice were on the C57BL/6 genetic background. Both male and female mice at the age of 8–12 weeks were employed in this study and no obvious sex differences were found between the sexes. *Rosa26-LoxP-STOP-LoxP-tdTomato* (*R26*$^{LSL\text{-}tdTomato}$) reporter mice were published previously.[18] The *Hoxb5*^{CreERT2} and *Hoxb5*^{mKO2} alleles were generated using CRISPR-Cas9 gene editing technology, using single strand donor DNA encoding the P2A-CRE-ERT2 protein and mKO2-P2A-mKO2-CAAX protein, respectively (please see Hoxb5CreERT2 and Hoxb5mKO2 mouse lines for further details). Congenic

CD45.1$^+$/CD45.2$^+$ recipient mice used for transplantation experiments were generated by intercrossing B6.SJL-Ptprca Pepcb/BoyJ mice with C57BL/6J mice (both strains were purchased from The Jackson Laboratory).

METHOD DETAILS

Hoxb5^{CreERT2} and Hoxb5^{mKO2} mouse lines

The *Hoxb5^{CreERT2}* and *Hoxb5^{mKO2}* alleles were generated using CRISPR-Cas9 gene editing technology employing fertilized 1-cell zygotes on the B6CBAF1/Crl genetic background. For the *Hoxb5^{CreERT2}* allele, we injected a single 15 ng/ul sgRNA (tcctccggatgggctca)[16] together with 25 ng/ul CAS9 mRNA and 17.5 ng/ul single strand donor DNA encoding the P2A-CRE-ERT2 protein flanked by 70 nucleotides of homology arms (Table S7). For the *Hoxb5^{mKO2}* allele, we used the same concentrations of sgRNA (tcctccggatgggctca), CAS9 mRNA and single strand donor DNA encoding the mKO2-P2A-mKO2-CAAX (Table S7). The F0 offspring was screened by PCR and Sanger sequencing. The *Hoxb5^{CreERT2}* and *Hoxb5^{mKO2}* lines were established from one founder animals, respectively, and back-crossed several times to the C57BL/6N genetic background. Mice were genotyped by PCR using primers detailed in Table S7.

Transplantation assays and hematopoietic reconstitution analysis

Primary and secondary transplanted recipient mice (CD45.1$^+$/CD45.2$^+$) were lethally irradiated with a split dose of 8 Gy (two doses of 4 Gy administered at least 4 hours apart). For primary transplantations, mice were tail-vein injected with 200 Hoxb5$^+$ or Hoxb5$^-$ HSCs (LSK CD48$^-$CD150$^+$) sorted from Hoxb5-mKO2 animals, together with 2x10^5 support CD45.1$^+$ unfractionated BM cells. For secondary transplantations, 3 000 CD45.2$^+$ LSK cells sorted from BM of primary recipients were mixed with 2x10^5 support CD45.1$^+$ unfractionated BM cells and re-transplanted. PB of all recipient mice was analyzed up to 21 weeks after primary and secondary transplantations. Leukocytes and HSCs (LSK CD48$^-$CD150$^+$) were stained as described below for flow cytometry analysis of PB and BM, except cells were also incubated with CD45.1-BV605 and CD45.2-PercP antibodies. For each mouse, the percentage of donor chimerism in the analyzed cell compartment was defined as the percentage of CD45.1$^-$/CD45.2$^+$ cells among total CD45.1$^-$/CD45.2$^+$ and CD45.1$^+$/CD45.2$^-$ cells, after exclusion of recipient fraction (CD45.1$^+$/CD45.2$^+$).

Induction of reporter gene expression by tamoxifen

Tamoxifen (1g) was dissolved in 10 mL absolute ethanol and 90 mL corn oil at 37°C. Aliquots of tamoxifen (10 mg/mL) were stored at -20 °C. 8-12 weeks *Hoxb5^{CreERT2}; tdTomato* mice were injected intraperitoneally (i.p.) with tamoxifen at 100 mg/kg body weight for 7 days. As controls for subsequent lineage tracing experiments, mice with same genotype were injected with equivalent volume of corn oil to determine whether any labelling was present in the absence of induction. *Hoxb5^{WT}; tdTomato* mice treated with tamoxifen were also analysed to confirm no background or tamoxifen-induced changes.

Flow cytometry

At end point analyses, the fraction of mKO2$^+$ and Tom$^+$ cells was determined in various hematopoietic compartments of BM, PB, spleens, thymi and lymph nodes. Cells from those tissues were prepared and analyzed as described previously.[57,58]

For HSC and progenitor cell analyses, unfractionated BM cells were incubated with Fc block, followed by biotin-conjugated anti-lineage marker antibodies (CD4, CD5, CD11b, B220, CD8a, Gr1 and Ter119), cKit-BV711, Sca1-APC/Cy7, CD48-APC and CD150-PE/Cy7 antibodies. Biotin-conjugated antibodies were then stained with Pacific blue-conjugated streptavidin. DAPI was used for dead cell exclusion.

For staining of megakaryocyte and erythroid progenitors, unfractionated BM cells were incubated with antibodies against lineage markers as described above, except Ter119 antibody was replaced by biotin-conjugated anti-CD19. Cells were stained together with cKit-BV711, Sca1-PB, CD150-PE/Cy7, CD16/32-APC/Cy7, CD41-BV605, CD105-APC and Ter119-FITC antibodies. Biotin-conjugated antibodies were then stained with PerCP-conjugated streptavidin.

For analyses of differentiated cells in the BM, cell suspensions were stained with B220-APC and CD19-APC/Cy7 antibodies for B cells, CD11b-PB and Gr1-PE/Cy7 for myeloid cells and Ter119-FITC for erythroid cells.

PB samples were collected from tail vein into EDTA-coated capillary tubes (Sarstedt). 1-2μL of unfractionated PB were used for analysis of erythrocytes, mixed with 10uL of platelet solution. Platelets were separated by centrifugation of PB samples at 100g for 10 min at room temperature. Platelets were identified as Ter119-PE/Cy5$^-$CD150-PE/Cy7$^+$CD41-BV605$^+$, and Ter119-PeCy5$^+$ cells were erythrocytes.

For analyses of leukocytes in PB, spleen and lymph node, myeloid cells were stained as above for BM cells, T cells with CD8a-APC and CD4-APC antibodies, and CD19-APC/Cy7 antibodies were used to detect B cells.

Cell suspensions from thymus were incubated with the biotin-conjugated anti-lineage marker antibodies described above together with CD4-APC, CD8b-APC/Cy7, CD25-PB and CD44-PE/Cy7 antibodies. Biotin-conjugated antibodies were then stained with PerCP-conjugated streptavidin.

Flow cytometry data were acquired by LSRFortessa (BD) and analysed with FlowJo software (TreeStar, v10).

Cell isolation for the scRNA-Seq experiments

Hoxb5-Tom experiments

All steps in this section (unless otherwise indicated) were performed on ice, and centrifugation steps performed at 300g, 4°C for 5 min. 8-12 weeks old mice carrying the Hoxb5-Cre and the Rosa26-LoxP-STOP-LoxP-tdTomato constructs were treated with 7 daily injections of tamoxifen (as described above) and sacrificed at indicated time-points. BM cells were extracted from ilia, tibiae and femora by grinding with mortar and pestle in PBS supplemented with 2% Fetal Bovine Serum (cell buffer). The suspension was filtered through a 50μm filter, centrifuged and resuspended in 3 ml of cell buffer. Red blood cells were removed using the ammonium chloride solution: 5 ml of 0.8% Ammonium Chloride was added to the suspension and incubated for 10 min with intermittent mixing. Afterwards cells were diluted with 7 ml of cell buffer, centrifuged and resuspended in 1 ml of cell buffer. Subsequently, lineage depletion was performed as follows: added 20 μl of the EasySep mouse hematopoietic progenitor cell isolation cocktail, incubated for 15 min, added 30 μl magnetic particles, incubated for 10 min, added 1.5 ml of cell buffer and placed tubes in a magnet, incubated for 3 min at room temperature and eluted cells twice (with additional 2.5 ml of cell buffer). Afterwards, cells were centrifuged, resuspended in 200 μl of cell buffer and stained with the antibody panel as follows: antibody mix was added, cells were incubated for 30 min, washed with 2 ml of cell buffer, centrifuged, resuspended in 200 μl cell buffer. For the secondary staining Streptavidin-BV510 was added, cells were washed with 2 ml of cell buffer, centrifuged, and resuspended in 1000 μl of cell buffer supplemented with 7AAD. Afterwards cells were sorted with BD influx sorter into either 96 well plates containing 2.3 μl lysis buffer (for the Smart-Seq2 protocol) or 100 μl of PBS with 0.04% BSA in eppendorf tubes ('droplet buffer') when used for the 10x Genomics scRNA-Seq protocol. The Smart-Seq2 plates were vortexed, centrifuged at 800g for 2 min and stored at -80°C.

Both Tom^+ or Tom^- cells within the Lin^- (cKit OR Sca1)$^+$ gate were sorted. (cKit OR Sca1)$^+$ is a superset of the cKit+ gate used previously[9] which contains more lymphoid progenitors and pDCs.

Hoxb5-mKO2 experiments

All steps in this section (unless otherwise indicated) were performed on ice, and centrifugation steps performed at 500g, 4°C for 5 min. 8-12 weeks old mice carrying the Hoxb5-mKO2 reporter were sacrificed and cells were isolated from BM (femurs and tibia) by grinding with mortar and pestle in PBS supplemented with 2% Fetal Bovine Serum (cell buffer). Cells were stained as described under Flow cytometry section for analysis of HSPCs. Cell suspension was filtered as above and sorted with BD influx sorter into 96-well plates containing 2.3 μl lysis buffer (for the Smart-Seq2 protocol). The Smart-Seq2 plates were vortexed, centrifuged at 800g for 2 min and stored at -80°C. The isolated populations were Lin^-, $Sca1^+$, $cKit^+$, $CD48^-$, $CD150^-$ (MPPs) and Lin, $Sca1^+$, $cKit^+$, $CD48^-$, $CD150^+$ (HSCs).

scRNA-seq data generation

Smart-Seq2

When cell numbers were limiting single cells were profiled with a modified version of the Smart-Seq2 protocol[59,60] rather than 10x Genomics kit. Single cells were sorted into 96-well plates with 2.3 μl lysis buffer containing 0.115 μl of SUPERase-In RNase Inhibitor at 20 U/μl concentration and 0.23 μl of 10% Triton X-100 solution, plates were vortexed and stored at -80°C. After thawing 2 μl of the annealing solution (0.1 μl of ERCC RNA Spike-In solution (1:300,000 dilution), 0.02 μl of the oligo-dT primer (100 μM stock concentration) and 1 μl of dNTP (10 mM stock concentration)) was added. The plate was incubated at 72°C for 3 min, cooled down on ice and reverse transcription was performed by adding 5.7 μl of RT buffer (0.1 μl of Maxima H minus reverse transcriptase at 200 U/μl concentration, 0.25 μl of SUPERase-In RNAse Inhibitor at 20 U/μl concentration, 2 μl of the Maxima enzyme buffer, 0.2 μl of TSO oligo at 100 μM concentration, 1.875 μl of PEG 8000 solution at 40% v/v concentration and 1.275 μl water) and incubation at 42°C for 90 min followed by incubation at 70°C for 15 min. Immediately after, cDNA was amplified by PCR by adding 1 μl of the Terra PCR Direct Polymerase (1.25 U/μl), 25 μl of the Terra PCR Direct buffer and 1 μl of the ISPCR primer (10 μM stock concentration) to a total volume of 50 μl using the following PCR conditions: 98°C for 3 min, 98°C for 15 s, 65°C for 30 s, 68°C for 4 min (21 cycles), 72°C for 10 min. The amplified cDNA was purified using AMPure XP beads, quantified using the PicoGreen assay (ThermoFischer P7589) and used for Nextera library preparation. The libraries were generated using either a standard protocol (batch 7d and mKO2 data) or modified protocol (batches 3d7d, 2w4w and 3dr2, see the corresponding metadata) described below. No obvious batch effects were observed among cells analyzed with either of the protocols.

The standard Nextera protocol: cDNA was diluted to approximately 50-150 pg/μl and 1.25 μl of the solution was used, 2.5 μl of Tagment DNA buffer 1.25 μl of Amplicon Tagment Mix (Nextera XT kit) were added, samples were incubated at 55°C for 10 min, and the reaction was stopped by addition of 1.25 μl of NT buffer. Tagmentation products were amplified by PCR by adding 1.25 μl of each N and S primers and 3.75 μl of NPM solution and using the following thermocycler settings: 72°C 3 min, 95°C 30 s, 12 cycles of 95°C 30s, 55°C 30s, 72°C 60s and a final extension at 72°C for 5 min.

The modified Nextera protocol follows the same principle as the standard Nextera protocol and includes the following steps: cDNA was diluted to approximately 50-150 pg/μl and 1.03 μl of the solution was used, 1.63 μl of Tagment DNA buffer and 0.6 μl Amplicon Tagment Mix was added, samples were incubated at 55°C for 10 min, the reaction was stopped by adding 0.82 μl of NT buffer. Tagmentation products were amplified by adding 1.23 μl of each N and S primers (as above but diluted 5 times), 2.3 μl of Phusion HF buffer, 0.1 μl of dNTP (25 mM stock concentration), 0.07 μl of Phusion polymerase and 2.5 μl of water and using the following thermocycler settings: 72°C 3 min, 98°C 3 min s, 12 cycles of 98°C 10s, 55°C 30s, 72°C 30s and a final extension at 72°C for 5 min.

Libraries were sequenced using the Illumina Hiseq4000 or NovaSeq instruments, obtaining an average of 1, 271,307 reads per cell.

10X genomics

For the 10x Genomics scRNA-Seq protocol up to 20,000 cells were pooled in pairs corresponding to male and female animals, centrifuged and resuspended in a volume of droplet buffer optimal for recovery of up to 10,000 cells and immediately processed with the 10x Genomics Single Cell 3′ v3 protocol following the manufacturer's instructions.

Libraries were sequenced using the Illumina NovaSeq instrument, obtaining at least 20,000 reads per cell in each run (33,843 reads per cell total average).

scRNA-Seq data analysis

Smart-Seq2 sequencing reads were aligned to the mouse genome (mm10) using the STAR aligner (version 2.7.3a) with default parameters. Reads mapping to exons were counted with featureCounts (version 2.0.0) using the ENSEMBL v93 annotation. Each cell was subjected to a quality control, cells with: <100,000 reads, <23% of reads mapped to exons, >8.5% of reads mapped to ERCC transcripts, >10% mitochondrial reads or <2000 genes detected above 10 counts per million were discarded. 1288 out of 1533 cells passed quality control. Data were normalized 10,000 total counts and ln(n+1) transformed.

10x genomics reads were pre-processed using cellranger (version 3.1.0, reference genome and annotation version 3.0.0) with default settings. Downstream analysis was performed mainly using the scanpy[61] framework with additional packages where indicated. Low quality barcodes with less than 1000 genes were excluded from the analysis, doublet scores were estimated using the scrublet tool (using 30 principal components), potential doublets were removed. Male and female cells were distinguished based on the expression of the Xist gene and Y chromosome genes. Cells with detectable Xist expression and undetectable Y chromosome gene expression were classified as female and vice versa, ambiguous cells or potential doublets were excluded. Data were normalised to 10,000 total counts and ln(n+1) transformed.

To determine highly variable genes, scanpy's highly_variable_genes function was used to select top 5000 genes within the 10x genomics data. From the list of highly variable genes, genes associated with cell cycle, Y-chromosome genes and the Xist were excluded. Genes associated with cell cycle were a union of cell-cycle genes from[9] and genes with at least 0.1 Pearson correlation with the following gene set: Ube2c, Hmgb2, Hmgn2, Tuba1b, Ccnb1, Tubb5, Top2a, Tubb4b, following previously established method.[12] Putative cell cycle phase was assigned using scanpy's 'score genes cell cycle' function to assign putative cell cycle phase to both 10x and Smart-Seq2 cells. Following that, 10x and Smart-Seq2 data were combined and subjected to Seurat CCA batch correction.[62] Among a variety of batch correction tools (Harmony,[63] Scanorama,[64] BBKNN,[65] fastMNN,[66] MNNcorrect) only Seurat CCA generated seamless integration best matching the cell frequencies based on flow cytometry analysis. After applying batch correction, we observed no obvious segregation of Smart-Seq2 and 10x scRNA-Seq profiles (Figure S2E). Corrected log-normalized counts were scaled and used to compute 50 principal components, find nearest neighbors and calculate a UMAP projection.[67] A minor batch effect between 10x samples was corrected using Harmony batch correction tool.[63] The corrected principal components were used to calculate 12 neighbors followed by cell clustering using the leiden algorithm[68] and calculation of the UMAP projection. Clusters were manually annotated based on the marker gene expression as described in Table S1. To reduce the complexity for the discrete model clusters with the following criteria were excluded from the further analysis: clusters that appeared disjointed from the main landscape body, represented low-quality/dying cells or with unclear origins based on the UMAP projection and PAGA analysis. This included: T cells, innate lymphoid cells (ILCs), cells with high mitochondrial gene counts, mature B cells, interferon-activated cells, cells with high complement expression and small clusters with unclear annotation, likely to represent doublet cells. Unfiltered landscape is displayed in Figure S2G.

To visualize the relative proportions of cells per cluster over time (Figure S4A), we averaged fractions of Tom^{+} cells in each cluster for each time-point and divided by the respective values for matching Tom^{-} cells.

mKO2 cell analysis

Smart-Seq2 sequencing reads were aligned to the mouse genome (mm10) using the STAR aligner (version 2.7.3a) with default parameters. Reads mapping to exons were counted with featureCounts (version 2.0.0) using the ENSEMBL v93 annotation. Cells with: <100,000 reads, <10% of reads mapped to exons, >10% of reads mapped to ERCC transcripts, >10% mitochondrial reads. 374 out of 384 cells passed quality control. Counts were normalized using the scran package in R and ln(n+1) transformed. Log-normalized counts were used to generate the corresponding violin plots, compute HSC-scores[17] and the projections on the[7] landscape. Particularly, the projections were performed within the scanpy module: log-normalized counts of the mKO2 experiment and of the published datasets were combined, subsetted to highly variable genes, and scaled. 50 PCs were then computed and corrected with the mnn_correct package.[66] Adjacency scores were determined based on the fraction of cells in the reference landscape that are neighbours of the cells to be projected according to the euclidean metric (method adapted from Dahlin et al.[9]).

Subclustering of cluster 0

To verify whether the HSC tip population has a constant labeling frequency, we subset cluster 0 from our landscape. We then focussed on 10x data only, to avoid artefacts deriving from the integration of different data types when it comes to very high detail. We then subclustered cluster 0 cells with higher resolution (1.3) of the Leiden algorithm. Among these subclusters, we identified the subcluster that has the highest HSC score[17] as the putative cluster 0a.

Embedding external datasets into the integrated HSPC landscape

For each external datasets the log-normalised counts for cells passing quality control were used as in the original work. Annotation was either obtained from the respective GEO repositories, literature or kindly provided by the authors.

Each dataset was integrated with the HSPC landscape (below denoted as reference) using the indicated batch correction tools and the Cellproject package as follows. Log-normalized counts for[7] were concatenated with the reference and batch effect was removed using Seurat CCA method[62] only highly-variable genes selected in the reference landscape were used. The corrected values were scaled and used to compute PCA (50 components) in the reference dataset. The correct values of Nestorowa et al.[7] dataset were fit into the reference PCA space, in which 15 nearest neighbors were identified between the datasets. These nearest neighbors were used for two purposes: (1) transfer the cluster identity to the new data (based on the most frequent label) and (2) to predict coordinates in the original reference PCA space (used as a basis for UMAP projection) using nearest-neighbor regression. Finally, the new PCA coordinates were used to embed the new data into UMAP space. As immunophenotypic populations we used the 'narrow' classification provided in the original study.

Bowling et al.[20] data was concatenated with the reference and a common PCA space was calculated, which was subsequently corrected with the Harmony batch correction tool. Within the corrected space 8 nearest neighbors were identified across the datasets, followed by label transfer and UMAP embedding as described above.

Weinreb et al.[12] and Upadhaya et al.[44] data were integrated analogously to the Nestorowa et al.[7] data. For Figures 3F, S3C only 'state-fate' clones were used, ie. cells captured at an early time-point (day2) with measured fate outcomes at later time-points. Only fates with more than 7 cells were considered for the analysis. To enable model predictions (Figures S7, E12-14) all cells and time-point were integrated using the same method.

Trajectory inference and selection

To pinpoint the most immature stem cells the HSC score was calculated (default parameters)[17] and denoised by averaging values over the nearest neighbors for each cell. As diffusion pseudotime the cell with the highest smoothed HSC score was selected, diffusion map was calculated and served as the basis for trajectory inference and continuous populations models described below.

To infer putative trajectories Tom^+ cells were used (matching the Pseudodynamics analysis below) for calculating cell transition probabilities using the Pseudotime Kernel method (based on the Palantir tool[33]) from the CellRank package.[32] To define the end states clusters 6, 7, 10, 11, 13, 14, 15, 16, 17, 18, 19 were selected and within them 50 cells with the highest pseudotime values. These states are largely consistent with an unsupervised method of macrostate selection Generalized Perron Cluster Analysis with Schur decomposition.[32] To assign cell fate probabilities Cellrank's compute_absorption_probabilities function was used.

Cells belonging to trajectories for the continuous models were selected as follows. In case of megakaryocytic trajectory cells belonging to cluster 0, 7 and 8 and with the respective fate probability >0.3 were chosen. For the erythroid trajectory cells with respective fate probability <0.2 and falling within the pseudotime range 0.015 and 0.294 (to exclude variable small number at the end of the trajectory) were used. Neutrophil and monocyte share a long stretch of progenitors with high probabilities towards both lineages, thus a different approach was used, motivated the apparent locations of bipotent cells with neutrophil and monocyte/DC potential based on cell fate assays (Figure 3F).[12] Neutrophil progenitors (terminal state 10) were selected with fate probability >0.24 and Mono/DC probability <0.38 and excluding a small number of cells falling into clusters 12, 17 and 14. Conversely for the Mono/DC progenitors (terminal state 6) cells were selected with Mono/DC fate probability >0.18 and neutrophil probability <0.49 and a small number of cells falling into clusters 12, 17 and 14 was excluded.

Differential expression analysis

For the DE analysis cells were selected to match the continuous model trajectories. The shapes of differentiation and net proliferation rates were inspected for potential regions of interests and respective ranges of pseudotime values were chosen. Prior to the analysis genes with low expression were filtered out, only genes detected in more than 2.5% cells and with overall mean expression above 0.05 (data normalized with logNormCounts from the scuttle package) were included. To select genes with dynamic expression in the chosen intervals the fitGAM function followed by startVsEndTest from the TradeSeq package were used. Genes were considered significant if they showed at least FDR of 0.1 and a $\log_2$(Fold change) of at least 1. Predicted and smoothed gene expression was used using the predictSmooth function from the same package. In heatmaps genes were clustered with hierarchical clustering using the hclust R function with default settings. Transcription factors were selected based on the gene list established in,[69] TF groups were established by cutting the tree at the level of 4. Gene enrichment was performed using GSEAPY interface to the enrichr tool.[70] The first derivative of the differentiation rate was calculated using interpolation at the same pseudotime points that were used to predict gene expression using the TradeSeq model described above.

QUANTIFICATION AND STATISTICAL ANALYSIS

Flow cytometry data analysis

Flow cytometry data was analysed using the FlowJo software followed by Prism 6 software for statistical testing and visualization. Statistical tests, relevant distributions parameters (means and standard errors of the mean) and replicate numbers are provided in the respective figures and figure legends. Effects exceeding p-values 0.05 and/or confidence interval of 95% were considered significant.

Discrete population model analysis

As input to the discrete models the estimated total number of Tom^+ or Tom^- cells per cluster was used (Table S2). The numbers were estimated based on the fraction of cells assigned to each cluster adjusted by the total number of cells (based on the flow cytometry analysis of the entire sample). One out of 5 mice analyzed at day 3 exhibited abnormally high labeling frequency, the sample was excluded to avoid introducing bias but we provide the corresponding data within the GEO submission files and source code for individual assessment.

To assess the kinetics of differentiation and growth of the different hematopoietic populations, we first considered a discrete compartments model, using the HSPC landscape clusters as compartments. To establish the available differentiation pathway, PAGA connections and pseudotime ordering were considered. We used a relaxed lenient PAGA connectivity threshold of 0.05 preserving the majority of connections between 'adjacent' clusters, consistent with the hematopoiesis models.[10,12,14,21] The relaxed PAGA connectivity threshold of 0.05 This reduced the number of model parameters and prohibited 'jumps' between distant states (e.g. from HSC to neutrophil progenitor directly), in line with the common assumptions of trajectory inference methods. Beyond identifying putative transitions the connectivity weights do not feed into our dynamics models. Furthermore, no back differentiation (ie. against pseudotime ordering) was permitted into cluster 0 and from most differentiated clusters with clear expression of commitment genes: 1, 3, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19. Other transitions above-threshold were considered potentially bidirectional. Each compartment is assigned a growth rate and as many differentiation rates as the number of its progeny compartments. Assuming the following:

- the label is neutral and stably propagated
- the kinetics parameters of each cluster are constant over time and independent of the size of any cluster
- the labeled and unlabeled cells have identical kinetics,

Population dynamics can be modeled as an ODE system of coupled equations:

$$\dot{x_i}(t) = \left(\beta_i - \sum_{j=1}^{nc} \alpha_{ij}\right) x_i(t) + \sum_{j=1}^{nc} \alpha_{j,i}\, x_j(t)$$

where $x_i(t)$ is the number of cells in population i, $\alpha_{j,i}$ is the differentiation rate from compartment j to i, and β_i the growth rate of population i. For the terminal and initial clusters, the equations take form respectively:

$$\dot{x_i}(t) = \beta_i\, x_i(t) + \sum_{j=1}^{nc} \alpha_{j,i}\, x_j(t)$$

$$\dot{x_0}(t) = \left(\beta_0 - \sum_{j=1}^{nc} \alpha_{0j}\right) x_0(t)$$

Please note that differentiation rates are set to zero if they have not passed the thresholding criteria as explained above. The differentiation rates were allowed to vary between 0 and 4 per day, with the exception of cluster 0a's rates, which were bounded to vary between 0 and 0.02 per day, based on previous knowledge of HSCs low activity.[19,42] The growth rates were bounded between -4 and 4 per day, to allow for death rate (negative values) or additional differentiation towards more mature cell states outside the presented HSPC landscape, or cell migration. The number of clusters, nc, is equal to 22, one per each of the 20 Leiden clusters, plus 2 additional subpopulations within cluster 0, the most immature cluster. The reason for this choice lays in 2 observed characteristics in the data: cluster 0 ratio of labelled to unlabeled cells (labelling frequency) grows over time, and some downstream clusters' labeling frequency overshoots the one in cluster 0. Based on Barile et al.[19] and Takahashi et al.,[22] this implies that the progenitor cluster must be heterogeneous. Indeed, the most immature HSCs occupy only the tip of cluster 0 (Figure 3C).

Particularly, we chose to add 2 more sub-compartments to allow for differentiation bias in the HSCs.[19,42] The growth rate in the most immature subcluster 0a was fixed in such a way to balance the differentiation rates, given the a priori knowledge that pure functional hematopoietic stem cells show only limited growth over time. The proliferation estimates range, we chose from one division per 145 days to in 50 days.[3,19,27,42] We accounted for this upon modelling cluster 0 overall number of cells with a logistic function, and thus added a logistic parameter ρ and a carrying capacity K. Both parameters are positive and unconstrained. Specifically, we implemented the following equations for cluster 0a:

$$\dot{x_0}(t) = \rho\, x_0(1 - x_0(t)/K)$$

$$\dot{x}_{0a}(t) = \dot{x_0}(t) - \dot{x}_{0b}(t) - \dot{x}_{0c}(t),$$

while the time evolution of clusters 0b and 0c is analogous to that of all other clusters. Since we calibrated the ODE system to both the labelled and unlabeled cells time courses, we also included as parameters 22*2 initial conditions (corresponding to labelling frequencies and cluster sizes), all positive and unbounded, except for the number of cells in cluster 0a, set to range between 500 and

1500 based on previous HSC number estimates[71] and factoring in cell isolation efficiency. The model allows the initial number of labelled cells to be greater than zero, thus accounting for any unspecific labeling.

We calibrated our model to 4 types of observables:

- The number of labeled cells in each cluster over time and relative to cluster 0 as computed via scRNA-Seq analysis
- The number of unlabeled cells in each cluster over time and relative to cluster 0 as computed via scRNA-Seq analysis
- The number of labeled cells in cluster 0 over time as computed via FACS sorting and scRNA-Seq analysis
- The number of unlabeled cells in cluster 0 over time as computed via FACS sorting and scRNA-Seq analysis

To estimate the parameters, we minimized a cost function of the squared sum of residuals. Each residual is weighted by the squared error, which was computed as pooled variance per time course. We computed the 95% confidence bounds on the parameters' best fit with the profile likelihood method as in.[19,72] To compute error bounds on the model, we ran ≈4000 bootstrap simulations, where data is resampled with replacement per time-point, and the cost function is re-minimized on the new dataset. For each simulation, a new parameter vector is found, and a model curve generated. 95% bootstrap confidence bounds are then determined cutting upper and lower 0.025 quantiles per time-point.

The bi-phasic model was generated analogously for data split into to the recovery phase (days 3-27) and the homeostasis phase (remaining time-points). We observe vast majority of changes in Tom^- cell abundance within the first 12 days, thus we conservatively chose day 27 as a boundary.

To simulate the ablation of any population, the initial condition of the unlabeled cells for the corresponding compartment can be set to 0. To ablate the HSCs, we simultaneously set to 0 the initial condition of all 3 subclusters.

To compute the journey times, we generated the model in the time interval 1–300 days with 1 day steps, assuming that cells are initially only in cluster 0 and with the unlabeled cells initial condition. We then computed the smallest time for which the number of cells in a population reaches one and dubbed that journey time.

Generalized model for testing alternative topologies

As explained in the main text and in the above section of the methods, our model constrains the topology based on the PAGA-predicted edges. In principle, though, one could test any topology, including backwards differentiation and unlikely connections such as HSC differentiating directly into a terminal compartment. We have thus implemented an additional code where the user can test the performance of any model upon setting to 1 the entries of a 22 X 22 table representing the existence of a differentiation rate from any cluster to any cluster.

Model selection for perturbed systems

To infer what parameters may change in non-homeostatic conditions, we developed a model selection-based method. We first fixed the parameters describing the challenged system to our best fit, and then allowed the parameters of specific populations to change. We considered 14 populations whose proliferation and differentiation rates may change, being 14 out of 20 the populations that have at least one progeny in the 'challenge' dataset. Out of these 14 populations, any subgroup may change its parameters or not, for a total of 2^{14} = 16384 models. These models were all fit to the transplantation data. In order to rank these models, we employed the Akaike information criterion, and retained only those models that simultaneously have the lowest possible number of populations whose parameters change in order to fit the data and whose corrected Akaike index is not greater than the best ranking Akaike index plus 10.

Continuous population model analysis

In order to compute pseudotime-dependent kinetic rates, we relied on the pseudodynamics framework.[31] Briefly, the compartment model explained in the previous section has a one to one correspondence to the continuous model if the compartment index is treated as a continuous variable, namely the diffusion pseudotime coordinate s, the number of cells is replaced by the cell density over pseudotime and real time $u(s,t)$, and the differentiation and net proliferation rates are replaced by the drift $v(s)$ and the growth rate $g(s)$, respectively. Given these substitutions, the ODE system becomes a PDE system. In addition, the Pseudodynamics framework also introduced an extra parameter $D(s)$ that allows for diffusion of the cells on the pseudotime axis to account for stochasticity in the differentiation process. The 3 kinetics parameters, drift, growth rate and diffusion, are modeled as natural cubic splines with 9 nodes. The nodes boundaries were kept as in the original publication: between 0 and 1 per day for drift and diffusion, and between -5 and 6 per day for the growth rate. To simplify the computation, we estimated such rates independently for 4 different trajectories, which avoids introducing parameters that describe the branching process. The trajectories were chosen based on the affinity to each terminal state as estimated by CellRank (see section 'Trajectory inference and selection'). For each trajectory, the PDE reads:

$$\frac{\partial u(s,t)}{\partial t} = \frac{\partial}{\partial s}\left(D(s)\,\frac{\partial u(s,t)}{\partial s}\right) - \frac{\partial}{\partial s}\left(v(s)\,u(s,t)\right) + g(s)\,u(s,t)$$

For the boundaries, we assumed no-flux Robin conditions, as in the original publication. To solve the PDE, we used the non-branching pseudodynamics model as compiled in MATLAB 2017b, with only one difference: we did not enforce differentiation to be 0 at the end of the trajectory which, together with the growth rates taking also negative values, accounts for the fact that the

populations in our landscape are all transient and that fully mature cells are not captured by our gating strategy. The model was calibrated to the time-dependent density and total number of labelled cells only. The error was computed as variance among replicates. For each trajectory, at least 240 simulations were launched, with regularization parameters 0, 1, or 10 to penalize big differences in the splines' nodes. The solution was chosen based on the highest log-likelihood, and the regularization parameter as the highest that visually fits the data well.

Transplantation data analysis

Dong et al. data[45] was integrated into the HSPC landscape analogously to the[7] data integration described in section 'Embedding external datasets into the integrated HSPC landscape'. Cells in each HSPC cluster were counted and used as an input into the discrete model prediction. Day 3 data was used as the initial condition and cell abundances per cluster were predicted from day 3 to day 7. The bootstrap confidence bounds were recomputed upon substituting the initial conditions. Given that the experimental data in relevant clusters vastly exceed the model prediction bounds, we concluded that the dynamics of perturbed hematopoiesis are different from normal conditions and suggest increased differentiation.