Dntt expression reveals developmental hierarchy and lineage specification of hematopoietic progenitors

Fabian Klein1, Julien Roux1,2, Grozdan Cvijetic1,3, Patrick Fernandes Rodrigues1, Lilly von Muenchow1,4, Ruth Lubin5, Pawel Pelczar6, Simon Yona5, Panagiotis Tsapogas1 and Roxane Tussiwand1,3✉

Intrinsic and extrinsic cues determine developmental trajectories of hematopoietic stem cells (HSCs) towards erythroid, myeloid and lymphoid lineages. Using two newly generated transgenic mice that report and trace the expression of terminal deoxynucleotidyl transferase (TdT), transient induction of TdT was detected on a newly identified multipotent progenitor (MPP) subset that lacked self-renewal capacity but maintained multilineage differentiation potential. TdT induction on MPPs reflected a transcriptionally dynamic but uncommitted stage, characterized by low expression of lineage-associated genes. Single-cell CITEd-seq indicated that multipotency in the TdT+ MPPs is associated with expression of the endothelial cell adhesion molecule ESAM. Stable and progressive upregulation of TdT defined the lymphoid developmental trajectory. Collectively, we here identify a new multipotent progenitor within the MPP4 compartment. Specification and commitment are defined by downregulation of ESAM which marks the progressive loss of alternative fates along all lineages.

Most blood cells have a short half-life and are regenerated throughout the life of an individual in a process referred to as hematopoiesis. Hematopoietic stem cells (HSCs) reside within the bone marrow (BM) at specific niches that provide the necessary cues for their maintenance and survival. Through proliferation and differentiation, the pool of HSCs is constantly self-renewed while it generates progeny that progressively expand, giving rise to all mature hematopoietic subsets. HSCs were originally described as Lineage−Sca1+− (LT-) and short-term (ST-)HSCs and MPPs. MPPs do not have self-renewal capacity, but will reconstitute lymphoid, myeloid and erythroid lineages. On the basis of the expression of FLT3, CD150 and CD48, the MPP compartment is currently split into erythroid-primed FLT3+CD150−MPP2s, myeloid-primed FLT3−CD48+CD150−MPP3s and lymphoid-primed FLT3+CD150−MPP4s, each characterized by developmental bias toward their respective lineages. However, the extent of their heterogeneity and plasticity and the stage at which lineage commitment becomes irreversible remain elusive. To dissect lineage restriction and specification at their earliest point along the lymphoid branch, we generated mouse models that directly report or trace the expression of the lymphoid-specific template-independent polymerase Dntt (encoding TdT), which is required for the insertion of random nucleotides at VDJ joining regions during B and T cell receptor rearrangement. TdT tracing surprisingly showed a broad expression profile, labeling all hematopoietic lineages. Using computational analysis on single-cell cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq), along with multiple functional assays, in these newly generated mouse lines, we resolved and re-defined early hematopoietic development from its most uncommitted precursor within the MPP4 subset up to each specific stage at which lymphoid, myeloid and erythroid lineage restriction occurred.

TdT labeling marks early T and B cell development

To isolate TdT-expressing cells and identify their progeny, we generated a TdT-reporter and a TdT-fate-mapping line (Extended Data Fig. 1a). The reporter line (hereafter TdT(iCre)) was constructed by inserting the self-cleaving peptide P2A followed by the extracellular domain of the human CD4 gene (hCD4), ensuring equimolar expression and surface detection of TdT. Similarly, TdT-fate mapping was achieved using P2A-iCre crossed to the Rosa26LSL-YFP line (hereafter TdT(YFP)). Faithful reporting of TdT by hCD4 was confirmed by quantitative PCR analysis (Extended Data Fig. 1b) of sorted LSK subsets using primers for Dntt, hCD4 and iCre or spanning the junctional regions, as well as by hCD4 and intracellular TdT co-staining (Extended Data Fig. 1c,d).

Expression of TdT, which inserts random nucleotides at junctional regions, is thought to be initiated on common lymphoid progenitors (CLPs) and maintained on T and B lymphocytes until cells have rearranged their corresponding T and B cell receptors. hCD4 expression is detected in developing T cells from CD4−CD8 double-negative (DN) to double-positive precursors (Fig. 1a,b). Consistently, yellow fluorescent protein (YFP) is detected starting from DN1 cells in TdT(YFP) mice (Fig. 1c). Along the B cell lineage, hCD4 in TdT(iCre) and YFP in TdT(YFP) mice were detectable in Lin−B220−CD11b+Ly6G−EPLMs (early BM progenitors with myeloid and lymphoid potential) up to CD19+Ki67 pro-B cells (Fig. 1d–g and Extended Data Fig. 1e), in agreement with the reported downregulation of Dntt expression upon rearrangement of the heavy chain. As expected, YFP expression remained high on all B cells.
cells (Fig. 1g), collectively confirming the origin of all T and B cells from a Dntt-expressing progenitor.

**TdT fate mapping labels across all hematopoietic lineages**

hCD4 expression on TdT<sup>HCD4</sup> was downmodulated on all mature cells (Extended Data Fig. 2a–d), except for plasmacytoid dendritic cells (pDCs), corroborating their lymphoid origin<sup>16,29</sup>, and a small fraction of splenic CD4<sup>+</sup>, CD8<sup>+</sup> and γδ T cells, likely representing recent thymic emigrants (Fig. 2a). Surprisingly, YFP expression was detected at different frequencies across all analyzed hematopoietic subsets, reaching 20–30% on platelets and pro-erythrocytes and 50–80% in myeloid subsets (Fig. 2b). Because P2A mediates in-frame translation of Dntt with hCD4 or iCre, we can exclude leakage and hypothesize that TdT or iCre expression occurred in a multipotent progenitor, resulting in YFP labeling across all lineages.

We therefore analyzed hCD4 and YFP expression in TdT<sup>HCD4</sup> and TdT<sup>iCre</sup> mice across all LSK cells. To remove residual lymphoid progenitors within the LSK fraction, we introduced an additional
Fig. 2 | TdT expression and tracing across mature cells and progenitors. a,b, Representative histogram plots (top) and cumulative bar graphs (bottom) showing hCD4 (platelets, n = 4; pro-erythrocytes n = 7; granulocyte, monocyte, cDC1, cDC2, pDC, NK cells, n = 6; CD4+ T cells, CD8+ T cells, γδ T cells, n = 12; B cells, n = 9) (a) and YFP (n = 6 for all subsets analyzed) (b) expression for different mature hematopoietic subsets in 6- to 8-week-old TdTCD4 reporter (a) and TdTYFP lineage tracer (b) mice. Shown are BM cells for pro-erythrocytes, blood mononuclear cells for platelets and splenocytes for all other subsets. The gating strategies to identify the corresponding cell types are shown in Extended Data Fig. 2a–d. Shown are cumulative results from three independent experiments. Error bars indicate s.e.m. c–g, BM progenitors gated as shown in c and in Extended Data Fig. 2f,g were analyzed for the expression of hCD4 (d,e) and YFP (f,g) in 6- to 8-week-old TdTCD4 reporter (d,e) and TdTYFP lineage tracer (f,g) mice. c, HSC, MPP subsets and CLPs were identified by the indicated gates. d–g, Representative histograms (left) and cumulative bar graphs (right) showing hCD4 (d,e) and YFP (f,g) expression for the indicated progenitor subsets. d, n = 12; e, n = 3; f, n = 6; g, n = 6. Shown are cumulative results from three independent experiments. Error bars indicate s.e.m.
IL-7Rγ gate and further subdivided LT- and ST-HSCs, MPP2s, MPP3s and MPP4s using CD48, CD150 and FLT3 (Fig. 2c and Extended Data Fig. 2e)15,16,29. As expected in TdTCD4 mouse, hCD4 expression in MPP4s was the highest, reaching about 80% of labeling (Fig. 2d). This percentage increased along the lymphoid branch, with almost 100% of CLPs expressing hCD4 (Fig. 2d). Across all other progenitor subsets, we observed 20% hCD4+ MPP3s and 35% hCD4+ monocyte/dendritic cell progenitors (MDPs) (Fig. 2de and Extended Data Fig. 2f). In lineage tracer TdTYFP mice, about 80% of MPP4s and all CLPs were labeled (Fig. 2f). Further, about 20–40% MPP2 megakaryocyte progenitors (MkPs) and colony-forming unit-erythrocytes (CFU-Es) were YFP+, despite all being hCD4– (Fig. 2fg), suggesting that iCre was initiated in a TdT+ progenitor upstream of these developmental stages. Along the myeloid developmental pathway, YFP labeling on mature subsets was consistent with their immediate precursors, as evidenced by the 50–70% labeling of granulocyte-monocyte progenitors (GMPs), monocyte progenitors (mMoPs), monocyte-dendritic cell progenitors (MDPs), common dendritic cell progenitors (CDPs) and their direct progeny (Fig. 2bg).

Given that all LT-HSCs were YFP-hCD4+, but YFP and hCD4 labeling could be detected across all lineages, we hypothesized that a TdT-hCD4+ progenitor existed. Since about 3% of ST-HSCs were YFP+, labeling could have occurred within this fraction. However, hCD4 that proceeded YFP labeling was evident in only a few mice and was at the limit of detection (Fig. 2d and Extended Data Fig. 2hi), while YFP was present on all analyzed mice, which excludes the possibility that labeling was initiated within this subset and suggests that YFP+ ST-HSCs were also downstream of a hCD4+ multipotent progenitor. About 20% of MPP2s expressed YFP; however, because hCD4 was almost undetectable (Fig. 2d and Extended Data Fig. 2hi), we could exclude the possibility that YFP labeling is initiated within this subset. Moreover, these results indicate that the expression of the lymphoid-specific gene Dntt was uncoupled from lymphoid lineage restriction on a fraction of MPPs and that at least two developmental pathways along the erythroid and the myeloid lineage are possible, one from a YFP− and one from a YFP+ progenitor.

**TdT**− **MPP4s are multipotent progenitors**

To understand the developmental pathways and the plasticity of the different MPP subsets in view of their TdT and YFP expression, we crossed the TdT+line with Rosa26mtmG mice (hereafter TdtmtmG). In these mice, induction of iCre excises the Tomato cassette, leading to the loss of constitutive Tomato expression with concomitant induction of green fluorescent protein (GFP). During a short window of time, cells are Tomato−GFP+, until Tomato is degraded or sufficiently diluted through proliferation. Developmental progression occurs from Tomato−GFP+ to Tomato+GFP+ and finally to Tomato−GFP− cells, enabling the earliest detection of iCre and therefore of TdT in TdtmtmG mice. LT-HSCs were exclusively Tomato−GFP+ (Fig. 3a,b and Extended Data Fig. 3a–d), confirming that they are upstream of all compartments. We then assessed Tomato expression within GFP− LSKs and observed that about 7% of MPP4s were Tomato+GFP+, which was the highest Tomato expression across all MPPs (Fig. 3a,b and Extended Data Fig. 3b–d). When back-gating on Tomato+GFP+ cells, 92% were MPP4s and about 2–3% were MPP3s (Fig. 3b and Extended Data Fig. 3b–d), suggesting that either both or one of the two subsets was responsible for the multilineage labeling. No Tomato+GFP− cells were detected within the MPP2 or ST-HSC gates (Fig. 3b and Extended Data Fig. 3b–d), corroborating that YFP labeling was not initiated within either subset, but rather GFP+ MPPs and ST-HSCs must have differentiated from Tomato+GFP+ MPP4s or MPP3s, where iCre expression was initiated.

To validate that multilineage potential was present within GFP− MPP3s and/or GFP− MPP4s in TdtmtmG mice or YFP+ MPP3s and/or YFP+ MPP4s in TdtYFP mice, we assessed their in vitro and in vivo differentiation potential. We established B and myeloid precursor frequency using limiting dilutions to directly compare YFP− and YFP+ MPP2s, MPP3s and MPP4s isolated from TdtmtmG mice. In contrast to previous reports15,16,19,29, B cell potential was confined to MPP4s, with higher precursor frequency for YFP− cells (Fig. 3c). The exclusion of IL-7Rγ LSKs using an additional gate likely removed residual lymphoid precursors from the MPP3 and MPP2 fractions (Extended Data Fig. 3e). Both YFP− and YFP+ MPP3s had myeloid potential (Fig. 3c), as previously reported20,19,29. Notably, myeloid precursors were also present in MPP4s, with YFP− MPP4s showing comparable frequency to myeloid-biased MPP3 subsets (Fig. 3c), suggesting a possible superior multilineage potential compared with that of other MPPs. YFP− and YFP+ MPP2s had limited but consistent in vitro myeloid potential (Fig. 3c), as previously shown19,29. To assess in vivo reconstitution potential across all hematopoietic branches, including platelets and erythrocytes, we used TdtmtmG instead of TdtYFP mice, in which Tomato traces pro-erythro- and megakaryocyte lineages. We transferred individual GFP− and GFP+ MPP2s, MPP3s or MPP4s into sublethally irradiated CD45.1 congenic mice and monitored the progeny of these cells, independently of GFP expression, every week for 4 weeks (Extended Data Fig. 3f–h). GFP− MPPs had an overall higher and broader reconstitution potential (Fig. 3d,e) than did their GFP+ counterparts. Within individual MPP subsets, GFP− MPP2s were mostly restricted to the megakaryocyte lineage, whereas GFP− MPP2s also generated myeloid progeny (Fig. 3d,e). Similarly, GFP− MPP3s were overall more efficient than their GFP+ counterparts at reconstituting the erythro-myeloid compartment (Fig. 3d,e). Independently of GFP
expression, both MPP2s and MPP3s lacked B cell potential in vivo, validating the above-mentioned results that were obtained in vitro. Only GFP− MPP4s showed multipotency, giving rise to all three lineages, erythroid, myeloid and lymphoid, whereas GFP+ MPP4s had no erythroid–megakaryocyte potential (Fig. 3d,e), suggesting that acquisition of GFP or YFP on MPP2s and MPP4s led to the extinction of their myeloid or platelet potential, respectively. As such, expression of TdT or GFP in TdTmTmG or YFP in TdTYFP mice marked the first step of lineage restriction.

Reconstitution of short-lived myeloid cells and pro-erythrocytes was maintained beyond 4 weeks post-transplantation from only GFP− MPP4s (Fig. 3d,e), suggesting that this subset was upstream of...
Fig. 4 | CITE-seq reveals heterogeneity and lineage bias within LSKs. LSK cells isolated from the BM of four 6- to 8-week-old TdThCD4/YFP double reporter mice were used for scRNA-seq in combination with CITE-seq as described in Methods. 

(a–c), Hierarchical clustering analysis was performed on 15,853 LSKs and results were projected in a 2D space using UMAP. Each color represents a specific cluster. 

(a), Hierarchical clustering identified eight clusters. 

(b), Cells were annotated on the basis of CITE-seq antibody labeling, and we applied an a posteriori gating strategy, as shown in Extended Data Fig. 4e. 

(c), Cells were annotated on the basis of transcriptional similarity to the ImmGen reference data set and annotated accordingly using the ImmGen subset definition applied for sorting. In (b) and (c), Cells that could not be assigned to a subset or a reference population are not shown. Contour lines display the 2D cell density obtained for (a) on the UMAP space. 

(d), Heatmap displaying the centered and scaled expression level of the top 30 markers upregulated in each cluster, defined as in (a). Shown are 104 markers. Cells were ordered following the hierarchical clustering tree. 

(e), Bar graphs showing the percentage of cells belonging to each cluster defined as in (a) and distributed according to the a posteriori gating strategy used in (b) or the ImmGen annotation as defined in (c). Compiled data showing the cluster distribution, defined as in (a), in relation to the similarity score to reference samples from the ImmGen dataset, defined as in (c), or the cell-type annotation based on the a posteriori gating strategy defined as in (b). The expression of CITE-seq markers, YFP and hCD4 expression are centered to the mean and scaled to the range of expression values.
all other MPPs and is possibly related to HSCs. GFP+ MPP4s constituted not only mature subsets across all lineages, but also all MPP subsets 2 and 4 weeks after transfer (Fig. 3f), while neither GFP+ MPP4s nor MPP3s and MPP2s, regardless of their GFP expression, gave rise to MPPs (Extended Data Fig. 3i).

To determine the long-term potential of YFP− MPP4s, we co-transferred them with equal numbers of CD45.1/2 LT-HSCs (Extended Data Fig. 3j) or ST-HSCs (Extended Data Fig. 3k) into CD45.1 congenic mice. As shown, myeloid progeny, which is devoid of self-renewal capacity, derived from YFP− MPP4s is extinguished after 4 weeks, suggesting multilineage potential but lack of self-renewal capacity.

**Single-cell profiling of MPP subsets reveals heterogeneity**

To define the heterogeneity within the MPP and HSC compartments, we used single-cell RNA sequencing, including CIT-eseq30, of LSKs isolated from TdThCD45.1 crossed with TdThYFP (hereafter TdThCD45.1/YFP). For the CITE-seq, we used oligonucleotide-coupled antibodies targeting hCD4, CD135 (FLT3), CD48 and CD150, to be able to back-gate on MPP subsets, and markers known to be expressed on progenitors, including CD9, CD41, CD55, CD105, CD115, CXCR4 and ESAM. Across 4 biological replicates, 15,853 LSK cells were retained, displaying an average of 3,999 detected genes/cell after filtering out proliferating cells and low-quality cells.

We performed a classical cell-type annotation based on the transcriptome similarity of each cell to reference bulk RNA-seq samples from the ImmGen platform (http://www.immgen.org) and from a progenitor-specific collection in ref. 12. Gated HSCs and MPPs were projected into the UMAP space (Fig. 4b), and their cluster distribution was analyzed (Fig. 4b), allowing us to perform a direct comparison of the transcriptional profiles with the functional data obtained above (Extended Data Fig. 4g).

This gating strategy confirmed that the excluded clusters of proliferating cells were enriched for the MPP2 and MPP3 subsets and were depleted of HSCs (Extended Data Fig. 4d). Further, to relate our work to previously published datasets, we performed a classical cell-type annotation based on the transcriptome similarity of each cell to reference bulk RNA-seq samples from the ImmGen platform (http://www.immgen.org) and from a progenitor-specific collection in ref. 12. Gated HSCs and MPPs were projected into the UMAP space (Fig. 4b). Using CD48, CD150 and CD135 together with the two-dimensional (2D) space (Fig. 4a), we were able to back-gate on MPP subsets, and markers known to be expressed on progenitors, including CD9, CD41, CD55, CD105, CD115, CXCR4 and ESAM. Across 4 biological replicates, 15,853 LSK cells were retained, displaying an average of 3,999 detected genes/cell after filtering out proliferating cells and low-quality cells.

The clustering analysis based on single-cell RNA sequencing (scRNA-seq) only partially overlapped with the analysis using the gated subsets (Fig. 4e and Extended Data Fig. 4g) or the ImmGen assignment (Fig. 4f,g). It confirmed the similarity of clusters 8 and 5 to LT- and ST-HSCs, but also revealed that a significant portion of MPP4s included cells belonging to these HSC-representing clusters (Fig. 4e and Extended Data Fig. 4g). Gated MPP2s were mostly represented by the transcriptional clusters 6 and 7, both highly enriched in erythroid-megakaryocyte transcripts, such as Gata1, Klf1, Wt1 and Pfd4 (ref. 34) (Fig. 4d–g and Extended Data Figs. 4g and 5b). However, when gated, MPP3s distributed predominantly across clusters 1, 3 and 4, revealing their transcriptional heterogeneity (Fig. 4e and Extended Data Figs. 4g and 5b). Cluster 2, which expressed lymphoid hallmark genes Ighm, Ighd, Notch1 and Lck (Fig. 4d and Extended Data Figs. 4g and 5c), contained exclusively MPP4s (Fig. 4e). However, when gated, lymphoid-biased MPP4s comprised multiple clusters besides cluster 2 (Fig. 4e and Extended Data Fig. 4g), validating their multilineage capacity. The ability of sorted MPP4s to generate myeloid progeny could be ascribed to the inclusion of clusters 3 and 4 (Fig. 4d). Similarly, the capacity of MPP4s to give rise to erythroid progeny, as well as their developing into all MPP subsets, could be explained by the presence of clusters 1, 5 and 8 (Fig. 4e and Extended Data Fig. 4g). The transcriptional profile of cluster 1, owing to its “central” position in the UMAP space, had a lineage-undefined profile (Fig. 4d), which was reflected in a mixed gating distribution (Fig. 4e,f). Collectively, this analysis showed the transcriptional heterogeneity of individual MPPs, but enabled us to identify within the gated MPP4 compartment a fraction of cells that transcriptionally aligned with HSCs.
Multilineage potential is present within MPP4s

Because YFP⁻ MPP4s had the broadest in vivo and in vitro potential, we specifically focused our computational analysis, dissecting the heterogenous expression of Dntt, hCD4 and YFP using transcript as well as CITE-seq antibody-mediated detection. Hypothetically, Dntt and hCD4 (CITE-seq) should be equally expressed; however, antibody tagging showed a higher detection (Extended Data Fig. 6a). To relate the functional data obtained above by sorting YFP⁺ and YFP⁻ MPPs, we directly compared the transcriptional profiles of each MPP subset on the basis of YFP expression (Fig. 5a–c). Within MPP2s, we did not observe any major differences in transcriptome aside from YFP (Fig. 5a). YFP⁻ MPP2s expressed slightly higher levels of CD41 (Fig. 5a), which has previously been associated with early hematopoiesis[6]. Both subsets were equally represented by clusters 6 and 7 (Fig. 5b,c), suggesting that the functional difference observed above for GFP⁺ and GFP⁻ MPP2s was not explained by...
a different cluster distribution for YFP− and YFP+ MPP2s or by a major transcriptional difference. It is, however, possible that more subtle differences may exist in the chromatin landscape. Gated MPP3s comprised clusters 1, 3 and 4; however, only YFP+ subtle differences may exist in the chromatin landscape. Gated major transcriptional difference. It is, however, possible that more transcriptionally uncommitted clusters 1 and 5 (Fig. 5c).

Within MPP4s, 3,596 DEGs characterized the YFP− and YFP+ fractions (Fig. 5a–c). YFP+ MPP4s expressed genes linked to lymphoid specification and loss of stemness (Cd48, Mpo, Ilf8, Ighm and Dnmt) and markers (Cd48 and hCd4), defined by clusters 2 and 3 (Fig. 5a–c). Consistent with the lack of erythroid potential, YFP+MPP4s had no cells from clusters 6 and 7 (Fig. 5c). YFP−MPP4s mostly contained clusters 1, 2 and 5, and a small fraction of cluster 8 (Fig. 5b,c). These results indicate that YFP− MPP4s were the most undifferentiated MPP subset and were transcriptionally characterized by a multilineage potential.

**Lineage gene induction is uncoupled from lineage restriction**

Because we had generated the heterozygous TdtCd4/YFP mice for the sequencing experiment and given that YFP expression was independent of lymphoid specification, we could computationally and functionally re-analyze all MPP subsets presuming the timeline of Dnmt expression was hCd4+ YFP+ to hCd4+ YFP− to hCd4+ YFP− to hCd4− YFP+. LT-HSCs were all hCd4+ YFP− (Fig. 5d), and 4% of ST-HSCs and 22% of MPP2s were hCd4+ YFP+ (Figs. 2i and 5d), validating the hypothesis that these cells originated from progenitors not included within these gates. MPP3s and MPP4s could be separated into four subsets on the basis of hCd4 and YFP expression (Fig. 5d). We next assessed their cluster distribution, UMAP localization and in vivo reconstitution potential. All fractions included within the MPP3 compartment contained a variable distribution of clusters 1, 3 and 4, displaying myeloid/lymphoid- and HSC-related transcripts (Extended Data Fig. 6b–d). YFP+ MPP3s correlated with higher similarity scores to HSCs, while YFP− MPP3s had higher similarity scores to the ImmGen-based MPP3 and MPP4 subsets (Extended Data Fig. 6b). In transplantation experiments, hCd4− YFP− MPP3s were the most immature, while hCd4+ YFP− MPP3s represented the most advanced population, with the lowest reconstitution capacity (Extended Data Fig. 6e). Since MPP3s were devoid of B cell potential (Extended Data Fig. 6e), we could assume that B cell precursors were only contained within the lymphoid cluster 2 or the HSC-related clusters 5 and 8.

hCd4− YFP+ MPP4s contained the uncommitted and HSC-related clusters 1, 5 and 8 (Fig. 5e and Extended Data Fig. 6f,g). Transition to hCd4+ YFP+ MPP4s was associated with an increased proportion of lymphoid cluster 2, whereas hCd4+ YFP+ MPP4s gained cluster 3 (Extended Data Fig. 6f,g), suggesting the initial induction of the lymphoid program and validating their ability to generate both myeloid and lymphoid progeny, respectively (Extended Data Fig. 6h).

Downregulation of Tdt in hCd4+ YFP− MPP4s was characterized by a re-distribution of clusters 1, 2, 3 and 5 frequencies and loss of HSC-related cluster 8 (Fig. 5e and Extended Data Fig. 6f,g). On the basis of the transplantation results that show a robust and multilineage reconstitution for hCd4+ YFP− cells (Fig. 5f and Extended Data Fig. 6h), we can hypothesize that MPP4s that remained hCd4+ YFP+ presumably continued their commitment to the lymphoid lineage (cluster 2), and a transition to the hCd4− YFP− stage reflected reversion to a more multipotent stage (clusters 1 and 5). hCd4− YFP+ MPP4s represented only a minor fraction (6%) of the YFP+ MPP4s (Fig. 5d), possibly explaining why erythro-megakaryocyte potential was not detected in GFP+ MPP4s from the Tdt+/- mice.

**Computational analysis using the Slingshot algorithm and Monocle 3**

We then validated the hypothesis that these cells originated from progenitors not included within these gates using computational methods (Fig. 5e). Shown are cumulative data from three independent experiments. Error bars indicate s.e.m.

**ESAM+ MPP4s are the only bona fide MPPs**

Among the most differentially expressed markers and genes between YFP− and YFP+ MPP4, we identified ESAM (Fig. 5a), which was previously shown to label all LT-HSCs and part of the MPP compartment (Fig. 5a and Extended Data Fig. 7a)13–15,20,37,38. In the context of UMAP projection, ESAM+ MPPs partially overlapped with YFP− MPPs from TdtCd4/YFP mice (Fig. 6b and Extended Data Fig. 7b,c). The enrichment in ESAM+ MPP4s for HSC transcripts and clusters 5 and 8 (Fig. 6c and Extended Data Fig. 7b) prompted us to test their in vivo and in vitro reconstitution potential. ESAM+ and ESAM− MPP2s, MPP3s and MPP4s were transferred into sublethally irradiated congenic Cd45.1 mice. For all subsets, downregulation of ESAM resulted in lineage restriction: ESAM−, as compared with ESAM+, MPP2s had no myeloid potential; ESAM−, as compared with ESAM+, MPP3s and MPP4s had no platelet potential (Fig. 6d). Only ESAM+ MPP4s reconstituted all lineages and all
ST-HSCs that correspond to YFP+ ST-HSCs clustered away from LT-HSCs in a t-distributed stochastic neighbor embedding distribution plot (Extended Data Fig. 7a,f), corroborating that they may not represent true stem cells. Expression of hCD4 is high on ESAM+ MPP4s, whereas ESAMhi MPP4s are hCD4+, indicating that lymphoid commitment was characterized by progressive stabilization and upregulation of the lineage-specific transcript Dntt (Fig. 6a,g and Extended Data Fig. 7e,g). Parallel to increased hCD4

ST-HSCs that correspond to YFP+ ST-HSCs clustered away from LT-HSCs in a t-distributed stochastic neighbor embedding distribution plot (Extended Data Fig. 7a,f), corroborating that they may not represent true stem cells. Expression of hCD4 is high on ESAM+ MPP4s, whereas ESAMhi MPP4s are hCD4+, indicating that lymphoid commitment was characterized by progressive stabilization and upregulation of the lineage-specific transcript Dntt (Fig. 6a,g and Extended Data Fig. 7e,g). Parallel to increased hCD4

ST-HSCs that correspond to YFP+ ST-HSCs clustered away from LT-HSCs in a t-distributed stochastic neighbor embedding distribution plot (Extended Data Fig. 7a,f), corroborating that they may not represent true stem cells. Expression of hCD4 is high on ESAM+ MPP4s, whereas ESAMhi MPP4s are hCD4+, indicating that lymphoid commitment was characterized by progressive stabilization and upregulation of the lineage-specific transcript Dntt (Fig. 6a,g and Extended Data Fig. 7e,g). Parallel to increased hCD4

ST-HSCs that correspond to YFP+ ST-HSCs clustered away from LT-HSCs in a t-distributed stochastic neighbor embedding distribution plot (Extended Data Fig. 7a,f), corroborating that they may not represent true stem cells. Expression of hCD4 is high on ESAM+ MPP4s, whereas ESAMhi MPP4s are hCD4+, indicating that lymphoid commitment was characterized by progressive stabilization and upregulation of the lineage-specific transcript Dntt (Fig. 6a,g and Extended Data Fig. 7e,g). Parallel to increased hCD4
induction was the progressive loss of erythroid potential, the reduced frequency of myeloid precursors and reduced myeloid in vivo reconstitution, but we observed increased early B cell potential (Extended Data Fig. 7b–k).

Transcriptionally, ESAM−hCD4⁺ MPP4s identified almost exclusively with the uncommitted clusters 1, 5 and 8 (Fig. 6f, g and Extended Data Fig. 7e), suggesting a profound overlap with HSCs. Similarly, ESAM−hCD4⁺ MPP4 also identified with the uncommitted clusters 1 and 5 besides the lymphoid primed cluster 2 (Fig. 6f, g), validating the observed labelling across all compartments in TdT-YFP mice. To define their potential at a clonal level, we sorted HSCs and MPP fractions (Fig. 6h), further corroborating that downregulation of ESAM mirrors lineage restriction. On the basis of these findings, we introduced a new gating strategy that considers the expression of ESAM for wild-type mice and hCD4 on TdT-YFP mice for the gating of HSCs and MPPs (Extended Data Fig. 8a–c).

Irradiation pauses the lymphoid transcriptional program

To define the changes that occur during emergency hematopoiesis, we monitored reconstitution in TdT-YFP mice after sublethal irradiation, which favors myelopoiesis. It required about 4 weeks to re-establish the steady-state frequency of YFP⁺ MPPs and YFP⁺ expressing mature subsets (Fig. 7a,b). Erythroid- and myelopoiesis had a transient shut down of the developmental pathway through YFP⁺ MPPs, suggesting that there is an overall downmodulation of transcripts related to lymphoid specification and of lymphopoiesis. Lymphoid development remained YFP⁺ (Fig. 7a) but was compromised beyond week 4 (Fig. 7b). These observations suggest that the increased erythroid and myeloid cell production after irradiation most likely occurred through the induction of environmental changes that affected early precursors and forced lineage specifications upon demand. We collectively propose a new hierarchy of early hematopoiesis at steady state and following perturbation (Extended Data Fig. 8d,e).

Discussion

Through the generation of two new mouse lines for reporting and tracing the expression of the lymphoid-specific gene Dntr, here we tracked key steps during early hematopoiesis beyond lymphopoiesis and identified a new MPP progenitor with an MPP4 profile and capable of multilineage reconstitution. Further, single-cell CITIP-seq of the LSK compartment in the dual reporter and lineage tracer mice revealed ESAM expression as the key marker for multipotency within ST-HSCs and MPP4 and for oligopotaecy in MPP2s and MPP3s.

Fate mapping, transposon and Cre-loxP-mediated barcoding systems studies have collectively contributed to our current understanding of early hematopoietic development. Clonal transposon tagging experiments revealed that, apart from HSCs, multilineage...
potential was primarily found in a fraction of MPP4s that could not be specifically identified\(^5\). It has been reported that MPP2s are capable of multipotent reconstitution\(^1\). While different multipotent progenitors have been proposed, there is general consensus that only a small fraction of HSCs generates most of the hematopoietic progeny\(^39,40\). In line with this view, we showed that only a minor fraction of HSCs was cycling and was therefore the likely source of most mature cells, potentially aligning with the recently described CD34\(^+\)CD135\(^−\)CD48\(^−\)CD150\(^−\) MPP5s (ref. \(^43\)).

Our main goal was to pinpoint progenitors at the bifurcation of lymphoid versus myeloid–erythroid lineage, which represents a major branchpoint during hematopoietic development\(^64,65\). In mice in which the expression of Dntt was traced, all hematopoietic lineages were labeled with YFP, including the erythroid-megakaryocyte and myeloid branch. We would exclude leakage on the basis of the genetic construct of the line. A detailed computational and functional analysis of hCD4- and YFP-expressing MPP2s, MPP3s and MPP4s allowed us to trace the earliest multipotent progenitors within the MPP4 compartment and to show that expression of lineage-specific genes is uncoupled from commitment. These findings reconcile with the idea that all hematopoietic cells are labeled in the transgenic line generated by crossing Flt3-Cre and Rosa26-YFP mice\(^46,47\). Transient induction of TdT led to the labeling of a small fraction of hCD4- YFP\(^+\) MPP4s that had multilineage potential, but were outcompeted by HSCs in in vivo reconstitution experiments, suggesting that they are developmentally downstream of the HSC compartment. The progeny derived from YFP\(^+\) MPP4s accounted for about 20–30% of the erythro-megakaryocyte lineage and about 60–80% of the myeloid lineage. Neither MPP2s nor MPP3s mediated B cell engraftment when an IL-7R exclusion gate was introduced. YFP labeling in 60% of myeloid cells and 30% of erythroid-megakaryocyte progeny suggested that one developmental pathway was marked by transient induction of lymphoid-associated transcripts, while the other one was independent. It is therefore possible to envision at least three developmental scenarios that would explain transient expression of lineage-specific genes in non-committed progenitors: in the first, the genomic landscape is plastic, and multipotency is maintained while lineage-specific genes such as Dntt can be turned on and off; in the second, there is simultaneous expression of lineage-specific genes that do not reach the necessary threshold of lineage regulators to ensure specification, or lineage branching is set in place, but the presence or lack of specific internal or external cues may re-direct cells to alternative fates. TdT was reported on immature leukemic blasts with both lymphoid as well as myeloid features, suggesting that, also in humans, TdT can be expressed in uncommitted precursor and that transient induction of lineage genes can occur independently of lineage specification\(^45,47\). Transient or low expression of TdT does reflect a permissive transcriptional state, in which exposure to cytokines (IL-7 for lymphoid, CSF-1 for myeloid and/or Epo for erythroid) or expression of selected transcription factors may influence commitment. The concept of lineage-defined niches is well known, and proliferation as well as migration will dictate which niche is likely to influence the fate of a given precursor. Gradients of cytokines and chemokines may intertwine, leading to the observed expression of lineage-specific transcripts in still uncommitted progenitors. It is possible that both intrinsic and extrinsic aspects are influencing HSCs, such as chromatin accessibility and receptor expression, as well as the cytokine or niche availability. The identification of ESAM as an ideal marker for multipotency and the observation that its downregulation is linked to lineage restriction may suggest that gene accessibility and chromatin landscape will mirror its expression. Collectively, we here redefine the hierarchy of early hematopoietic progenitors, validating experimentally and transcriptionally key stages that associate with multipotency and progressive lineage restriction across all three lineages.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41590-022-01167-5.

Received: 13 July 2021; Accepted: 18 February 2022; Published online: 30 March 2022

References
1. Sankaran, V. G. et al. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science 322, 1839–1842 (2008).
2. Porgera, C. M. et al. Hematopoietic stem cells are the major source of multilineage hematopoiesis in adult animals. Immunity 45, 597–609 (2016).
3. Eaves, C. J. Hematopoietic stem cells: concepts, definitions, and the new reality. Blood 125, 2605–2613 (2015).
4. Ikuta, K. & Weissman, I. L. Evidence that hematopoietic stem cells express mouse c-kit but do not depend on steel factor for their generation. Proc. Natl Acad. Sci. USA 89, 1562–1566 (1992).
5. Morrison, S. J. & Weissman, I. L. The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. Immunity 1, 661–673 (1994).
6. Ogawa, M. et al. B cell ontogeny in murine embryo studied by a culture system with the monolayer of a stromal cell clone. ST2: B cell progenitor develops first in the embryonal body rather than in the yolk sac. EMBO J. 7, 1337–1343 (1988).
7. Adolfsson, J. et al. Upregulation of Flt3 expression within the bone marrow Lin−Sca1−c−kit+ stem cell compartment is accompanied by loss of self-renewal capacity. Immunity 15, 659–669 (2001).
8. Christensen, J. L. & Weissman, I. L. Flk-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells. Proc. Natl Acad. Sci. USA 98, 14541–14546 (2001).
9. Kiel, M. J. et al. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. Cell 121, 1109–1121 (2005).
10. Yang, L. et al. Identification of Lin−Sca1−c−kit−CD34−FLt3− short-term hematopoietic stem cells capable of readily reconstituting and rescuing myeloablated transplant recipients. Blood 105, 2717–2723 (2005).
11. Arinobu, Y. et al. Reciprocal activation of GATA-1 and PU.1 marks initial specification of hematopoietic stem cells into myeloid and myelolymphoid lineages. Cell Stem Cell 1, 416–427 (2007).
12. Cabezas-Wallscheid, N. et al. Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. Cell Stem Cell 15, 307–322 (2014).
13. Oguro, H., Ding, L. & Morrison, S. J. SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. Cell Stem Cell 13, 102–116 (2013).
14. Ooi, A. G. et al. The adhesion molecule esam1 is a novel hematopoietic stem cell marker. Stem Cells 27, 653–661 (2009).
15. Pietras, E. M. et al. Functionally distinct subsets of lineage-biased multipotent progenitors control blood production in normal and regenerative conditions. Cell Stem Cell 17, 35–46 (2015).
16. Rodriguez-Fraticelli, A. E. et al. Clonal analysis of lineage fate in native haematopoiesis. Nature 553, 212–216 (2018).
17. Wilson, A. et al. Hematopoietic stem cells reversibly switch from dormancy to self renewal during homeostasis and repair. Cell 135, 1118–1129 (2008).
18. Wilson, N. K. et al. Combined single-cell functional and gene expression analysis resolves heterogeneity within stem cell populations. Cell Stem Cell 16, 712–724 (2015).
19. Yamamoto, R. et al. Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. Cell 154, 1112–1126 (2013).
20. Yokota, T. et al. The endothelial antigen ESAM marks primitive hematopoietic progenitors throughout life in mice. Blood 113, 2914–2923 (2009).
21. Ng, S. Y., Yoshida, T., Zhang, J. & Georgopoulos, K. Genome-wide lineage-specific transcriptional networks underscore Ikaros-dependent lymphoid priming in hematopoietic stem cells. Immunity 30, 493–507 (2009).
22. Mansson, R. et al. Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors. Immunity 26, 407–419 (2007).
23. Herman, J. S., Sagar & Grun, D. FateID infers cell fate bias in multipotent progenitors from single-cells RNA-seq data. Nat. Methods 15, 379–386 (2018).
24. Gilliland, S., Durisch, A., Lemeur, M., Benoist, C. & Mathis, D. Mice lacking TdT: mature animals with an immature lymphocyte repertoire. Science 261, 1175–1178 (1993).
25. Alberti-Servera, L. et al. Single-cell RNA sequencing reveals developmental heterogeneity among early lymphoid progenitors. *EMBO J.* 36, 3619–3633 (2017).
26. Balciunaite, G., Ceredig, R., Massa, S. & Rolink, A. G. A B220^+CD117^+CD19^− hematopoietic progenitor with potent lymphoid and myeloid developmental potential. *Eur. J. Immunol.* 35, 2019–2030 (2005).
27. Klein, F. et al. Accumulation of multipotent hematopoietic progenitors in peripheral lymphoid organs of mice over-expressing interleukin-7 and Flt3-ligand. *Front Immunol.* 9, 2258 (2018).
28. Dress, R. J. et al. Plasmacytoid dendritic cells develop from Ly6D^+ lymphoid progenitors distinct from the myeloid lineage. *Nat. Immunol.* 20, 852–864 (2019).
29. Rodrigues, P. F. et al. Distinct progenitor lineages contribute to the heterogeneity of plasmacytoid dendritic cells. *Nat. Immunol.* 19, 711–722 (2018).
30. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* 14, 865–868 (2017).
31. Melania Barile, K. B. et al. Hematopoietic stem cells self-renew symmetrically or gradually proceed to differentiation. Preprint at CellPress https://doi.org/10.2139/ssrn.3787896 (2021).
32. Gazit, R. et al. Transcriptome analysis identifies regulators of hematopoietic stem and progenitor cells. *Stem Cell Rep.* 1, 266–280 (2013).
33. Heng, T. S. et al. The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.* 9, 1091–1094 (2008).
34. Carrelha, J. et al. Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. *Development* 129, 2003–2013 (2002).
35. Street, K. et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* 19, 477 (2018).
36. Ishibashi, T. et al. ESAM is a novel human hematopoietic stem cell marker associated with a subset of human leukemias. *Exp. Hematol.* 44, 269–281 e261 (2016).
37. Sudo, T. et al. The endothelial antigen ESAM monitors hematopoietic stem cell status between quiescence and self-renewal. *J. Immunol.* 189, 200–210 (2012).
38. Sun, J. et al. Clonal dynamics of native haematopoiesis. *Nature* 514, 322–327 (2014).
39. Busch, K. et al. Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature* 518, 542–546 (2015).
40. Sommerkamp, P. et al. Mouse multipotent progenitor 5 cells are located at the interphase between hematopoietic stem and progenitor cells. *Blood* 137, 3218–3224 (2021).
41. Pei, W. et al. Polylox barcoding reveals hematopoietic stem cell fates realized in vivo. *Nature* 548, 456–460 (2017).
42. Boyer, S. W., Schroeder, A. V., Smith-Berdan, S. & Forsberg, E. C. All hematopoietic cells develop from hematopoietic stem cells through Flk2/Flt3-positive progenitor cells. *Cell Stem Cell* 9, 64–73 (2011).
43. Buza-Vidas, N. et al. FLT3 expression initiates in fully multipotent mouse hematopoietic progenitor cells. *Blood* 118, 1544–1548 (2011).
44. Drexler, H. G., Sperling, C. & Ludwig, W. D. Terminal deoxynucleotidyl transferase (TdT) expression in acute myeloid leukemia. *Leukemia* 7, 1142–1150 (1993).
45. Cuneo, A. et al. Clinical review on features and cytogenetic patterns in adult acute myeloid leukemia with lymphoid markers. *Leuk. Lymphoma* 9, 285–291 (1993).
46. Campagnari, F., Bombardieri, E., de Braud, F., Baldini, L. & Maiolo, A. T. Terminal deoxynucleotidyl transferase, TdT, as a marker for leukemia and lymphoma cells. *Int. J. Biol. Markers* 2, 31–42 (1987).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2022.
Methods

Mice. C57BL/6 wild-type (CD45.1, CD45.1/2 and CD45.2), TdT−/−, Rosa26CreERT2−/− and Rosa26loxPlox mice were bred and maintained in our animal facility under specific-pathogen-free conditions, according to institutional guidelines (Veterinäramt BS, license number 2786_26606 and ASP license number 19-896). All mice used as donors in transplantations and for analysis were 5–16 weeks old and recipient mice were 8–15 weeks old, and all of the C57BL/6 strain. TdT−/− and TdT−/− mice were generated at the Center for Transgenic Models in Basel using CRISPR–Cas9 technology. All Cas9 reagents were purchased from IDT. Briefly, RNPs consisting of Cas9 protein (40 ng μl −1), trcRNA (20 ng μl −1) and cRNAs (10 ng μl −1 each) targeting the last exon of the Dntt gene were injected before the stop codon, together with a single-stranded DNA template (IDT) encoding the P2A self-cleaving peptide in front of the human CD4 or iCre coding sequence flanked by 200-base-pair-long homology arms, were microinjected into C57BL/6 zygotes, essentially as described in ref. 51. Embryos that survived the DNA and Cas9 RNA microinjection were transferred into recipient females mated with genetically vasectomized males and the offspring were allowed to develop to term. Extended Data Fig. 1a illustrates the strategy used to generate the TdT−/− and TdT−/− mice by Cas9-mediated homology-directed repair. Genotyping was performed by PCR using different sets of primers. To detect hCD4 and iCre integration, a forward primer was located in the Dntt gene right before the transgenes and a reverse primer located in the untranslated region of the Dntt gene right after the transgenes. For reverse PCR, hCD4 FW1 + hCD4 RV1 (200 bp product); iCre FW1 + iCre RV1 (258 bp product) (Supplementary Table 1). To distinguish between homozygous and heterozygous mice, a forward primer located in the Dntt gene right after the transgenes and a reverse primer located in the untranslated region of the Dntt gene right after the transgenes were used: hCD4 Drt1 + hCD4 Drt2, Drt2 FW1 + Drt1 FW1 (291 bp product) (Supplementary Table 1). In mice heterozygous for hCD4 or iCre insertion, both PCR products are positive; for homozygous animals, PCR2 is negative (product too large for amplification). Furthermore, combinations of the primers allowed us to confirm transgene integration at the designated site: PCR3 + 4: hCD4 FW1 or iCre FW1 + hCD4 RV1; Drt2 FW1 + hCD4 RV1 or iCre RV1. PCR3 were performed with GoTaq Green Master Mix (Promega), according to the manufacturer’s instructions.

Cell collection and flow cytometry. Analysis and sorting, BM cells were flushed or extracted through fragmentation with a mortar and pestle from femurs and/or tibiae and/or pelvic bones of the two hind legs of mice with fluorescence-activated cell sorting (FACS) buffer (PBS containing 0.5% BSA and 5 mM EDTA) and single-cell suspensions of spleen and thymus cells were made. Debris was removed by filtration through a 70 μm filter. Cells were washed, fixed and permeabilized after cell-surface staining using a Fix/Perm buffer set (BD Biosciences). Single-cell suspensions were made from flush or extracted through fragmentation with a mortar and pestle from BM. For analysis and sorting, BM cells were fixed for 20 minutes at room temperature, 2 ml of FACS buffer was added to the platelet staining, which was then rapidly analyzed. To lyse red blood cells, 2 ml of FACs lysing solution (BD Biosciences) was added to the B cell and myeloid cell staining before analysis. For intracellular staining, cells were fixed and permeabilized after cell-surface staining using a Fix/Perm buffer set (BD Biosciences) according to the manufacturer’s protocol. Enrichment of progenitor cell populations prior to staining was performed by magnetic-activated sorting (Miltenyi Biotec) using biotin-labeled antibodies directed against lineage markers (CD3, CD19, B220, Ter119, NK1.1 and Ly6G) and anti-biotin MicroBeads (Miltenyi Biotec), according to the manufacturer’s protocol. For sorting, a BD FACSAria II instrument (BD Biosciences) with a custom built-in violet laser was used. Cells were sorted into Iscove’s modified Dulbecco’s medium (IMDM) supplemented with 5% FBS, 5 × 10−5 M β-mercaptoethanol, 1 mM glutamine, 0.03% (wt/vol) penicillin, 100 units ml −1 streptomycin and 100 pg ml −1 human FLT3-ligand (produced in-house) and 25 ng ml −1 murine FLT3-ligand (produced in-house) and 25 mg ml −1 murine stem cell factor (produced in-house) for OP9 co-cultures, at 37 °C in a humidified atmosphere containing 10% CO2 in the air. After 14 or 18 days in culture, for ST2 or OP9 co-cultures, respectively, wells were inspected under an inverted microscope, and wells containing colonies of more than 50 cells were scored as positive.

Methylcellulose cultures. For BFU-E methylcellulose assays, 500–2,000 cells in 1 ml SF M3436 (StemCell Technologies) supplemented with 100 units ml −1 penicillin and 100 μg ml −1 streptomycin were cultured in a 3 cm Petri dish. For simultaneous assessment of multilineage CFU-GE/M, CFU-GM, CFU-G, CFU-GM, CFU-G, CFU-GM, and CFU-GM, 200 cells were cultured in 1 ml M3231 (StemCell Technologies) supplemented with 5% FBS, 1-glutamine (2 mM), 100 units ml −1 penicillin, 100 μg ml −1 streptomycin and the following cytokines: SCF (25 ng ml −1), FLT3-ligand (25 ng ml −1), GM-CSF (10 ng ml −1), EPO (25 ng ml −1), TPO (25 ng ml −1), IL-3 (10 ng ml −1) and IL-11 (25 ng ml −1). Colonies were counted under an inverted microscope after 7 days of culture. Colonies of monocytes, granulocytes and erythroid cells (M) or megakaryocyte megakaryocytes (cMoPs) were defined as cKithiLy6C− and granulocytes, CFU-GM (mixed granulocyte and macrophage colonies), CFU-G (granulocyte colonies), CFU-M (macrophage colonies) and CFU-E (erythroid colonies).

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

CITe-seq. BM cells from four TdT−/− mice were sorted and stained for enriched progenitor cells by MACS by the use of antibodies directed against CD3, CD19, B220, Ter119 and Ly6G. Subsequently, cells were stained with antibodies directed against additional lineage markers (CD11b, CD11c, NK1.1, CD4, CD8, CD16/32, CD150, CX3CR1 or CD11b) and distinguished as cKithi and cKitlo/int, and cKithiLy6C− or cKithiLy6C+ and counted of UMIs for cDNAs and CITE-seq antibody barcodes. Cells were stained with antibodies coupled to oligonucleotides directed against CD4, CD8, CD48, CD150, CD115, CXCR4 and ESAM (Biolegend, see Supplementary Table 1). LSK cells were stained and an estimate of 4,000–6,000 cells per mouse were loaded on one well each of a single 10x Genomics Chromium Single Cell Controller. Single-cell capture and RNA library preparation were performed at the Genomics Facility Basel of the ETH Zurich, Basel, with a Single-Cell 3 v3 Reagent Kit (10x Genomics), according to the manufacturer’s instructions with the changes as described in ref. 53 to capture cDNA and produce libraries from antibody derived oligos (ADT). Sequencing was performed on 4 lanes (2 flow-cells) of an Illumina Novaseq 6000 instrument, with a mix of 90% cDNA library and 10% ADT library for the 2 first lanes, and 95% cDNA library and 5% ADT library for the 2 last lanes, to produce 91nt-long R2 reads.

The dataset was analyzed by the Bioinformatics Core Facility, Department of Biomedicine, University of Basel. Read quality was controlled with the FastQC tool (version 0.11.5). Sequencing files of both cDNA and ADT libraries were separately processed using cell ranger software (10x Genomics). Feature Barcoding Analysis instructions (https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/feature-bc-analysis) were followed to perform quality control, sample demultiplexing, cell barcode processing, alignment of cDNA reads to the mm10 genome with STAR (version 2.6.1.a) and counting of UMIs for cDNAs and CITe-seq antibody barcodes. Covariate file (Cell Ranger), which contains features and relevant metadata was downloaded by the tools outSAMmultiMax set to 1 and alignIntronMax set to 10,000. The reference transcriptome refdata-cellranger-mm10-3.0.0 using Ensemble 93 gene models.
Articles
NATURE IMMUNOLOGY
(https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest) was used, and supplemented by the sequences of the YFP, human CD4 and d2Cre constructs from the iT7-cre double reporter mice.

Filtering for high-quality cells was done on the basis of library size (at least 1,000 UMI counts per cell), the number of detected genes (at least 1,000 genes detected) and the percentage of reads mapping to mitochondrial genes (larger than 0% and less than 7%), based on the distribution observed across cells. Low-abundance genes with average counts per cell lower than 0.015 were filtered out. After quality filtering, the resulting dataset consisted of 12,165 genes and 20,595 cells, ranging from 3,993 to 6,286 per sample.

Further analyses were performed using R (version 3.6.6), and Bioconductor (version 3.10) packages, notably dropletUtils (version 1.6.1)74, scan (v1.14.6)75 and scater (v1.14.6)7, and the Seurat package (v4.0.5)72, mostly following the steps of the workflow presented at https://bioconductor.org/packages/release/ONCAsR/. Clusters of cells were performed on an normalized59 and denoised log-count values, with hierarchical clustering on the Euclidean distances between cells (with Ward’s criterion to minimize the total variance within each cluster);72 package cluster version 2.1.0). The number of clusters used for following analyses was identified by applying a dynamic tree cut (package dynamicTreeCut, version 1.63-1)72, resulting in 12 clusters and an average silhouette width of 0.09. As complementary lines displaying the 2D cell density on the UMAP space were calculated with the 10 components of the denoised principal component gradient, leading us to use also the

4e,f). For classification of YFP+ cells, ranging from 3,081 to 4,849 per sample. Re-clustering of these cells resulted in 8 clusters and an average silhouette width of 0.1. The findMarkers function of the scan package was used to find markers (genes, constructs or CITExseq antibodies) upregulated in any of the clusters. The top 30 markers for each cluster were extracted and pooled from a list of 104 markers (Fig. 4c). DEGs are displayed in G1 phase (Extended Data Fig. 4a,b,d). Cells from an additional cluster were filtered out because it was heterogeneous and composed of cells with elevated percentage of reads mapping to mitochondrial genes (for example, likely of lower quality; Extended Data Fig. 4c). The final filtered dataset was composed of 15,853 cells, resulting from 3,081 to 4,849 per sample. Re-clustering of these cells resulted in 8 clusters and an average silhouette width of 0.1. The findMarkers function of the scan package was used to find markers (genes, constructs or CITExseq antibodies) upregulated in any of the clusters. The top 30 markers for each cluster were extracted and pooled from a list of 104 markers (Fig. 4c). DEGs are displayed in Supplementary Table 2.

The Bioconductor package SingleR (version 1.0.5) was used for cell-type annotation of the cells8, using as a reference the relevant samples from the ImmunoGenome Project ImmGen mouse RNA-seq dataset (LTHSC,34–36, BM, LTHSC,34–BM, STHSC,150–BM, MPP2,150–BM, MPP3,150–BM, MPP4,135–BM, MPP4,135–BM, MPP4,135–BM, MPP4,135–BM, and the HSC, MPP1, MPP2, MPP3, and MPP4 bulk RNA-seq samples from Cabezás-Wallscheid et al.14. For the visualization of SingleR scores across cells on heatmaps, the scores were scaled between zero and one across populations for each cell and cubed to improve dynamic range next to one (ref. 2). A posteriori gating of cells to the LT-HSC, ST-HSC, MPP2, MPP3 and MPP4 populations was performed based on the surface protein and the CITE-seq antibodies (except for FLT3/CD115 which displayed a continuous gradient, leading us to use also the FLt3 transcript expression level to recover gating results most similar to the FACS analyses as shown in Extended Data Fig. 4e,f). For classification of YFP+/YFP−, hCD4+/hCD4− and ESAM+/ESAM− cells, a similar three-population approach was used, and the findMarkers function of the scan package was used to find differentially expressed markers between positive and negative populations at a false-discovery rate of 1% (in both directions).

A UMAP dimensionality reduction was used for visualizing single cells on 2 dimensions1, calculated using the runUMAP function from the scater package and default parameters (using the 10 components of the denoised principal component analysis as input, the 500 most variable genes and a neighborhood size of 15). For visualization, the y axis coordinates were adjusted which led to exclusion of eight cells separating from the bulk of other cells on the second dimension. Contour lines displaying the 2D cell density on the UMAP space were calculated with the MASS package (version 7.3–51.5).

Trajectory analysis was performed with the Bioconductor package Slingshot (version 1.4.0)1, a choice based on the very good performances of this tool in a recent benchmark of 45 single-cell trajectory inference72. We ran the analysis using the UMAP coordinates and the hierarchical clustering labels. Cluster 8 (HSCs) was set up as the start cluster. The cluster-based minimum spanning tree and the reconstructed smooth curves are shown in Fig. 5g. We compared this trajectory to the Monocle 3 results, where a cell from cluster 8 was also set as the starting point of the trajectory (Fig. 5g)3.

Integration of our dataset with a scRNA-seq dataset of sorted subsets from the Monocle 3 results, where a cell from cluster 8 was also set as the starting point of the trajectory (Fig. 5g)3.

Integration of our dataset with a scRNA-seq dataset of sorted subsets from the Monocle 3 results, where a cell from cluster 8 was also set as the starting point of the trajectory (Fig. 5g)3.

Conclusion of our dataset with a scRNA-seq dataset of sorted subsets from the Monocle 3 results, where a cell from cluster 8 was also set as the starting point of the trajectory (Fig. 5g)3. A newly generated UMAP projection of the joint dataset is shown as Extended Data Fig. 4i. Statistical analysis. A two-tailed unpaired Student’s t-test was performed comparing frequency of YFP+ subsets in BM and spleen at steady state and following sublethal irradiation (Fig. 7a). *P<0.05; **P<0.01; ***P<0.001; ****P<0.0001. A multiple two-tailed unpaired Student’s t-test was performed for experiments shown in Fig. 6d. *P<0.05; **P<0.01; ***P<0.001; ****P<0.0001. Error bars indicate s.e.m.

Acknowledgements
We dedicate this work to the memory of T. Rolink, who has been a great mentor and a friend to all of us. His vision and passion for research will remain. We acknowledge C. Engdahl, G. Capoforti, M. Burgunder and S. Sikanci for their contributions. We thank...
A. Offinger and L. Davidson and both teams of animal caretakers at the DBM Basel and NIDCR USA for constant support. Further we would like to acknowledge the Genomics Facility Basel (D-BSSE ETH Zürich) for generating the CITE-seq dataset. Calculations were performed at scICORE (http://scicore.unibas.ch/) scientific computing center at the University of Basel. We also thank Y. Belkaid, G. Trinchieri, A. Bhandoola and C. Dunbar for their inputs and discussion. This work was in part supported by the SNF grants PP00P3_179056, 310030_185193 and by the Research Fund of the University of Basel for the promotion of excellent junior researchers (FK). This research was in part supported by the Intramural Research Program of the NIH, NIDCR (ZIADE000752-02).

**Author contributions**
F.K., G.C., P.F.R., L.v.M., P.T., S.Y., R.L., P.P. and R.T. designed and performed experiments; F.K., G.C. and P.P. generated the TdThCD4 and TdTiCre mouse lines; F.K., P.F.R., J.R. and R.T. analyzed data; F.K. and R.T. conceived the project and wrote the manuscript.

**Competing interests**
The authors declare no competing interests.

**Additional information**
Extended data is available for this paper at https://doi.org/10.1038/s41590-022-01167-5.
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41590-022-01167-5.
Correspondence and requests for materials should be addressed to Roxane Tussiwand.

**Peer review information** Nature Immunology thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Ioana Visan, in collaboration with the Nature Immunology team. Peer reviewer reports are available.

**Reprints and permissions information** is available at www.nature.com/reprints.
Extended Data Fig. 1 | Generation of TdT^hCD4 and TdT^iCre mice. (a) Schematic Fig. illustrating the generation of TdT^hCD4 and TdT^iCre mouse lines. First Cas9, guided by a specific crRNA and the common trcrRNA, introduced a double-strand break before the stop-codon of the Dntt gene. Homology directed repair resulted in the integration of ssDNA repair templates containing 200 base-pair homology regions and the P2A self-cleaving peptide followed by either the hCD4 or iCre coding sequence. Small arrows spanning the final constructs indicate primers used in (b). (b) LSK cells isolated from the BM of 6-8 weeks old TdT^hCD4/YFP mice were sorted based on the expression of hCD4 as indicated. Shown are bar graphs depicting the relative mRNA expression of Dntt, hCD4, iCre as well as inter-spanning transcripts of Dntt-hCD4 and Dntt-iCre using primer pair combinations shown in (a) Error bars indicate s.e.m. (n=5-6). (c-d) LSK cells subdivided based on the expression of hCD4 as indicated were analyzed for intracellular TdT expression. Shown are representative histograms (c) and cumulative bar graphs of the mean fluorescence intensities (d) depicting intracellular TdT (left, n=5) and surface hCD4 (right, n=6) in TdT^hCD4 BM LSK. Error bars indicate s.e.m. (e) Shown is the gating strategy used to identify Ly6D+ EPLMs in the BM.
Extended Data Fig. 2 | See next page for caption.
Extended Data Fig. 2 | Gating strategies to identify progenitors and mature cells. (a-d) Shown are representative FACS plots depicting the gating strategies used to identify (a) lymphocytes as B cells, CD4/CD8 T cells, gd-T cells and NK cells in the spleen; (b) pro-erythrocytes in the BM; (c) platelets in peripheral blood; (d) pDC, pDC-like, cDC1, cDC2, monocytes and granulocytes in the spleen. (e) Shown is a representative FACS plots depicting the expression of CD150 and CD48 on MPP4s gated as shown in Fig. 2c. (f, g) Shown are representative FACS plots depicting the gating strategies used to identify (f) MkPs, CFU-Es, GMPs and (g) cMoPs, MDPs and CDPs in the BM. (h-i) Shown are representative FACS plots depicting the expression of (h) intracellular TdT against hCD4 (h) and surface hCD4 plotted against CD150 (i) on the indicated progenitor subsets gated as shown in Fig. 2c.
Extended Data Fig. 3 | See next page for caption.
Extended Data Fig. 3 | Reduced self-renewal capacity of YFP− MPP4s. (a-d) Shown is the distribution of Tomato+ and GFP+ in TdT+flk2 mice on the indicated progenitor subsets gated as in Fig. 2c. Data are derived from 2 independent experiments (n=3). Error bars indicate s.e.m. (a) Shown is the distribution of Tomato+ (red), Double positive (orange) and GFP+ single positive (green) cells within each subset as indicated. (b) Shown is the percent distribution of GFP+Tomato+++; GFP+Tomato++ and GFP+Tomato++ as indicated. (c) Shown is the percent distribution of cells that are Tomato+++; Tomato++ and Tomato++ within GFP+ progenitor subsets as indicated. (d) Shown is the expression of Tomato and GFP for the indicated subsets. (e) Shown is the expression of CD150 and CD48 on IL7Rα cells pre-gated as LSK (Lin−FL−T3−CD117+ Sca1+). (f-h) Shown are the gating strategies to identify and quantify (f) donor derived B cells, (g) myeloid cells, (h) and platelets in peripheral blood of recipient mice. (i) 4000 GFP+ MPP4, GFP+ or GFP− were transferred into sub-lethally irradiated WT recipients. Shown are percent reconstitution in the bone marrow of recipient animals of Tomato+ (red bars) and GFP+ (green bars) MPP2s, MPP3s and MPP4s after 2 and 4 weeks. Cumulative from 2-3 independent experiments for each timepoint (n=4–8). Error bars indicate s.e.m. (j-k) 1500 LT− (j) and ST-HSCs (k) were sorted from the BM of 6–8 weeks old WT mice and transferred i.v. in competition with 1500 YFP− MPP4s sorted from the BM of 6–8 weeks old TdT−/−CD45.1/2 mice into lethally irradiated WT CD45.1 recipient mice. 3x10⁵ WT CD45.1 BM cells were co-injected as support cells. Shown is the percent peripheral blood reconstitution for CD11b−CD3−NK1.1− myeloid and CD19−CD11b− B cells at the indicated timepoints after transfer. Data were collected from 2 independent experiments ((j) n=7; (k) n=8). Error bars indicate s.e.m.
Extended Data Fig. 4 | See next page for caption.
Extended Data Fig. 4 | HSC and MPP transcriptional profiling using a posteriori gating. LSK cells isolated from BM of four 6–8 weeks old TdTomato/COPA/YFP double reporter mice were used for single-cell RNA sequencing in combination with CITE-Seq labelling as described in the methods. (a–c) Hierarchical clustering analysis was performed and projected in a 2-dimensional space using UMAP as explained in the methods. Each color represents a specific cluster as indicated. (a) Hierarchical clustering identified 12 clusters. (b) Cell-cycle phase, represented by the colors yellow (G1), orange (S) and red (G2), of each cell was determined as described in the methods. (c) Unique molecular identifiers (UMIs) coming from the mitochondrial (MT) genome were quantified across cells, reflected by the color intensity. Contour lines display the 2D cell density on the UMAP space. (d) Shown is a bar plot distribution indicating the frequency of cells in the different phases of the cell cycle across subsets obtained for the “a posteriori gating defined as in Fig. 4b. (e) UMAP and bar graphs illustrating the scaled expression of the Flt3, Slamf1, and Cd48 mRNA (left panels), as well as the expression of their corresponding surface markers used for CITE-Seq (right panels). The colors represent cells from the different clusters. Dot size and color intensity indicate expression levels. Bar height in bar graphs indicate the average expression across cells from each biological replicate across clusters. (f) A posteriori gating strategy used to define HSC and MPP populations within the CITE-Seq data. For MPP4: Flt3 > 2.5, CD135 > 3.5, and CD150 < 4. For MPP3: Flt3 < 2, CD135 < 3, CD150 < 4, and CD48 > 7. For MPP2: Flt3 < 2.5, CD135 < 3.5, CD150 > 5.5, and CD48 > 7. For LT-HSC: Flt3 < 2.5, CD135 < 3.5, CD150 > 5.5, and CD48 < 6.5. For ST-HSC: Flt3 < 2, CD135 < 3, CD150 < 4, and CD48 < 6.5. The colors represent cells from the different clusters. (g) Heatmap displaying the centered and scaled expression of the top differentially expressed genes between the gated populations defined as in (e), resulting in a list of 96 markers. Cells were ordered following the hierarchical clustering tree. Cluster assignment and similarity score of each cell to reference ImmGen RNA-seq samples is shown on top of the heatmap. (g) Shown is the UMAP distribution of our single cell dataset assigned using the previously published bulk RNA-Seq obtained from sorted progenitors (see gates below) from ref. 28. (i) Shown is the UMAP distribution for the integrated analysis of our dataset with the previously published scRNA-Seq obtained from ref. 34 obtained using the Seurat package (findIntegrationAnchors function)35. On the left the overlay of the two data sets on the right the scRNA-Seq obtained from ref. 34.
Extended Data Fig. 5 | Single-cell gene expression profiling of LSKs. (a-d) UMAP and bar graphs illustrating the scaled expression for selected (a) stem cell-related markers and surface receptors (Cd34, CD48, CD9, ESAM); (b) erythroid (Gata1, Klf1, Vwf, and Pf4), (c) myeloid (Mpo, Irf8, Ctsg, Elane), (d) lymphoid (Ighm, Ighd, Notch1, Lck). The colors represent cells from the different clusters. Dot size and color intensity indicate expression levels. Below is the distribution across clusters where bar height indicates the average expression across cells from each biological replicate across clusters.
Extended Data Fig. 6 | See next page for caption.
Extended Data Fig. 6 | Single cell profiling and functional characterization of MPP3s. (a) UMAP and bar graphs illustrating the scaled expression of the Dntt mRNA, hCD4 surface marker, and YFP. The colors represent cells from the different clusters. Dot size and color intensity indicate expression level. Bar height in bar graph indicate the average expression across cells from each biological replicate across clusters. (b) Compiled data showing hCD4−YFP−, hCD4+YFP+, and hCD4−YFP− MPP3s, annotated based on a posteriori gating as in Fig. 4b, in relation to their cluster distribution defined as in Fig. 4a; the similarity score to reference samples from the ImmGen dataset defined as in Fig. 4c; centered and scaled expression for the top 26 markers differentially expressed between subsets. (c) UMAP plot illustrating the distribution of hCD4−YFP−, hCD4+YFP−, hCD4+YFP+, and hCD4−YFP+ MPP3 cells. (d) Bar graph showing the distribution of the subsets as in (c) across clusters defined as in Fig. 4a. The colors represent the different clusters. (e) hCD4−YFP−, hCD4+YFP+, and hCD4−YFP+ MPP3s were sorted from the BM of 6–8 weeks old TdThCD4/YFP mice and 1500 cells were transferred i.v. in competition with 1500 ST-HSCs sorted from the BM of 6–8 weeks old WT-CD45.1/2 mice into sub-lethally irradiated WT-CD45.1 recipients. Shown are the percent peripheral blood reconstitutions for B cells (CD19+CD11b−) and myeloid cells (CD11b+, CD3−, NK1.1−) at the indicated timepoints after transfer. Data were collected from 2 independent experiments (hCD4−YFP− MPP3 n = 5; hCD4+YFP+ MPP3 n = 6; hCD4−YFP− MPP3 n = 6). A multiple two-tailed unpaired Student’s t test was performed. (Myeloid cells: hCD4−YFP− and hCD4+YFP+ day 10 P = 0.026, day 14 P = 0.03, day 18 P = 0.049; hCD4−YFP− and hCD4−YFP+ day 10 P = 0.005, day 14 P = 0.026; hCD4+YFP+ and hCD4−YFP+ day 10 P = 0.004). *, P < 0.05; **, P < 0.01; ***, P < 0.001. ***, P < 0.0001. Error bars indicate s.e.m. (f) UMAP plot illustrating the distribution of hCD4−YFP−, hCD4+YFP−, hCD4+YFP+, and hCD4−YFP+ (orange) MPP4 cells. (g) Bar graph showing the distribution of the subsets as in (f) across clusters defined as in Fig. 4a. The colors represent the different clusters. (h) hCD4−YFP−, hCD4+YFP−, and hCD4+YFP+ MPP4s were sorted from the BM of 6–8 weeks old TdThCD4/YFP mice and 1500 cells were transferred i.v. in competition with 1500 ST-HSCs sorted from the BM of 6–8 weeks old WT-CD45.1/2 mice into sub-lethally irradiated WT-CD45.1 recipients. Shown are the percent peripheral blood reconstitutions for B cells (CD19+CD11b−) and myeloid cells (CD11b+, CD3−, NK1.1−) at the indicated timepoints after transfer. Data were collected from 2 independent experiments (hCD4−YFP− MPP4 n = 5; hCD4+YFP+ MPP4 n = 6; hCD4−YFP− MPP4 n = 7). A multiple two-tailed unpaired Student’s t test was performed (B cells: hCD4−YFP− and hCD4+YFP+ day 14 P = 0.008; hCD4−YFP− and hCD4−YFP+ day 14 P = 0.002, day 18 P = 0.03; Myeloid cells: hCD4−YFP− and hCD4−YFP+ day 10 P = 0.011, day 14 P = 0.0003, day 18 P = 0.000004; hCD4+YFP+ and hCD4−YFP+ day 10 P = 0.00021, day 14 P = 0.0012, day 18 P = 0.002). *, P < 0.05; **, P < 0.01; ***, P < 0.001; ****, P < 0.0001. Error bars indicate s.e.m.
Extended Data Fig. 7 | See next page for caption.
Extended Data Fig. 7 | ESAM defines the developmental hierarchy of MPPs. (a) Two color histograms depicting the expression of YFP and ESAM from BM cells of 6–8 weeks old TdTC4YFP mice. Cells are pre-gated as shown in Fig. 2c. (b–c) Progenitors were defined using the “a posteriori” gating (Extended Data Fig. 4f) as ESAM negative and positive MPP2s (top) MPP3s (middle) and MPP4s (bottom). (b) Shown is the distribution with the frequency across clusters for each subset. (c) Shown are volcano plots projecting the difference in gene expression for each subset. Genes downregulated in the ESAM− fractions with an FDR < 0.01 are marked in blue and genes upregulated in the ESAM+ fractions are marked in red. Genes with an abs log2 FC > 1 are labeled. (d) 4000 ESAM negative and positive MPP2s (top) MPP3s (middle) and MPP4s (bottom) were isolated from the BM of 6–8 weeks old Rosa26mT mG mice and transferred i.v. into sub-lethally irradiated WT recipients. Shown are cumulative data with percent reconstitution of recipient animals for mature subsets (gated as shown in Extended Data Fig. 2a-d) after four weeks in the bone marrow (pro-Erythrocytes), peripheral blood (platelets) and spleen (all other subsets). Data were collected from 3 independent experiments. A multiple two-tailed unpaired Student’s t test was performed (MPP4: B cells P = 0.003, CD4+ T cells P = 0.0097, CD8+ T cells P = 0.005, pDCs P = 0.000005, monocytes P = 0.000001, pro-erythrocytes P = 0.006). **, P < 0.01; ****, P < 0.0001. Error bars indicate s.e.m. (e) Compiled data showing ESAM+hCD4−, ESAM+hCD4+, ESAM−hCD4+, and ESAM−hCD4− MPP4s, annotated based on the “a posteriori” gating as in (Extended Data Fig. 4f), in relation to their cluster distribution defined as in Fig. 4a; the similarity score to reference samples from the ImmGen dataset defined as in Fig. 4c; The top 38 differentially expressed markers and genes between subsets is centered and scaled. (f) Representative histograms illustrating the expression levels of hCD4 within LT-HSCs, ESAM− and ESAM+ MPP4s, and CLPs isolated from the BM of 6–8 weeks old TdTC4 mice. (g) Shown is the expression of Sca-1 and ESAM for LT- (in black) and ST-HSCs (color scale) are projected into a t-SNE plot. ST-HSCs are projected indicating the expression of ESAM. (h-k) CD48+ LSKs isolated from the BM of 6–8 weeks old TdTC4 mice were sorted based on the expression of hCD4 as indicated. (h) 2000 cells were analyzed for erythroid colony forming potential (BFU-E). Shown are the number of colonies obtained. Data were collected from 3 independent experiments (n = 6). Statistical analysis was done with two-tailed unpaired Student’s t test (hCD4-neg and hCD4-int P = 0.002). *, P < 0.05; **, P < 0.01; ***, P < 0.001; ****, P < 0.0001. Error bars indicate s.e.m. (i) Myeloid precursor frequency was determined for the indicated subsets by in vitro limiting dilution analysis, as described in methods. Shown is one representative experiment (n = 3). Error bars indicate s.e.m. (j, k) 4000 cells were transferred i.v. into sub-lethally irradiated WT-CD45.1 recipients in competition with 4000 hCD4-neg CD48+ LSK cells sorted from TdTC4-CD45.1/2 mice. Shown is the percent peripheral blood reconstitution for CD11b+CD3−NK1.1 myeloid (j) and CD19+CD11b− B cells (k) at the indicated timepoints after transfer. Data were collected from 3 independent experiments. Error bars indicate s.e.m.

ARTICLES

NATURE IMMUNOLOGY | www.nature.com/natureimmunology
Extended Data Fig. 8 | A new functional classification of early hematopoietic progenitors. (a,b) Shown is the improved gating strategy depicting the use of ESAM to identify LT-, ST-HSCs, bona fide MPPs and oligopotent progenitors (OPP) with prominent myeloid and limited erythroid OPP-Em potential. Lin− progenitors are pre-gated as IL-7R− and referred as LEKs (Lineage ESAM+ cKit+). LEKs are further subdivided based on FLT3 expression as bona fide MPPs (FLT3+ LEK), ST-HSCs (FLT3−CD48−CD150− LEK), LT-HSCs (FLT3−CD48−CD150− LEK), Oligopotent progenitors with prominent Myeloid- and limited erythroid (OPP-Me, FLT3−CD48−CD150− LEK) or prominent Erythroid and limited myeloid potential (OPP-Em, FLT3−CD48−CD150− LEK) (c). Progenitors pre-gated as IL-7R− and referred as LSK (Lineage Sca1+ cKit+) can be further subdivided based on FLT3 expression as MPP4s (FLT3+). Within this MPP4 subset the expression of ESAM in C57BL/6 mice identifies bona fide MPPs (FLT3+ESAM+ LSK), whereas the combined use of ESAM and hCD4 in TdT− reporter mice allows for the identification of Lymphoid Progenitors (LP, FLT3+hCD4lowESAM− LSK), and Oligopotent Progenitors with Lymphoid- and Myeloid potential (OPP-LM) (FLT3+hCD4+ESAM− LSK) besides MPPs (FLT3−ESAM− LSK). FLT3−LSK can be further subdivided as shown into ST-HSCs (FLT3−ESAM−CD48−CD150− LSK), LT-HSCs (FLT3−CD48−CD150− LSK), OPP-Me (FLT3−ESAM−CD48−CD150− LSK), OPP-Em (FLT3−ESAM−CD48−CD150− LSK), Myeloid Progenitors (MP, FLT3−ESAM−CD48−CD150− LSK) and Erythroid Progenitors (EP, FLT3−ESAM−CD48−CD150− LSK) as shown (d). Highlighted are the newly identified subsets. (c) Each pre-gated subset as obtained from Extended Data Fig. 8a,b is color coded according to their transcriptionally most similar cluster as defined in Fig. 4a and gated using selected markers (upper panel) or projected in a two-dimensional t-SNE (lower panel) plot. (d,e) Schematic model of steady state (d) and emergency (e) hematopoiesis illustrating the proposed hierarchy as observed for HSCs and MPPs in TdT−reporter and lineage tracer mice. Progenitors are labeled for hCD4, YFP, ESAM and FLT3 expression allowing for the new subset’s definition: OPP-LM, OPP-Me, OPP-Em, LP, MP and EP.
Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
  Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection
Flow cytometry and cell sorting were performed using FACS Diva using BD LSORTFotessa and BD FACS Aria IIu respectively. For CITE-seq single-cells were captured using a single 10X Genomics Chromium Single Cell Controller and libraries were sequenced on an Illumina NovaSeq 6000 instrument.

Data analysis
Flow cytometry data was analyzed on FlowJo v10.
Graphs and statistics were created and analyzed using GraphPad Prism Version 8 with the exception of the CITE-seq dataset. Read quality of the CITE-seq was controlled with the FastQC tool version 0.11.5 and sequencing files were further processed using the Cell Ranger Software v3.1.0. Subsequent analysis was performed using R version 3.6 (R studio as platform) and Bioconductor (version 3.10) packages, including dropletUtils (version 1.6.1), scran (v1.14.6), and scater (v1.14.6). The SingleR package was used for cell-type annotations and the Slingshot package (version 1.4.0) was used to perform trajectory analysis.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.
Data
Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The CITE-Seq dataset is available at the Gene Expression Omnibus database under accession number GSE145491.

Field-specific reporting
Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design
All studies must disclose on these points even when the disclosure is negative.

Sample size
Sample size for each experiment is stated in the figure legends. Sample size was not predetermined by statistical methods, but chosen based on pilot experiments and previous experience with similar experiments. Data were collected from two or more biological replicates.

Data exclusions
Data from experiment involving sorting were excluded if the isolated populations were not pure, which resulted in aberrations in transplantation experiments, trackable based on GFP and Tomato signals.

Replication
All experiments were repeated and reliably reproduced. Data were collected from three biological replicates if not stated differently. Number of mice used and independent experiments performed are indicated in the figure legends.

Randomization
The experimental groups were determined based on their genotype. Litter mates, where applicable, or age and gender matched mice were used as controls. For transplantation and irradiation experiments animals were selected randomly from the same cage for different treatments.

Blinding
Investigators were not blinded during data collection if the obtained data was not subjective. Methylcellulose assay was performed blinded. Counting was performed from another person than the one plating the cells, therefore not knowing the corresponding populations plated.

Reporting for specific materials, systems and methods
We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | Methods |
|----------------------------------|---------|
| n/a (Involved in the study)      | n/a (Involved in the study) |
| □ Antibodies                     | □ ChIP-seq |
| □ Eukaryotic cell lines          | □ Flow cytometry |
| □ Palaeontology and archaeology  | □ MRI-based neuroimaging |
| □ Animals and other organisms    |         |
| □ Human research participants    |         |
| □ Clinical data                  |         |
| □ Dual use research of concern   |         |

Antibodies
Antibodies used
- anti-CD3, Pacific Blue, clone 17A2, Biolegend 100214
- anti-CD3, BV421, clone 17A2, Biolegend 100228
- anti-CD3ε, Biotin, clone 145-2C11, Biolegend 100304
- anti-CD3ε, AF488, clone 145-2C11, In-house
- anti-CD3ε, PE, clone 145-2C11, Biolegend 100308
- anti-CD3ε, PECy7, clone 145-2C11, Biolegend 100320
anti-CD4, APC/Fire 750, clone GK1.5, Biolegend 100460
anti-CD4, AF648, clone OKT4, In-house
anti-CD8a, Bv510, clone 53-6.7, Biolegend 100752
anti-CD11b, Biotin, clone M1/70, In-house
anti-CD11b, Bv785, clone M1/70, Biolegend 101243
anti-CD11b, PerCP-Cy5.5, clone M1/70, Biolegend 101228
anti-CD11c, Biotin, clone N418, In-house
anti-CD11c, APC/Fire 750, clone N418, Biolegend 117352
anti-CD16/32, unlabeled, clone 2.4G2, Biolegend 101302
anti-CD16/32, PE, clone 2.4G2, BD Biosciences 561727
anti-CD19, Biotin, clone 1D3, In-house
anti-CD19, AF647, clone 1D3, In-house
anti-CD19, Pacific Blue, clone 1D3, Biolegend 115523
anti-CD19, Bv785, clone 1D3, BD Biosciences 563333
anti-CD19, PE, clone 1D3, Biolegend 102016
anti-CD19, Bv785, clone 1D3, Biolegend 117352
anti-CD21, PerCP-Cy5.5, clone 7E9, Biolegend 123416
anti-CD23, PE, clone B3B4, Biolegend 101614
anti-CD25, AF647, clone PC61, Biolegend 102008
anti-CD25, PE, clone PC61, Biolegend 102016
anti-CD28, Biotin, clone 37.51, Biolegend 102104
anti-CD28, PE, clone 37.51, Biolegend 102016
anti-CD41, Bv510, clone MWReg30, Biolegend 133923
anti-CD44, Bv785, clone IM7, Biolegend 103059
anti-CD45.1, PAC Blue, clone 1D3, Biolegend 115523
Validation

Antibodies were validated by the manufacturer and if produced in house they were validated and optimized by titration prior to use.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s) | OP9 and ST2 stromal cells were expanded and grown in our facility.
Authentication | B cell colony formation (OP9) and myeloid colony formation (ST2) from sorted BM progenitors.
Mycoplasma contamination | Cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register) | No misidentified cell lines were used in this study.

Animals and other organisms

Policy information about studies involving animals: ARRIVE guidelines recommended for reporting animal research

Laboratory animals | Species: Mus Musculus. Strains: WT C57BL/6 (CD45.1, CD45.1/2 and CD45.2), TdT-hCD4, TdT-iCre, Rosa26-LSL-YFP, Rosa26-LSL-mTmG. TdT-iCre mice were crossed to Rosa26-LSL-YFP and Rosa26-LSL-mTmG to obtain TdT-YFP and TdT-mTmG mice, respectively. All mice used (males and females) as donors in transplantations and for analysis were 6-10 and recipient mice were 8-15 weeks old, and all were of the C57BL/6 strain.
Wild animals | The study did not involve wild animals.
Field-collected samples | The study did not involve samples collected from the field.
Ethics oversight | All animals were bred and maintained in a specific pathogen-free animal facility according to institutional guidelines (authorization number 2786_26606 from cantonal veterinarian office, Canton Basel-Stadt).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☑️ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).
☑️ All plots are contour plots with outliers or pseudocolor plots.
☑️ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation | Bone marrow cells were flushed or extracted through fragmentated with a mortar and pestle from femurs and/or tibiae and/or pelvic bones of the two hind legs of mice with FACS buffer (PBS containing 0.5% BSA and 5 mM EDTA) Spleen and thymus were smashed. Blood was collected from the tail vein. Debris was removed by filtration through a 70 μm strainer. Red blood cells were lysed with ACK lysis buffer. Cells were counted and stained for analysis and cell sorting.
Instrument | BD LSRFortessa was used for analysis. BD FACSArria Iiu with a custom built in violet laser was used for sorting.
Software | FACS DIVA (BD Biosciences) for acquisition and FlowJo v10 (TreeStar) for analysis.
Cell population abundance | All sorted samples were checked for post-sorting purity. The abundance of the sorted populations were >95%.
Gating strategy | Samples were gated on size by FSC-A/SSC-A and doublets were excluded using SSC-W/SSC-H. Dead cells were removed by propidium iodide or 7AAD stain. Please see the detailed gating strategies used provided within the methods, figure legends and figures of the manuscript.

☑️ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.