Epigenetic programming defines haematopoietic stem cell fate restriction.

Yiran Meng[1], Joana Carrelha[1,2], Roy Drissen[1], Xiying Ren[1], Bowen Zhang[1], Adriana Gambardella[1], Simona Valletta[1], Supat Thongjuea[3], Sten Eirik Jacobsen[1,2,4,5,6] and Claus Nerlov[1,7].

[1]MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DS, UK. [2]Haematopoietic Stem Cell Laboratory, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DS, UK. [3]MRC Centre for Computational Biology, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DS, UK. [4]Department of Cell and Molecular Biology, Wallenberg Institute for Regenerative Medicine, Karolinska Institutet, Stockholm SE-17177, Sweden. [5]Department of Medicine Huddinge, Center for Hematology and Regenerative Medicine, Karolinska Institutet, Stockholm SE-17177, Sweden. [6]Karolinska University Hospital, Stockholm SE-17177, Sweden. [7]Corresponding author: claus.nerlov@imm.ox.ac.uk.

Abstract.

Haematopoietic stem cells (HSCs) are multipotent, but individual HSCs can show restricted lineage output *in vivo*. Currently, the molecular mechanisms and physiological role of HSC fate-restriction remain unknown. We here show that lymphoid fate is epigenetically, but not transcriptionally, primed in HSCs. In multi-lineage HSCs that produce lymphocytes, lymphoid-specific upstream regulatory elements (LymUREs), but not promoters, are preferentially accessible compared to platelet-biased HSCs that do not produce lymphoid cell types, providing transcriptionally silent lymphoid lineage priming. *Runx3* is preferentially expressed in multi-lineage HSCs, and re-instating Runx3 expression increases LymURE accessibility and lymphoid-primed MPP4 progenitor output in old, platelet-biased HSCs. In contrast, platelet-biased HSCs show elevated levels of epigenetic platelet-lineage priming, and give rise to MPP2 progenitors with molecular platelet-bias, which generate platelets with faster kinetics, and through a more direct cellular pathway, compared to MPP2s derived from multi-lineage HSCs. Epigenetic programming therefore predicts both fate-restriction and differentiation kinetics in HSCs.

Main.

Haematopoietic stem cells (HSCs) are rare bone marrow-resident cells with extensive self-renewal capacity that sustain definitive haematopoiesis lineage throughout the mammalian lifespan. While HSCs as a population continuously generate all definitive haematopoietic cell types, single cell analysis has identified HSCs that preferentially or selectively produce a subset of blood cell types[1-3], and shown that such fate-restrictions are hierarchically organised[4]. Identification of molecular traits associated with HSC fate-restriction remains challenging, as we currently lack molecular markers that allow HSC subtypes to be prospectively isolated at a level of purity sufficient for accurate molecular

characterization. Combined barcoding and single cell RNA sequencing (scRNAseq) has been successfully used to identify transcriptional signatures associated with HSC lineage bias[5,6]. However, single cell approaches are limited by the uncertainty associated with measuring lineage output from single HSCs and the difficulty in obtaining multiple, deep genomic data sets from single cells[7]. Fluorescent barcoding of HSC clones has shown correlation between epigenetic lineage priming and lineage output, but did not allow measurement of platelet or erythroid output from HSCs, precluding identification of platelet-biased HSCs[8]. We have therefore used chromatin and transcriptome profiling of clonal HSC populations derived from single cell transplantation to define the chromatin accessibility and transcriptional programs associated with HSC fate-restriction, and their role in establishing the downstream differentiation pathways specific to HSC subtypes.

Results.

Fate-restricted HSC clones are transcriptionally distinct and coherent.

Clonal HSC populations that arise from transplantation of single HSCs in principle allow multi-modal in-depth molecular profiling of HSCs with precisely defined fate-restriction. However, a prerequisite for such an approach is that clonal HSCs populations are molecularly homogeneous and distinct between different types of fate-restriction. We therefore generated fate-restricted HSC clones by single cell transplantation as described[4] (Fig. 1a, Extended Data Fig. 1a), and performed Smart-seq2–based scRNAseq[9] on phenotypic HSCs (Lin–c-Kit+Sca-1+CD150+CD48–*Gata1*-EGFP–; Extended Data Fig. 1b) isolated from clones that were platelet (PLT) or platelet-erythroid-myeloid (PEM) restricted in their lineage output, as well as multi-lineage (MUL) HSCs (Fig. 1b). Classifier genes were identified from random forest analysis (Fig. 1c; Extended Data Fig. 2). We selected the top 500 genes to cluster the single cell transcriptomes (Fig. 1d), as genes below this cut-off had low importance scores (Fig. 1c). In this analysis clones with similar fate-restriction formed separate clusters of highly significant purity (Supplementary Table 1), showing that clonal, fate-restricted HSC populations were molecularly distinct and coherent. We therefore performed both RNAseq and ATACseq on clonal PLT-, PEM- and MUL-HSCs bulk populations. The genes differentially expressed across HSC subtypes contained a significant proportion of genes with PLT+PEM- and PEM+MUL-specific expression, but few genes with PLT+MUL-specific expression (Fig. 2a; Supplementary Table 2). Similarly, ATACseq identified significant accessibility differences between PLT- and MUL-HSCs, with PEM-HSCs having very few uniquely accessible (or uniquely non-accessible) chromatin sites (Fig. 2b; Supplementary Table 3). Correlation analysis using these variable features showed that by both gene expression (Fig. 2c) and chromatin accessibility (Fig. 2d) the PLT- and MUL-HSC subtypes were the more distantly related. PEM-HSCs are therefore molecular intermediates between PLT- and MUL-HSC that share features with both these HSC subtypes, consistent with PEM-HSCs being functionally intermediate between MUL- and PLT-HSCs. To comprehensively identify differential molecular properties between HSC subtypes we therefore compared MUL- and PLT-HSCs using Gene Set Enrichment Analysis[10]. Using gene sets associated with stemness[11,12], proliferation[13,14], lineage bias and cellular output[6] we found that

PLT-HSCs were enriched in signatures associated with HSC quiescence, stemness, platelet-bias and low output, whereas MUL-HSCs showed enrichment for signatures associated with cell cycle progression, and multi-lineage and high cellular output (Fig. 2e), a pattern similar to that seen when comparing platelet-biased and multi-lineage HSCs using barcoding[6]. To identify cellular pathways underlying these differences we performed similar analysis using MSigDB Hallmark signatures (https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp#H). This showed that the high expression of proliferative/high output signatures in MUL-HSCs was associated with enrichment of genes regulated by E2F and MYC, both canonical transcription factors (TFs) that promote cell cycle progression, as well as DNA repair genes, required for correction of replication errors (Fig. 2f). In contrast, PLT-HSCs were highly enriched for gene signatures associated with inflammation, and interferon response in particular, consistent with the increased sensitivity to interferon signalling observed in malignant HSCs in Polycythemia Vera being a consequence of their platelet bias[15].

Epigenetic priming of UREs predicts lymphoid fate in HSCs.

While most of the chromatin preferentially accessible in MUL-HSCs was unique, the majority of PLT- and PEM-HSC–selective chromatin was shared between these HSC subtypes, indicating a close regulatory relationship. To explore this further, we used ChromVAR[16] to identify TF motifs with differential accessibility across subtypes. This showed that the same TF motifs were enriched in PLT- and PEM-HSCs compared to MUL-HSCs, with some of these showing higher overall accessibility in PLT-HSCs, whereas MUL-HSCs–specific TF motifs were clearly distinct (Fig. 3a; Supplementary Table 4). To identify transcriptional and epigenetic lineage programming underlying lymphoid and myeloid fate in HSCs we therefore compared PLT- and MUL-HSCs. Previous analysis of lineage-specific gene expression in single HSCs showed that while platelet- and myeloid-lineage gene expression was readily detectable, lymphoid lineage-specific gene expression was essentially absent[17,18]. Using GSEA and previously described lineage-specific gene sets[19] to identify differential lineage priming between PLT- and MUL-HSCs we found expression of platelet lineage-specific genes (MkP signature) to be enriched in PLT-HSCs, and myeloid lineage-specific genes (preGM signature) in MUL-HSCs, predictive of their lineage bias. However, consistent with the general lack of lymphoid-specific gene expression in HSCs, no difference in transcriptional lymphoid priming (CLP signature) was seen between PLT- and MUL-HSCs (Fig. 3b). To investigate a potential role of epigenetic lineage priming in HSC fate-restriction, we first performed ATACseq of progenitor populations committed to platelet (MkP), erythroid (CFU-E), neutrophil/monocyte (NMP) and lymphoid fates (proB-cells, DN3 thymocytes), and identified chromatin regions selectively accessible in each of these lineages. As a general lymphoid-specific signature we used ATAC peaks selectively accessible in both proB-cells and DN3 thymocytes (Fig. 3c; Supplementary Table 5). We determined the proportions of lineage-specific chromatin that was accessible across all HSC subtypes, using the combined peaks from PLT-, PEM- and MUL-HSCs, and found that the majority of MkP-specific chromatin features were accessible in HSCs, myeloid- and lymphoid-specific chromatin showed moderate accessibility, whereas erythroid-

specific chromatin was largely inaccessible (Fig. 3d), consistent with the presence of epigenetic priming of lymphoid gene expression prior to activation of transcription. Quantitative comparison of the accessibility of lineage-specific chromatin using GSEA (focusing on chromatin accessible in HSCs) showed that both myeloid- and lymphoid-specific chromatin was significantly more accessible in MUL-HSCs compared to PLT-HSCs (Fig. 3e). Therefore, epigenetic, rather than transcriptional, lineage priming accurately predicts the differential lineage output from PLT- and MUL-HSCs. To determine the basis for the dissociation between transcriptional and epigenetic lymphoid priming in MUL-HSCs we generated separate lineage-specific chromatin accessibility signatures for promoters and upstream regulatory elements (UREs) (Extended Data Fig. 3a,b, Supplementary Table 6) and used PWMEnrich[20] to perform sequence-based motif-enrichment analysis to examine which of the TF motif identified as differentially accessible between HSC subtypes (Figure 3a) were enriched in lymphoid-specific promoters and enhancers, respectively. This showed that promoter- and URE-enriched TF motifs were clearly distinct (Fig. 4a; Extended Data Fig. 3c,d), and in particular that UREs with lymphoid-specific accessibility were highly enriched for TF motifs (RUNX1-3, Gfi1) preferentially accessible in MUL-HSCs, whereas lymphoid-specific promoters were primarily enriched for motif preferentially accessible in PLT- and PEM-HSCs. This would be consistent with lymphoid-specific UREs, rather than promoters, being involved in epigenetic lymphoid priming in MUL-HSCs. Supporting this, separate GSEA analysis of promoters and UREs showed that lymphoid HSC fate was associated with enhanced accessibility of lymphoid-specific UREs, but not promoters (Fig. 4b). In contrast, both myeloid-specific promoters and UREs showed enhanced accessibility in MUL-HSCs compared to PLT-HSCs (Fig. 4c). To determine if the lymphoid-specific UREs that were selectively accessible in MUL-HSCs predicted the activation of the lymphoid program during HSC differentiation we identified the genes associated with the LymURE GSEA leading edge ATAC peaks (Fig. 4c) (102 genes; Supplementary Table 7). These genes were not differentially expressed between HSC subtypes (Fig. 4d), consistent with the general lack of transcriptional lymphoid priming in HSCs. However, using previously published RNAseq profiling of HSCs and multi-potent progenitors (MPPs)[21] we found that their expression was selectively up-regulated in the MPP4 population (Fig. 4e), the earliest MPP population biased towards lymphopoiesis[22,23]. Further analysis using microarray profiling of haematopoietic progenitors[24] showed that their expression is sustained in fully lymphoid-committed CLPs (Fig. 4f). Analysis of the accessible chromatin associated with LymURE leading edge peaks identified RUNX1- and RUNX3-binding motifs as significantly enriched (Fig. 4g). ATAC-based footprinting analysis confirmed that there was increased occupancy of RUNX binding sites in MUL-HSCs compared to PLT-HSCs (Fig. 4h). Selective epigenetic priming of lymphoid UREs in MUL-HSCs therefore identifies the lymphoid lineage genes primed for activation and expression during initial commitment to a lymphoid fate, with selective access and increased binding of RUNX1 or RUNX3 to these UREs as the putative mechanism. Notably, the preferential accessibility of lymphoid-specific UREs was generated in the absence of expression of TFs involved in B- and T-lineage commitment, such as *Pax5*, *Ebf1*, *Lef1* and *Tcf7* (Extended Data Figure

4a), suggesting that lymphoid fate-restriction is established prior to and independently of lymphoid lineage commitment.

HSC ageing is associated with epigenetic platelet bias that is counteracted by RUNX3.

Ageing is associated with an increased prevalence of PLT-HSCs and a corresponding decrease in MUL-HSCs[18]. To validate the epigenetic signatures in native haematopoiesis we therefore performed bulk RNAseq and ATACseq on HSCs isolated from young (2-months old) and old (24-months old) mice and used GSEA to compare young and old HSCs using the PLT- and MUL-HSC–specific gene expression and chromatin signatures. This analysis showed highly significant enrichment of both epigenetic and transcriptional PLT-HSC signatures in old HSC, and conversely enrichment of MUL-HSC signatures in young HSCs (Extended Data Figure 4b,c), showing that the HSC subtype-specific signatures are present in native haematopoiesis, and can capture alterations to the composition of the HSC population. As the MUL- and PLT-HSC chromatin signatures showed very significant changes during ageing we performed single cell ATACseq of young and old HSCs to determine if these signatures could be used to computationally identify and quantify HSC subtypes. We quantified accessible PLT-, PEM- and MUL-HSC–specific chromatin in each cell and assigned them to the subtype with the highest proportion of open chromatin (Figure 5a), reproducing the shift in PLT-HSC and MUL-HSC prevalence observed in previous single cell transplantation studies[4,18,25]. Overall, the ratio between PLT- and MUL-HSCs changed 7.4-fold during ageing (Figure 5b). Deconvoluting bulk ATAC analysis of young and old HSCs using Cibersort[26] showed a similar change in the ratio between the PLT- and MUL-HSC ATAC signatures during ageing (Figure 5c). The ATAC signatures identified in HSC clones generated by single cell transplantation could therefore be used to assess the composition and lineage bias of native HSC populations using both bulk and single cell chromatin profiling. During ageing the MPP compartment is skewed away from lymphoid-biased MPP4s, which are produced by MUL-HSC, but not PLT- or PEM-HSC, and towards Mk/E-biased MPP2s which are the only MPP population generated by PLT-HSCs[4,27]. This was paralleled by increased accessibility of MkP-specific chromatin and decreased accessibility of lymphoid-specific chromatin, including LymUREs, in old mice (Figure 5d). As observed when comparing PLT- and MUL-HSCs, the preferentially accessible LymUREs were enriched for RUNX binding motifs (Figure 5e), which by footprinting analysis showed higher occupancy in young HSCs (Figure 5f), again consistent with RUNX factor binding being rate-limiting for lymphoid chromatin accessibility in HSCs. We therefore examined the gene expression of RUNX factors present in haematopoietic cells using the RNAseq data set. No difference in *Runx1* and *Runx2* expression was observed between young and old HSCs (Extended Data Fig. 5a) or across HSC subtypes (Extended Data Fig. 5b). In contrast, *Runx3* was expressed at a higher level in MUL-HSCs compared to both PEM- and PLT-HSCs (Figure 6a), and in young compared to old HSCs (Figure 6b). To test if restoring RUNX3 expression would increase LymURE accessibility we ectopically expressed *Runx3* in old HSCs (CD45.2 allotype) by *ex vivo* lentiviral transduction (Extended Data Fig. 5c) followed by transplantation into lethally irradiated recipients (CD45.1 allotype) along with CD45.1 total

bone marrow cells to provide haematopoietic rescue, and re-isolated the transduced HSCs after 14 weeks for chromatin profiling (Extended Data Fig. 5d,e). Comparison of *Runx3* and mock transduced HSCs by ATACseq showed that *Runx3* expression increased lymphoid-specific chromatin and LymURE accessibility (Figure 6c), and at the same time decreased accessibility of MkP-specific chromatin (Figure 6d). The LymURE leading edge from this comparison was also enriched for RUNX binding sites (Figure 6e), and *Runx3*-expressing HSCs showed increased accessibility of RUNX3 motifs (Figure 6f). Finally, ectopic RUNX3 expression in old HSCs re-balanced the old MPP compartment towards MPP4 and away from MPP2 production (Figure 6g) consistent with RUNX3 playing a role in promoting accessibility of lymphoid chromatin in MUL-HSCs and subsequent generation of lymphoid primed MPP4 progenitors. Consistent with previous reports[28,29], *Runx3* expression led to a general suppression of mature lineage outputs from HSCs, and it was therefore not possible to determine if restoring LymURE accessibility improved lymphopoiesis or decreased platelet bias of old HSC output.

Progenitors generated by PLT-HSCs have accelerated kinetics of platelet formation.

Both transcriptional and epigenetic platelet linage priming was higher in PLT-HSCs compared to MUL-HSCs (Fig. 3b,e), and involved increased chromatin access to both platelet lineage-specific UREs and promoter regions (Fig. 4b,c). Platelet generation from HSCs is generally thought to occur through a series of progenitors that include MPP2s, preMegEs and MkPs[7,12]. However, PLT-HSCs generate fewer MPP2s compared to MUL-HSCs, and only sparsely populate the MEP and MkP populations[4], indicating that PLT-HSCs use a distinct or accelerated differentiation pathway compared to MUL-HSC. Using the *Gata1*-EGFP reporter to prospectively isolate *Gata1*-expressing (*Gata1*-EGFP+ or GE+) cells we previously found that within the HSC-MPP compartment expression of *Gata1* defines both loss of HSC self-renewal and commitment to a *Gata1*-expressing (i.e. megakaryocyte/erythroid/basophil/eosinophil/mast cell) fate[30]. Examination of the HSC-MPP compartment in PLT-, PEM- and MUL-HSC clones using the combined *Gata1*-EGFP and *Vwf*-tdTomato reporters showed that GE+ progenitors were present within the phenotypic HSC compartment and predominant in the MPP2 populations (Fig. 7a,b). However, whereas GE– PLT-HSCs and the GE+ progenitors they generate were predominantly *Vwf*-tdTomato+ (VT+), GE– MUL-HSCs and the derived GE+ MPP2s were predominantly VT–, with PEM-HSC clones showing an intermediate phenotype. Overall, this indicated that MUL-HSCs generate platelets via GE+VT– MPP2s and the classical cellular pathway, whereas PLT-HSCs generate GE+VT+ MPP2s, and subsequently platelets through a more direct route (Fig 7c). A corollary of this model is that the GE+VT+ and GE+VT– MPP2 populations are functionally and molecularly distinct. In steady-state haematopoiesis, both GE+VT+ and GE+VT– populations are present within the phenotypic HSC and MPP2 compartments (Fig. 7d,e). If GE+VT+ and GE+VT– MPP2s are derived from PLT- and MUL-HSCs, respectively, in native haematopoiesis they would likely display molecular properties associated with their parental HSCs. To address this we performed ATACseq on GE+VT+ and GE+VT– MPP2s, identifying 1336 differentially accessible peaks (Fig. 8a,b). Compared to HSCs (Fig. 4d), GE+ MPP2s showed higher overall

accessibility of Meg and Ery signatures and decreased accessibility of Myl and Lym signatures (Fig. 8c), consistent with commitment towards megakaryocyte/erythroid fates. Comparing the accessibility of lineage-specific chromatin between GE+VT+ and GE+VT– MPP2s using GSEA showed that both Meg promoter and MegURE signatures were significantly more accessible in GE+VT+ MPP2s compared to GE+VT– MPP2s, whereas MylUREs and promoters, as well as LymUREs (but not Lym promoters) showed higher accessibility in GE+VT– MPP2s (Fig. 8d,e), the same pattern observed when comparing PLT- and MUL-HSCs. These results are therefore consistent with GE+VT+ and GE+VT– MPP2s being to a significant extent derived from PLT- and MUL-HSCs, respectively, and inheriting the lineage-specific chromatin accessibility patterns of the parental HSCs. *In vitro*, the more HSC-like GE– MPP2 sub-populations predominantly generated mixed megakaryocyte-myeloid colonies indicative of multi-potency, whereas GE+VT+ and GE+VT– MPP2s both generated predominantly megakaryocyte colonies *in* vitro, and did not differ in their level of megakaryocyte-lineage commitment (Figure 8f). However, the distinct population of the Mk progenitor hierarchy by the parental PLT- and MUL-HSCs raised the possibility that the kinetics of platelet formation from the two MPP populations was different. As accelerated kinetics of platelet formation by PLT-HSCs could negatively influence their platelet output, as it would allow fewer transit amplifying cell divisions, we compared engraftment of the HSCs and platelet compartments between MUL- and PLT-HSCs in single HSC-transplanted mice, which showed that MUL-HSCs on average generated ~6-fold more platelets per HSC (Fig. 8g).
In order to measure the time required for GE+ progenitors derived from distinct HSC subtypes to differentiate, we generated a *Gata1*-CreERT2 knock-in allele, where Tamoxifen (TAM) inducible CreERT2 is expressed from the *Gata1* locus. To maintain *Gata1* function, the knock-in was made in the 3'UTR, using an IRES to drive CreERT2 translation (GCET2 allele; Extended Data Fig. 6a). In the presence of a Cre-conditional *Rosa26*-tdTomato reporter (Ai9 allele[31]), activation of *Gata1*-CreERT2 by TAM administration quantitatively lineage traced platelet-forming progenitors leading to similar quantitative Tomato labelling of peripheral blood platelets (Extended Data Fig. 6b). Once these progenitors have differentiated, *Rosa26*-tdTomato+ (RT+) platelets are rapidly cleared from the blood, due to the short half-life of circulating platelets[32], and replaced by RT– platelets generated from GE+ progenitors that were not formed at the time of tamoxifen treatment, with turnover complete after 20 days (Extended Data Fig. 6c,d). By measuring the kinetics by which RT+ platelets are replaced by RT– platelets, the speed with which the RT-labelled GE+ progenitors differentiate can therefore be measured. In order to compare the differentiation kinetics of PLT- and MUL-HSCs, we performed single cell transplantation to generate PLT- and MUL-HSCs clones. For this purpose, the GCERT2/Ai9 line was crossed to the GE line to generate GCET2/GE/Ai9 mice. This was done to efficiently EGFP-label platelet output, so that the GCERT2/Ai9/GE-derived platelets could be identified in single HSC-transplanted mice, regardless of whether they are RT+ or RT–. After identification of GCRET2/Ai9/GE-derived PLT- and MUL-HSC clones we performed TAM induction of the GCET2 transgene, and measured the proportion of GE+ platelets that were also RT+ over time (Fig. 8h; Extended Data Fig. 6e). This showed that platelets generated by GE+ progenitors derived from PLT-

HSCs were replaced with a median time of 12.2 days, whereas replacement by MUL-HSCs–derived GE+ progenitors required a median 14.7 days, demonstrating more rapid platelet differentiation kinetics of VT+GE+ MPP2s derived from PLT-HSCs, compared to VT–GE+ MPP2s derived from MUL-HSCs. Platelet generation from PLT- and MUL-HSCs therefore occurs via molecularly and functionally distinct cellular pathways, with accelerated platelet generation a distinct physiological function of PLT-HSCs.

Discussion.

In conclusion, by systematic transcriptome and chromatin accessibility profiling we show that HSCs with defined fate-restriction are transcriptionally and epigenetically distinct, and that their lineage output is accurately predicted by the accessibility of lineage-specific UREs. This allows transcriptionally silent priming of the lymphoid lineages in MUL-HSCs, with RUNX3 identified as a putative mediator. Such epigenetic priming as a latent predictor of stem cell fate is a concept that may apply to other haematopoietic stem- and progenitor cell populations, and to multipotent tissue stem cells in general. Furthermore, decreased LymURE priming in HSCs was observed during ageing, and may contribute to the decline in lymphoid output from aged HSCs, and the associated decline in adaptive immunity. Runx3 expression in old HSCs increased LymURE accessibility and lymphoid-primed MPP4 output. However, while Runx3 may play an important role in establishing lymphoid fate in HSCs, additional factors are likely to be involved in HSC fate restriction, and further investigation is clearly warranted. In addition, we find that platelet-lineage fate-restriction of PLT-HSCs involves elevated transcriptional and epigenetic platelet-lineage priming compared to MUL-HSCs, allowing PLT-HSCs to produce selectively platelet-primed MPP2 progenitors that generate platelets with faster kinetics than MUL-HSC–derived MMP2s. This identifies molecularly and functionally distinct cellular pathways of platelet generation, a concept supported by recent single cell transcriptome-based modelling that identified a putative direct pathway from HSCs to MkPs, which shows resistance to 5-fluorouracil–induced myeloablation[33]. This accelerated platelet differentiation of PLT-HSC–derived progenitors exemplifies how stem cell diversity allows expedited production of short-lived mature cell types without compromising multi-lineage differentiation, potentially providing resilience against tissue injury. Therefore, epigenetic priming predicts both HSC fate-restriction and the trade-off between speed and quantity during platelet differentiation from HSCs. These findings highlight the importance of the epigenome in cellular fate decisions, and provide a paradigm for the study of multi-potent stem cell populations of other tissues, including intestine[34], brain[35] and skin[36].

Acknowledgements.

We thank the WIMM FACS facility for assistance with cell sorting, and the WIMM CBRG for computational support. This research was funded in whole or in part by BBSRC Project Grant BB/V002198/1 and MRC Unit Grant MC_UU_12009/7 to C.N., and MRC Unit Grant MC_UU_12009/5 and Swedish Research Council Grant 538-2013-8995 to S.E.J. For the purpose of

 The WIMM FACS Core Facility is supported by the MRC HIU, MRC MHU (MC_UU_12009), NIHR Oxford BRC and the John Fell Fund (131/030 and 101/517), the EPA fund (CF182 and CF170) and by WIMM Strategic Alliance awards (G0902418 and MC_UU_12025).

Author contributions.

C.N. and S.E.J. conceived the study and supervised experiments, Y.M., J.C., R.D., X.R., B.Z., A.G. and S.V. performed the experiments, Y.M., J.C. and C.N. analysed the data, Y.M., B.Z., S.T. and C.N. performed the bioinformatics analysis, and C.N. and Y.M. wrote the manuscript.

Competing interests.

The authors have no financial and non-financial competing interests.

Figure legends.

Figure 1 | Fate-restricted HSC clones are internally coherent and molecularly distinct.

a) Experimental workflow for profiling fate-restricted HSCs (defined as Lin–Sca-1+c-Kit+CD150+CD48+CD34–Vwf+; Extended Data Fig. 1a) by single cell RNAseq.

b) Mean peripheral blood reconstitution over time of platelets, erythrocytes, myeloid, B and T cells in single-HSC transplanted mice used for scRNAseq. Independent animals: PLT-HSC: N=2. PEM-HSC: N=2. MUL-HSC: N=2.

c) Importance scores of the top 5000 classifier genes identified by random forest analysis of single HSC transcriptomes. PLT-HSC: 177 cells from 2 mice; PEM-HSC: 118 cells from 2 mice; MUL-HSC: 163 cells from 2 mice. The 10 genes with highest importance score are shown. HSCs were defined as Lin–Sca-1+c-Kit+CD150+CD48–*Gata1*-EGFP–.

d) HSC single cell transcriptomes from (c) were clustered by UMAP using the top 500 genes from (c). HSC subtypes and sample identity are indicated by point colour and shape, respectively.

Figure 2 | Identification of transcriptional and epigenetic programs of fata-restricted HSCs.

a) Heatmap of genes identified as differentially expressed between HSC subtypes ($P<0.05$; log2(fold change)>0.5) using bulk RNAseq. Each sample represents a distinct, single HSC-derived clone. Regulons are defined by pairwise comparing of HSC subtypes and identification of genes selectively upregulating in a single subtype (e.g. the PLT regulon is defined by PLT>PEM and PLT>MUL) or up-regulated in two subtypes compared to the third subtype (e.g. the PLT+PEM regulon is defined by PLT>MUL and PEM>MUL). Biological replicates: PLT-HSC: N=3. PEM-HSC: N=6. MUL-HSC: N=4. Scale shows row z-score of transformed gene expression levels.

b) Heatmap of chromatin features differentially accessible between HSC subtypes ($P<0.05$; log2(fold change)>0.5) using bulk ATACseq, using the same single HSC-derived clones and regulon definition

as in (a). Scale shows row z-score of transformed chromatin accessibility levels. Biological replicates: PLT-HSC: N=3. PEM-HSC: N=6. MUL-HSC: N=4.

c) Spearman correlation of HSC clones using the differentially expressed genes from (e). Samples are ordered by hierarchical clustering. The inset shows the means of the correlation coefficients of all pairwise comparisons between the indicated HSC subtypes. Biological replicates: PLT-HSC: N=3. PEM-HSC: N=6. MUL-HSC: N=4.

d) Spearman correlation of HSC clones as in (g) using the differentially accessible chromatin features from (f). Biological replicates: PLT-HSC: N=3. PEM-HSC: N=6. MUL-HSC: N=4.

e) Comparison of PLT- and MUL-HSCs using GSEA analysis of HSC/MPP-associated gene sets sets[6,11,12,13,14] (left panel) and MSigDB hallmark gene sets (right panel). Gene sets with P<0.05 and normalised enrichment score (NES)>1 or <-1 are shown. Biological replicates: PLT-HSC: N=3. PEM-HSC: N=6. MUL-HSC: N=4.

Figure 3 | HSC fate is predicted by epigenetic priming.

a) Heatmap of transcription factor (TF) motifs differentially enriched in HSC subtype regulons. Differentially accessible motifs were identified using chromVar. The values shown are deviation z-scores for motifs significantly up-regulated in each regulon (P<0.05). Biological replicates: PLT-HSC: N=3. PEM-HSC: N=6. MUL-HSC: N=4.

b) Comparison of lineage-specific gene expression in PLT- and MUL-HSCs by GSEA analysis using gene sets from haematopoietic progenitors committed to a lymphoid (CLP: common lymphoid progenitor), myeloid (preGM: pre-granulocyte/macrophage progenitor) or megakaryocyte/platelet fate (MkP: megakaryocyte progenitor). The normalised enrichment score (NES) and P-value are shown for each comparison. Biological replicates: PLT-HSC: N=3. MUL-HSC: N=4. ES: enrichment score. RLM: rank list metric.

c) Heatmap of chromatin accessibility signatures specific for MkPs (Meg signature), CFU-Es (Ery), NMPs (Myl) and proB/DN3 cells (Lym). Each signature contains the top 2000 features identified as selectively accessible in each of the 4 cell populations/combined cell populations using DESeq2 (Padj<0.05 and |log2(fold change)| >1). N=3 biological replicates for all populations. Scale shows row z-score of transformed chromatin accessibility levels.

d) The number of lineage-specific peaks from (c) detected in HSC populations for each of the 4 signatures. For this purpose the union of peaks from all HSC subtypes was used. Differences in the level of accessibility was analysed using the $\chi 2$-test (one-sided, df=1), and P-values are shown for the key comparisons.

e) Comparison of the accessibility of lineage-specific chromatin features from (c) between PLT- and MUL-HSCs using GSEA. The normalised enrichment score (NES) and P-value is shown for each comparison. PLT-HSC: N=3. MUL-HSC: N=4.

Figure 4 | Lymphoid gene expression during HSC commitment is driven by primed enhancers.

a) Enrichment of HSC subtype-enriched TF motifs in lymphoid-specific promoters and UREs using PWMEnrich. P-values <0.00001 were considered significant.

b) Comparison of accessibility of lineage-specific promoter signatures from Extended Data Fig. 3a between PLT- and MUL-HSCs using GSEA of bulk ATACseq. The normalised enrichment score (NES) and P-value are shown. Biological replicates: PLT-HSC: N=3. MUL-HSC: N=4. ES: enrichment score. RLM: rank list metric.

c) Comparison of the accessibility of lineage-specific URE signatures from Extended Data Fig. 3b between PLT- and MUL-HSCs using GSEA of bulk ATACseq. Normalised enrichment score (NES) and P-value are shown. Leading edge is indicated. Biological replicates: PLT-HSC: N=3. MUL-HSC: N=4. ES: enrichment score. RLM: rank list metric.

d) Whisker plots comparing expression of genes associated with LymURE leading edge features in HSC subtypes using ssGSEA. P-values from two-tailed Welch's t-test; no adjustment for multiple comparisons. The box centres at median and ranges from Q1 to Q3. Whiskers extend to 1.5x IQR. Biological replicates: PLT: N=3. PEM: N=6. MUL: N=4.

e) Plot as in (d) comparing the expression LymURE leading edge-associated genes between HSCs and multi-potent progenitor populations using ssGSEA. P-values show comparison to HSCs (two-tailed Welch's t-test; no adjustment for multiple comparisons). The box centres at median and ranges from Q1 to Q3. Whiskers extend to 1.5x IQR. Biological replicates: HSCs: N=4; MPP populations: N=3.

f) Plot as in (d) comparing the expression LymURE leading edge-associated genes between the indicated lineage-committed progenitor populations using ssGSEA. Mean value indicated. N=2 biological replicates for all populations.

g) Enrichment of HSC subtype-enriched TF motifs in LymURE leading edge peaks calculated using PWMEnrich. P-values <0.00001 were considered significant. Dot size indicates motif enrichment.

h) TF motifs differentially occupied between MUL- and PLT-HSC in LymURE genomic regions by footprinting analysis with the TOBIAS package. Biological replicates: PLT-HSC: N=3. MUL-HSC: N=4.

Figure 5 | Epigenetic priming identifies increased abundance of PLT-HSCs in aged mouse bone marrow.

a) The percentage of computationally defined PLT-, PEM- and MUL-HSC in HSCs (defined as LSKCD150+CD48-CD34-*Vwf-EGFP*+) in young and old bone marrow by analysis of scATAC data with subtype-specific chromatin signatures. Young: 332 single cells from 2 independent experiments. Old: 1060 single cells from 3 independent experiments. Statistical analysis was performed using the χ^2 test (one-sided, df=2).

b) The ratio of computationally defined PLT-HSC vs MUL-HSC in young and old *Vwf*+ HSCs in (a).

c) Young and old HSCs were profiled by bulk ATAC and deconvoluted for the abundance of PLT- and MUL-HSC using CIBERSORT. The ratio of PLT-HSC vs MUL-HSC in young and old HSC pool was calculated by dividing the percentage of PLT-HSC by the percentage of MUL-HSC predicted by

CIBERSORT. Statistical analysis was performed using two-tailed Welch's t-test (N=4 biological replicates for both groups). Mean values with SD are shown.

d) Comparison of the accessibility of Meg-, Lym- and LymURE-specific chromatin signatures between young and old HSCs using GSEA. The normalised enrichment score (NES) and P-value are shown for each comparison. N=4 for both groups. ES: enrichment score. RLM: rank list metric.

e) Enrichment of HSC subtype-enriched TF motifs in LymURE leading edge peaks from young vs old GSEA (panel d) calculated using PWMEnrich. TF motifs with P-value < 0.00001 were defined as significantly enriched. Dot size indicates the breadth of motif enrichment.

f) TF motifs that are differentially occupied in LymURE regions between young and old HSCs in footprinting analysis using TOBIAS package. N=4 for both groups.

Figure 6 | Ectopic expression of *Runx3* rescues epigenetic lymphoid priming and corrects Meg-bias in aged HSCs.

a) The expression level of *Runx3* in MUL-, PEM- and PLT-HSC. The RPKM value from bulk RNAseq is shown. P-values from two-tailed Welch's t-test. Biological replicates: PLT: N=3. PEM: N=6. MUL: N=4. Mean values with SD are shown. No adjustment for multiple comparisons.

b) Expression level of *Runx3* in young and old HSCs from bulk RNAseq is shown. Statistical analysis was performed using two-tailed Welch's t-test. Biological replicates: Young: N=4. Old: N=3. Mean values with SD are shown.

c) GSEA analysis of the accessibility of Lym- and LymURE-specific chromatin signatures in HSCs transduced with mock or *Runx3*-expressing lentivirus and transplanted into lethally irradiated recipients. Control: N=4. Runx3: N=3. ES: enrichment score. RLM: rank list metric.

d) GSEA analysis of the accessibility of Meg-specific chromatin signatures in donor-derived HSCs transduced with mock or *Runx3*-expressing lentivirus and transplanted into lethally irradiated recipients. Control: N=4. Runx3: N=3. ES: enrichment score. RLM: rank list metric.

e) Enrichment of HSC subtype-enriched TF motifs in LymURE leading edge peaks from *Runx3* vs control GSEA (panel c) calculated using PWMEnrich. TF motifs with P-values < 0.00001 were defined as significantly enriched. Dot size indicates the breadth of motif enrichment.

f) Whisker plots showing the RUNX3 motif accessibility in donor-derived HSCs that were transduced with mock or RUNX3-expressing lentivirus and transplanted into lethally irradiated recipients. Motif deviation scores from chromVAR analysis were used for differential accessibility comparison. Statistical analysis was performed using two-tailed Welch's t-test. Biological replicates: Control: N=4. Runx3: N=3. The box centres at median and ranges from Q1 to Q3. Whiskers extend to 1.5x IQR.

g) Ratio of CD45.2+hCD2+ MPP2, MPP3 and MPP4 reconstituted by old HSCs transduced with mock or RUNX3-expressing lentivirus. Statistical analysis was performed using two-tailed Welch's t-test (N=9 biological replicates for both groups; 2 independent experiments). Mean values with SD are shown.

Figure 7 | Distinct MPP2 pathways arise from PLT- and MUL-HSCs.

a) Representative flow cytometry plots showing VT and GE expression in the phenotypic HSC and MPP2 compartments of donor (CD45.2) cells in CD45.1 mice transplanted with single CD45.2 HSCs with the indicated fate-restriction. The percentage of cells in each quadrant is shown.

b) Mean value of the quadrant population percentages defined in (a). Biological replicates: PLT-HSC: N=3; PEM-HSC: N=6; MUL-HSC: N=5. Mean values with SD are shown.

c) Model of the distinct cellular pathways of platelet production emanating from PLT- and MUL-HSCs, respectively.

d) Representative flow cytometry plots showing *Vwf*-tdTomato and *Gata1*-EGFP expression in the phenotypic HSC and MPP2 compartments of native mouse bone marrow. The percentage of cells in each quadrant is shown.

e) Mean value of the gated populations defined in (d). N=6 biological replicates. Mean values with SD are shown.

Figure 8 | Accelerated platelet production by PLT-HSC–derived MPP2s.

a) The number of chromatin features differentially accessible between VT+GE+ and VT–GE+ MPP2 populations ($P<0.05$; log2(fold change)>0.5) using bulk ATACseq (N=4/population). VT+GE+ regulon: VT+GE+>VT–GE+. VT–GE+ regulon: VT–GE+>VT+GE+.

b) Heatmap of chromatin features differentially accessible between VT+GE+ and VT–GE+ MPP2 populations. Regulon is indicated as in (f). Scale shows row z-score of transformed chromatin accessibility levels. N=4 biological replicates/population.

c) The number of lineage-specific peaks (from Fig 2c) detected in the VT+GE+ and VT–GE+ MPP2 populations combined.

d) Comparison of the accessibility of lineage-specific promoter features (from Extended Data Figure 3a) between VT+GE+ and VT–GE+ MPP2 populations using GSEA. The normalised enrichment score (NES) and P-value is shown for each comparison. N=4/population. ES: enrichment score. RLM: rank list metric.

e) Comparison of the accessibility of lineage-specific URE features (from Extended Data Figure 3b) between VT+GE+ and VT–GE+ MPP2 populations using GSEA. The normalised enrichment score (NES) and P-value is shown for each comparison. N=4/population. ES: enrichment score. RLM: rank list metric.

f) *In vitro* lineage potential of the indicated MPP2 subpopulations assessed in single cell culture. Percentages of single cell-derived colonies with the indicated lineage output are shown. MK: megakaryocyte only; GM: granulocyte/macrophage only; Mixed: both megakaryocyte and granulocyte/macrophage output. For all populations, N=3 from 3 independent animals from 3 independent experiments. The number of cultures analysed: VT+GE–: 58; VT–GE+: 336; VT+GE+: 277; VT–GE–: 347. Mean values with SD are shown.

g) The ratio between single HSC-derived platelets and HSCs as a percentage of the entire HSC and platelet populations, respectively, in PLT-HSC (N=8) and MUL-HSC (N=5) transplanted mice from 8 independent experiments. Mean values with SD are shown. P-value (two-sided Wilcoxon ranked-sum test) for PLT- vs MUL-HSC comparison is shown.

h) Tomato+ platelets as a percentage of GE+ platelets in mice reconstituted with single PLT- or MUL-HSCs from GE/GCET2/Ai9 donor mice at the indicated days after Tamoxifen treatment. Biological replicates: PLT-HSC: N=4; MUL-HSC: N=10. 3 independent experiments. Mean values with standard error are shown. P-values were from Wilcoxon ranked-sum test. Vertical lines indicate the median time of replacement of Tomato+ platelets by Tomato– platelets for each HSC subtype.

References.

1 Muller-Sieburg, C. E., Cho, R. H., Thoman, M., Adkins, B. & Sieburg, H. B. Deterministic regulation of hematopoietic stem cell self-renewal and differentiation. Blood 100, 1302-1309 (2002).

2 Dykstra, B. et al. Long-term propagation of distinct hematopoietic differentiation programs in vivo. Cell stem cell 1, 218-229, doi:10.1016/j.stem.2007.05.015 (2007).

3 Yamamoto, R. et al. Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. Cell 154, 1112-1126, doi:10.1016/j.cell.2013.08.007 (2013).

4 Carrelha, J. et al. Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. Nature 554, 106-111, doi:10.1038/nature25455 (2018).

5 Pei, W. et al. Resolving Fates and Single-Cell Transcriptomes of Hematopoietic Stem Cell Clones by PolyloxExpress Barcoding. Cell Stem Cell 27, 383-395 e388, doi:10.1016/j.stem.2020.07.018 (2020).

6 Rodriguez-Fraticelli, A. E. et al. Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis. Nature 583, 585-589, doi:10.1038/s41586-020-2503-6 (2020).

7 Jacobsen, S. E. W. & Nerlov, C. Haematopoiesis in the era of advanced single-cell technologies. Nat Cell Biol 21, 2-8, doi:10.1038/s41556-018-0227-8 (2019).

8 Yu, V. W. C. et al. Epigenetic Memory Underlies Cell-Autonomous Heterogeneous Behavior of Hematopoietic Stem Cells. Cell 167, 1310-1322 e1317, doi:10.1016/j.cell.2016.10.045 (2016).

9 Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc 9, 171-181, doi:10.1038/nprot.2014.006 (2014).

10 Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

11 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, doi:10.1016/j.stem.2015.04.004 (2015).

12 Pietras, E. M. et al. Re-entry into quiescence protects hematopoietic stem cells from the killing effect of chronic exposure to type I interferons. J Exp Med 211, 245-262, doi:10.1084/jem.20131043 (2014).

13 Cabezas-Wallscheid, N. et al. Vitamin A-Retinoic Acid Signaling Regulates Hematopoietic Stem Cell Dormancy. Cell 169, 807-823 e819, doi:10.1016/j.cell.2017.04.018 (2017).

14 Lauridsen, F. K. B. et al. Differences in Cell Cycle Status Underlie Transcriptional Heterogeneity in the HSC Compartment. Cell Rep 24, 766-780, doi:10.1016/j.celrep.2018.06.057 (2018).

15 Tong, J. et al. Hematopoietic stem cell heterogeneity is linked to the initiation and therapeutic response of myeloproliferative neoplasms. Cell Stem Cell 28, 780, doi:10.1016/j.stem.2021.02.026 (2021).

16 Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat Methods 14, 975-978, doi:10.1038/nmeth.4401 (2017).

17 Mansson, R. et al. Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors. Immunity 26, 407-419, doi:10.1016/j.immuni.2007.02.013 (2007).

18 Grover, A. et al. Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells. Nat Commun 7, 11075, doi:10.1038/ncomms11075 (2016).

19 Bereshchenko, O. et al. Hematopoietic stem cell expansion precedes the generation of committed myeloid leukemia-initiating cells in C/EBPalpha mutant AML. Cancer Cell 16, 390-400, doi:10.1016/j.ccr.2009.09.036 (2009).

20 Stojnic, R. & Diez, D. PWMEnrich: PWM enrichment analysis. R package version 4.30.0., doi:10.18129/B9.bioc.PWMEnrich (2021).

21 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, 507-522, doi:10.1016/j.stem.2014.07.005 (2014).

22 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, doi:10.1016/j.stem.2015.05.003 (2015).

23 Adolfsson, J. et al. Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. Cell 121, 295-306, doi:10.1016/j.cell.2005.02.013 (2005).

24 Choi, J. et al. Haemopedia RNA-seq: a database of gene expression during haematopoiesis in mice and humans. Nucleic Acids Res 47, D780-D785, doi:10.1093/nar/gky1020 (2019).

25 Benz, C. et al. Hematopoietic stem cell subtypes expand differentially during development and display distinct lymphopoietic programs. Cell Stem Cell 10, 273-283, doi:10.1016/j.stem.2012.02.007 (2012).

26 Newman, A. M. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol 37, 773-782, doi:10.1038/s41587-019-0114-2 (2019).

27 Young, K. et al. Progressive alterations in multipotent hematopoietic progenitors underlie lymphoid cell loss in aging. J Exp Med 213, 2259-2267, doi:10.1084/jem.20160168 (2016).

28 Wong, W. F. et al. Over-expression of Runx1 transcription factor impairs the development of thymocytes from the double-negative to double-positive stages. Immunology 130, 243-253, doi:10.1111/j.1365-2567.2009.03230.x (2010).

29 Menezes, A. C. et al. RUNX3 overexpression inhibits normal human erythroid development. Sci Rep 12, 1243, doi:10.1038/s41598-022-05371-z (2022).

30 Drissen, R. et al. Distinct myeloid progenitor-differentiation pathways identified through single-cell RNA sequencing. Nat Immunol 17, 666-676, doi:10.1038/ni.3412 (2016).

31 Madisen, L. et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. Nat Neurosci 13, 133-140, doi:10.1038/nn.2467 (2010).

32 Fujii, Y. et al. A novel mechanism of thrombocytopenia by PS exposure through TMEM16F in sphingomyelin synthase 1 deficiency. Blood Adv 5, 4265-4277, doi:10.1182/bloodadvances.2020002922 (2021).

33 Morcos, M. N. F. et al. Fate mapping of hematopoietic stem cells reveals two pathways of native thrombopoiesis. Nat Commun 13, 4504, doi:10.1038/s41467-022-31914-z (2022).

34 Noah, T. K., Donahue, B. & Shroyer, N. F. Intestinal development and differentiation. Exp Cell Res 317, 2702-2710, doi:10.1016/j.yexcr.2011.09.006 (2011).

35 Taupin, P. & Gage, F. H. Adult neurogenesis and neural stem cells of the central nervous system in mammals. J Neurosci Res 69, 745-749, doi:10.1002/jnr.10378 (2002).

36 Blanpain, C. & Fuchs, E. Epidermal homeostasis: a balancing act of stem cells in the skin. Nat Rev Mol Cell Biol 10, 207-217, doi:10.1038/nrm2636 (2009).

Methods.

Animals

All mice were bred and maintained under SPF conditions, and all experimental procedures were performed in accordance with UK Home Office regulations, and were approved by the University of Oxford Medical Sciences Division Animal Welfare and Ethical Review Board under project licenses 30/3359, 30/3103 and PP2240412. *Vwf*-tdTomato/*Gata1*-EGFP (VT/GE) dual reporter mice expressing the CD45.2 allotype were generated by crossing *Vwf*-tdTomato (VT) mice[4] to *Gata1*-EGFP (GE) mice[30] to produce donors for single cell transplantations as previously described[4]. *Gata1*-CreERT2 knock-in (GCET2) mice were generated by Cyagen and crossed with GE mice and Ai9 Rosa26-tdTomato reporter mice[31] to generate donors with three transgenic alleles (GCET2/GE/Ai9) for single cell transplantations and lineage tracing experiments. The GCET2 allele was genotyped by PCR using the following primers in the same reaction: 5'-TCTGCTGGGATCGCCTACAAC, CACACCGGCCTTATTCCAAGC and 5'-GGAGGTAGAGGCAGGAGAATG. The thermal cycles

are: 97°C for 2min; 31 cycles of 94°C for 30sec, 60°C for 30sec, 72°C for 30sec; 72°C for 10min. The expected product of the GCET2 allele is at 156bp and WT at 225bp. All strains were on a C57Bl/6J background. Wild-type (WT) B6.SJL-Ptprc^{a} Pepc^{b}/BoyJ (CD45.1) mice were used as transplantation recipients.

Flow cytometry analysis and cell sorting

For analysis and sorting of HSCs and progenitors, bone marrow (BM) cells were isolated from mice by crushing leg bones into fluorescence-activated cell sorting (FACS) media that consists of PBS (Gibco) supplemented with 5% fetal calf serum (FCS, Gibco) and 2mM ethylenediaminetetraacetic acid (EDTA, Sigma-Aldrich). Isolation of thymocytes was carried out by crushing thymus into PBS with 1% FCS and 2mM EDTA. In some cases, kit-enrichment of BM was performed using anti-mouse CD117 microbeads (Miltenyi Biotec) prior to Fc block. Total BM or kit-enriched cells were incubated in FACS media containing Fc block antibody and subsequently stained with antibody cocktail for 20-30 minutes at 4°C. Samples were washed with media by centrifugation for 5 minutes at 4°C, 500g. Analysis of stained cells was performed on LSRII, LSR Fortessa or LSR X-20 flow cytometers (BD Biosciences). Cell sorting of stained samples was carried out using a FACSAriaII, FACSAriaIII or FACSAria Fusion cell sorter (BD Biosciences). FACS data was collected using BD FACS DIVA software (v9.0). For analysis of mature haematopoietic lineages, peripheral blood (PB) samples were collected from mouse tail vein into lithium heparin-coated microvettes (Sarstedt). PB samples were centrifuged at 100x*g* for 10 minutes at room temperature and the supernatant was collected for platelets analysis. A small aliquot of the red pellet was aliquoted out for analysing erythroid cells. Samples were then incubated 1:1 with 2% w/v Dextran (Sigma-Aldrich), for 30 minutes in a 37°C incubator and the supernatant was pelleted for erythrocyte lysis with ammonium chloride solution (Stem Cell Technologies) for 2 minutes at room temperature before leukocytes were collected by centrifugation. Antibody staining was carried out for 15-20 minutes at 4°C in PBS 1% FCS 2mM EDTA and samples were analyzed on LSRII, LSR Fortessa or LSR X-20 flow cytometers. Antibodies used for flow cytometry analysis and cell sorting are listed in Supplementary Table 8.

Single cell transplantations

Single cell transplantations were performed as previously described[4]. Single *Vwf*+ HSCs (LSKCD150+CD48-CD34-VT+) and CD229– HSCs (LSKCD150+CD48-CD34-CD229-/lo) were sorted from the bone marrow (BM) of adult (8-16 weeks) VT/GE (both males and females) and GCET2/GE/Ai9 mice (females only), respectively. Cells were sorted into Iscove's Modified Dulbecco's Medium (IMDM, Gibco) with 20% BIT-9500 Serum Substitute (Stem Cell Technologies), 100U/mL Penicillin and 0.1mg/mL Streptomycin (100x Pen/Strep, PAA Laboratories), 2mM L-Glutamine (PAA Laboratories) and 0.1mM 2-Mercaptoethanol (Sigma-Aldrich). Sorted single HSCs were combined with 2x10^{5} WT CD45.1 unfractionated BM support cells and transplanted into lethally irradiated (10 Gy, split dose) CD45.1 mice by intravenous injection. The lineage bias of transplanted single HSCs

was determined by 3 consecutive peripheral blood analyses over 16-18 weeks post transplantation, as previously described[4].

Analysis of platelet production *in vivo*

Single CD229– HSCs were sorted from 8-12 weeks old GCET2/GE/Ai9 female mice and transplanted into 8-12 weeks old CD45.1 female mice. Lineage reconstitution patterns were determined by 8-week and 12-week PB readouts. Mice with fate-restricted lineage output were treated with 4mg tamoxifen per day for 2 consecutive days by oral gavage. PB samples were analysed by flow cytometry and the fraction of *Rosa26*-tdTomato+ cells in the donor-derived *Gata1*–EGFP+ platelet population quantified by flow cytometry. The relative platelet output from HSCs for PLT- and MUL-HSCs was calculated from previously published data[4].

In vitro differentiation assays

Megakaryocyte (MK) and granulocyte-macrophage (GM) lineage potentials of MPP2 subsets were determined *in vitro*. MPP2 (defined as LSK CD150+CD48+, and subdivided based on VT and GE expression) from BM of 8-12 weeks old VT/GE female mice were bulk sorted into X-VIVO15 liquid medium with Gentamycin and L-Glutamine (Lonza), supplemented with 10% FCS, 0.1mM 2-Mercaptoethanol and the following cytokines: 50ng/mL mouse stem cell factor (mSCF, PeproTech), 50ng/mL human FLT3 ligand (hFL, Immunex), 50ng/mL human thrombopoietin (hTPO, PeproTech) and 20ng/mL mouse interleukin 3 (mIL-3, PeproTech). Cells were manually plated into Terasaki microplates (Thermo Fisher Scientific) at 20µl/well to ensure single cell was deposited in each well. Culture confluency in Terasaki microplates was scored at day 9 to determine the presence of MK/GM colonies. Colony morphology was subsequently confirmed by cytospin followed by May Grünwald Giemsa staining.

Single cell RNAseq of lineage-restricted HSCs

Single *Vwf*+ HSCs were sorted from 8-16 weeks old male or female VT/GE mice, and transplanted into lethally irradiated 8-12 weeks old sex-matched CD45.1 recipients. Single donor-derived HSCs (LSK CD150+CD48-CD45.2+) were index sorted from BM of platelet-restricted (PLT-), platelet/erythroid/myeloid-restricted (PEM) and multilineage (MUL) mice that received single cell transplantations. Details of the transplanted mice and sorted cell numbers for each clone sorted are shown in Supplementary Table 9. Library preparation was performed using the SMARTer Ultra Low RNA kit for Illumina Sequencing (Clontech) and sequenced on Illummina HiSeq2500 (51bp, single end).

Single cell RNAseq analysis

Short reads (51 bp) from RNA sequencing of mouse HSCs were mapped using Tophat (version 2.0.10) [37] to the mouse genome (mm10) with a supplied set of RefSeq gene model. The mapping parameters '-

g 1' was used to allow one alignment to the reference for a given read. Cells with <500,000 mapped reads, with percent of mapping to the mitochondrial chromosome >10% or with <1,000 detected genes were excluded from further analysis. A total of 458 HSCs fulfilled these criteria. FeatureCounts (v1.6.2)[38] was used to count reads using the RefSeq gene model as a reference. Reads per kilobase of transcript per million mapped reads (RPKM) expression values were calculated using a custom R script, and further normalised into the log2(RPKM) scale. Random forest analysis was performed using R package 'randomForest' (v4.6-14) with 'ntree = 2000'. A model for MUL-, PEM- and PLT-HSCs was trained. The normalized expression values were donor and gender corrected, and the HSC subtype was preserved, using 'removeBatchEffect' function in Limma R package (v3.40.2). Genes were ranked by MeadnDecreaseGini in random forest analysis and top 500 genes were used for UMAP analysis with the 'uwot' R package (v0.1.8).

Bulk RNAseq and ATACseq

Bulk RNAseq was carried out with 100 cells per replicate. For lineage-restricted HSCs, single *Vwf*+ HSCs were sorted from 8-12 weeks old female VT/GE mice and transplanted into lethally irradiated 8-12 weeks old CD45.1 female recipients. Bulk donor-derived GE– HSCs (LSK Flt3-CD150+CD48-CD45.2+GE–) were sorted from single cell transplanted mice reconstituted with BM of PLT-, PEM- and MUL-HSCs. For young and old HSCs, LSK CD150+CD48- BM cells were sorted from female wild type C57BL/6J mice at 2 months (young) or 24 months (old) of age. Details of the transplanted mice and sorted cell numbers for each clone sorted are shown in Supplementary Table 9. Library preparation was performed using the Smart-seq+ as previously described[39]. Briefly, cells were sorted into 4ul lysis buffer containing 0.18% Triton X-100, 4U RNase Inhibitor (Clontech) 2.5mM dNTPs and 2.5uM oligo(dT) (Biomers.net). Reverse transcription was performed using SMARTScribe (Clontech) reserve transcriptase and cDNA amplification was performed using SeqAMP polymerase (Clontech). PCR product was purified with AMPure XP beads (Beckman) and libraries were constructed using the Nextera XT DNA Library Preparation Kit (Illumina). Libraries were sequenced on Illumina NextSeq 500 (76bp, single-end read). Bulk ATACseq was performed using the Omni-ATACseq[40]. For lineage-restricted HSCs, 500 donor-derived GE– HSCs per replicate were sorted from recipients reconstituted with single PLT-, PEM- or MUL-HSCs. For young and old HSCs, 1000 LSK CD150+CD48- BM cells per replicate were sorted from wild type C57Bl/6J female mice at 2 months (young) or 24 months (old) of age. For multipotent progenitors, 500 VT+GE+ and VT–GE+ MPP2 (LSK Flt3-CD150+CD48+) per replicate were sorted from adult (8-12 weeks old) female VT/GE mice. For committed progenitors, 1000 cells per replicate were sorted from adult (8-12 weeks old) female GE mice for following populations. MkP: LKCD150+CD41+, CFU-E: LKCD41-CD16/32-CD150-CD105+, NMP: LKCD41-CD150-CD16/32+GE–, Pro-B: Lin-B220+CD19+cKit+IgM-, DN3: Lin-CD4-CD8-CD44-CD25+. Sorted cells were tagmented using Nextera DNA Library Prep Kit (Illumina) and fragmented DNA was amplified using the NEBNext High-Fidelity 2x PCR Master Mix (New England BioLabs). PCR

products were purified by AMPure XP beads (Beckman) and final libraries were sequenced on Illumina NextSeq 500 (40bp, paired-end read).

Bulk RNAseq analysis

FASTQ files were inspected using FastQC (v0.11.7; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) followed by Nextera adapter sequence removal using trimmomatic (v0.32)[41]. Reads were aligned to the mm9 transcriptome using STAR (v2.6)[42]. Unique reads were counted using featureCounts (v1.6.2)[38] and the mm9 genome. All output files were quality assessed using MultiQC (v1.9)[43]. Read counts were imported into R Studio and differentially expression genes (DEGs) were analysed using DESeq2 (v1.24.0)[44], with cutoffs indicated in the figure legend for each dataset. P-values were calculated using the DESeq2() function. For gene-set enrichment analysis (GSEA), genes were ranked by stat from DESeq2 to generate a single rank-ordered list that was imported into the javaGSEA application (v4.0.1)[45]. Single sample GSEA (ssGSEA)[46] was used to quantify the expression of lymphoid leading-edge genes in HSCs and progenitors.

Bulk ATACseq analysis

FASTQ files were inspected using FastQC (v0.11.7; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) followed by Nextera adapter sequence removal using TrimGalore (v0.5.0; http://www.bioinformatics.babraham.ac.uk/projects/trim_galore). Reads were aligned to the mm9 genome using Bowtie2 (v2.3.5)[47] and the resulting SAM files converted to BAM files using SAMtools (v1.9)[48]. PCR duplicate reads were filtered out using the MarkDuplicates function from the Picard tool (v2.18.17) (http://broadinstitute.github.io/picard). Reads shift and mitochondrial DNA removal were performed using the ATACseqQC R package (v1.8.1)[49]. Peak calling was performed using the MACS2 software (v2.1.2; -q 0.05 --nomodel --shift -100 --extsize 200 --keep-dup=1 --call-summits)[50] to identify open chromatin regions. Peak calling results were converted into a list of open chromatin regions and unique reads mapped into these regions were quantified using featureCounts (v1.6.2)[38]. All output files were quality assessed using MultiQC (v1.9)[43]. Read counts were imported into R Studio and differentially accessible chromatin features (DACs) identified using DESeq2 (v1.24.0)[44], with the cutoffs indicated. Annotation of ATACseq peaks by promoter vs upstream regulatory element (URE) was performed using ChIPseeker (v1.20.0)[51] and a custom script. Promoters were defined as -1kbp to +1kbp of the transcription start site (TSS) and UREs were defined as peaks outside the TSS regions. For GSEA analysis, peaks were ranked by stat from DESeq2 to generate a single rank-ordered list that was imported into the javaGSEA application[45] and P-values calculated using 1000 permutations. Leading edge analysis was performed by inputting the GSEA results into the javaGSEA application leading-edge peaks identified (Supplementary Table 10), and associated genes were defined as genes nearest to the input peaks on the mm9 genome. Motif analysis was performed by chromVAR R package (v1.6.0)[16] and PWMEnrich R package (v4.20.0)[20] using the JASPAR2020 motif

database[52]. In chromVAR, a bed file containing a list of genomic regions was used as the input peak list. Bam files of individual samples were used to calculate the motif accessibility within the input peak regions. Differential motif accessibility between cell types was calculated on motif deviation scores using the differentialDeviations() function. In PWMEnrich, the genomic background motif distributions were first calculated using the mm9 genome and JASPAR2020 motif database using makeBackground() function. Motif enrichment within selective genomic regions was determined by the motifEnrichment() function using the pre-compile background distribution and JASPAR2020 motif. TF footprinting analysis was performed using the TOBIAS python software (v0.12.1)[53]. Individual bam files from the replicates of the same cell type were aggregated into a cell-type specific bam file and was used for Tn5 insertion bias correction with the ATACorrect function. TF footprint scores within selective genomic regions were calculated on the corrected files by the ScoreBigwig function. P-values for differential TF binding were calculated by the BINDetect function. The web-based tool CIBERSORTx was used for the deconvolution analysis[26]. The bulk ATACseq TPM (transcripts per kilobase million) matrix of lineage-restricted HSCs were used to build a signature matrix file (q-value = 0.1). The signature matrix file was then applied to deconvolute bulk ATACseq TPM matrix of young and old HSCs to predict the fraction of HSC subtypes (permutations = 1000).

Single cell ATACseq of young and old HSCs

A plated-based method was used to perform single cell ATAC with modifications[54]. Briefly, ~2500 *Vwf*+ HSCs (LSK CD150+CD48-CD34-Vwf-tdTomato+) from young (2 months) or old (24 months) female VT/GE mice were sorted in bulk into 25µl tagmentation reaction, incubated at 37ºC for 30 minutes and stopped by incubating with 25µl tagmentation stop buffer on ice for 10 minutes. Single tagmented cells were sorted into 96 well plates with lysis buffer and Index Primers. Plates were incubated at 65ºC for 15 minutes and quenched with 10% TWEEN-20. NEBNext High-Fidelity 2x PCR Master Mix (New England BioLabs) was used for library amplification. PCR products were purified by Qiagen MinElute PCR Purification Kit (Qiagen) and size selection on eluates was carried out with 1.8x AMPure XP beads (Beckman). Sequencing was performed on Illumina NextSeq 500 (75bp, pair end). Preprocessing of single cell ATACseq data was carried out in the same manner as bulk data, including adaptor trimming, genome alignment, duplicates removal, reads shift, mitochondrial reads removal and featureCoutns. Cells with ≥ 1000 unique fragments and ≥ 30% featureCounts ratio were selected for downstream analysis. The number of reads mapped to PLT-, PEM- and MUL-HSC specific regions was calculated for each single cell. Regions with ≥ 1 mapped read were defined as open and the ones with 0 read were defined as closed. To annotate single cells with an HSC subtype, the percentages of PLT-, PEM- and MUL-specific regions that are open in each single cell were calculated by dividing the number of open PLT- PEM- and MUL-specific regions by the total number of PLT-, PEM- and MUL-specific regions, respectively. Cells with higher percentage of open PLT-specific regions than PEM- and MUL-specific regions were defined as PLT-HSC. PEM- and MUL-HSC were annotated in the similar manner.

Transplantation of *Runx3*-overexpressing aged HSPCs

Bone marrow were isolated from old (24 months) female VT/GE mice for cKit-enrichment using LS column (Miltenyi Biotec). The mouse *Runx3* cDNA (NM_001369050.1) was obtained from GenScript. The Lentiviral vector used for cloning was a kind gift from Professor Hugh JM Brady from Imperial College London. Transduction of mock or *Runx3*-expressing lentivirus was carried out at MOI 100 in StemSpan SFEM (Stem Cell Technologies) supplemented with 100U/mL Penicillin and 0.1mg/mL Streptomycin, 2mM L-Glutamine, 100mg/ml murine SCF (Peprotech), 100ng/ml human Flt3-Lignad (Peprotech) and 100ng/ml human Thrombopoietin (Peprotech). Cells were spinfected with lentivirus at 20ºC, 700g for 1 hour and cultured in the incubator for another 7 hours. $1x10^5$ transduced cells (CD45.2) were combined with $5x10^5$ WT CD45.1 unfractionated BM support cells and transplanted into lethally irradiated (10 Gy, split dose) 8-12 weeks old female CD45.1 mice by intravenous injection. Sex matched recipients were randomly allocated to experimental groups. 14-16 weeks after transplantation, bone marrow from recipients were isolated to determine the engraftment of transduced aged bone marrow. The percentages of transduced donor HSC (LSK Flt3-CD150+CD48-CD45.2+hCD2+), MPP2 (LSK Flt3-CD150+CD48+CD45.2+hCD2+), MPP3 (LSK Flt3-CD150-CD48+CD45.2+hCD2+) and MPP4 (LSK Flt3-CD150-CD45.2+hCD2+) were determined by FACS. For bulk ATAC, 500 CD45.2+hCD2+ HSCs were sorted from mock or *Runx3*-overexpressing group. The ratio of CD45.2+hCD2+ MPP2, MPP3 and MPP4 was calculated to determine the lineage bias rescue by the ectopic expression of *Runx3*.

Statistics and Reproducibility

Statistical analysis and visualization in this study was performed in R software and Graphpad Prism 9. For normally distributed data two-tailed t-tests were applied with Welch's correction that does not assume equal variance. Normality was tested using the Shapiro-Wilk test. For non-normal data distributions Wilcoxon's ranked-sum test was used. The specific test used is indicated in the legend of the corresponding figure. No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those used in our previous studies[55,56]. Data collection and analysis were not performed blind to the conditions of the experiments.

Data availability.

All raw and processed high throughput sequencing data generated in this study have been deposited in the Gene Expression Omnibus (GEO) database under accession number GSE188119. Public datasets used in this study include the JASPAR2020 motif database[52], RNA sequencing of HSC/MPPs[21], available in the ArrayExpress database (http://www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-2262, and haematopoietic progenitors[24], available in the Haemospere database (https://www.haemosphere.org/). All other data generated in this study are available upon reasonable request. Requests for materials and manuscript correspondence should be directed to C.N.

Methods-only references.

37 Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36, doi:10.1186/gb-2013-14-4-r36 (2013).

38 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).

39 Rodriguez-Meira, A. *et al.* Unravelling Intratumoral Heterogeneity through High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing. *Mol Cell* **73**, 1292-1305 e1298, doi:10.1016/j.molcel.2019.01.009 (2019).

40 Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**, 959-962, doi:10.1038/nmeth.4396 (2017).

41 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).

42 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

43 Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048, doi:10.1093/bioinformatics/btw354 (2016).

44 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

45 Subramanian, A., Kuehn, H., Gould, J., Tamayo, P. & Mesirov, J. P. GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics* **23**, 3251-3253, doi:10.1093/bioinformatics/btm369 (2007).

46 Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108-112, doi:10.1038/nature08460 (2009).

47 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

48 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

49 Ou, J. H. *et al.* ATACseqQC: a Bioconductor package for post-alignment quality assessment of ATAC-seq data. *Bmc Genomics* **19**, doi:10.1186/s12864-018-4559-3 (2018).

50 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

51 Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383, doi:10.1093/bioinformatics/btv145 (2015).

52 Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **48**, D87-D92, doi:10.1093/nar/gkz1001 (2020).

53 Bentsen, M. *et al.* ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat Commun* **11**, 4267, doi:10.1038/s41467-020-18035-1 (2020).

54 Chen, X., Miragaia, R. J., Natarajan, K. N. & Teichmann, S. A. A rapid and robust method for single cell chromatin accessibility profiling. *Nat Commun* **9**, 5345, doi:10.1038/s41467-018-07771-0 (2018).

55 Di Genua, C. *et al.* C/EBPalpha and GATA-2 Mutations Induce Bilineage Acute Erythroid Leukemia through Transformation of a Neomorphic Neutrophil-Erythroid Progenitor. *Cancer Cell* **37**, 690-704 e698, doi:10.1016/j.ccell.2020.03.022 (2020).

56 Valletta, S. *et al.* Micro-environmental sensing by bone marrow stroma identifies IL-6 and TGFbeta1 as regulators of haematopoietic ageing. *Nat Commun* **11**, 4075, doi:10.1038/s41467-020-17942-7 (2020).

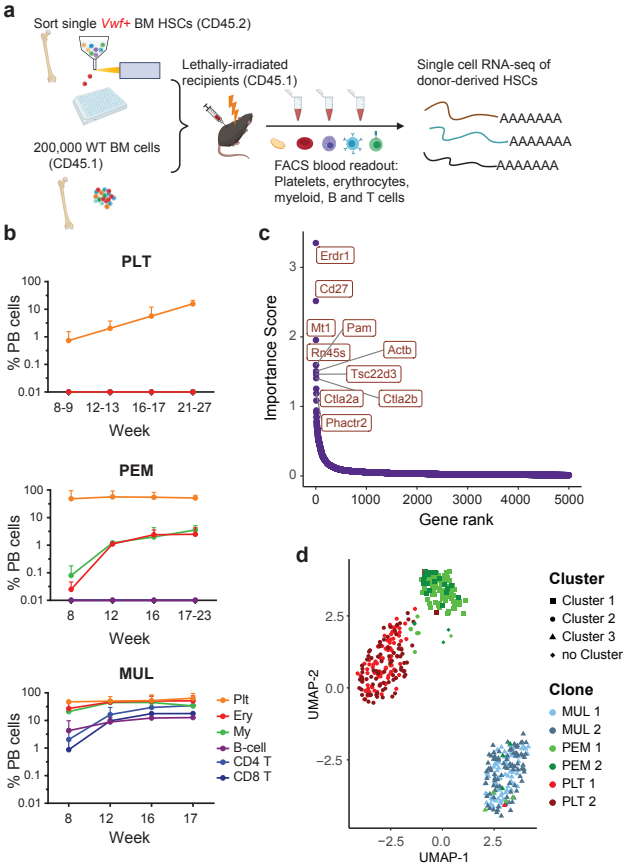
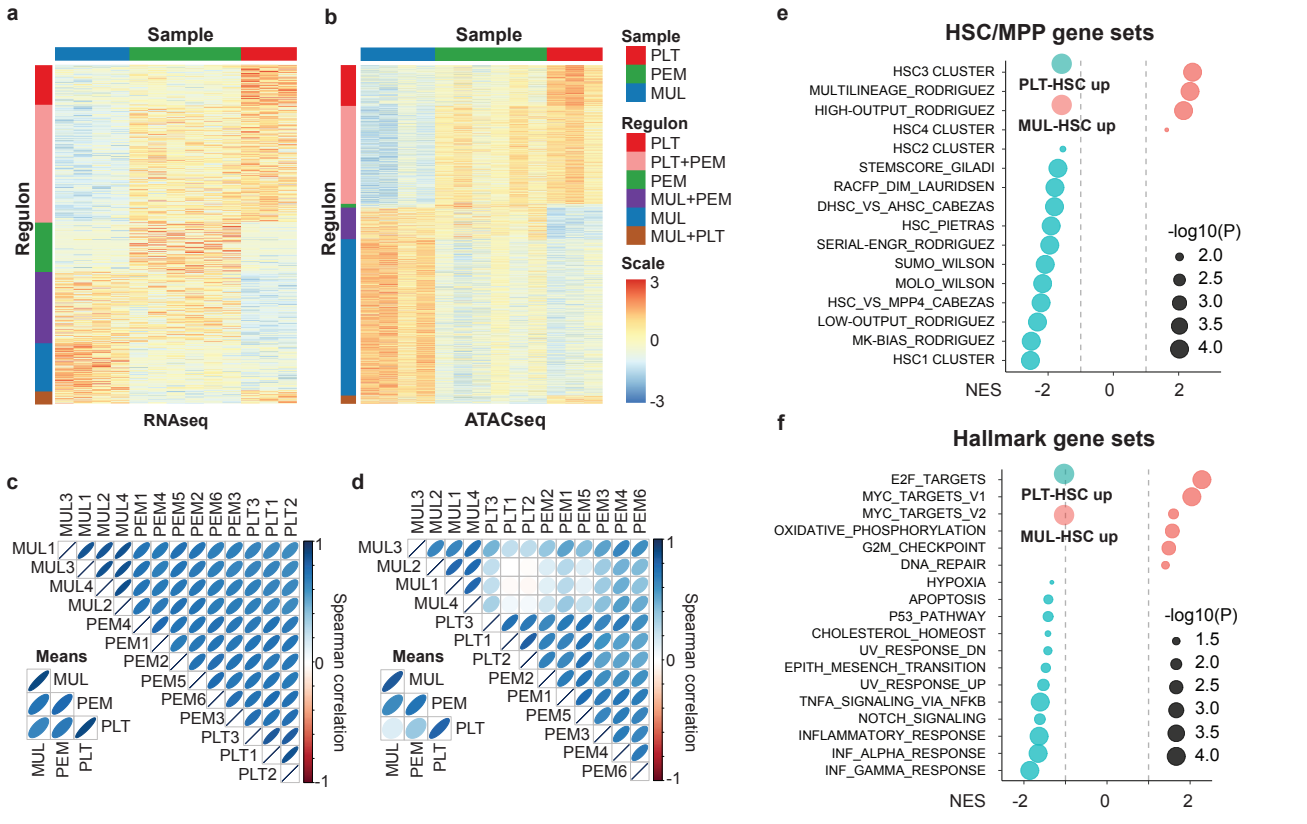
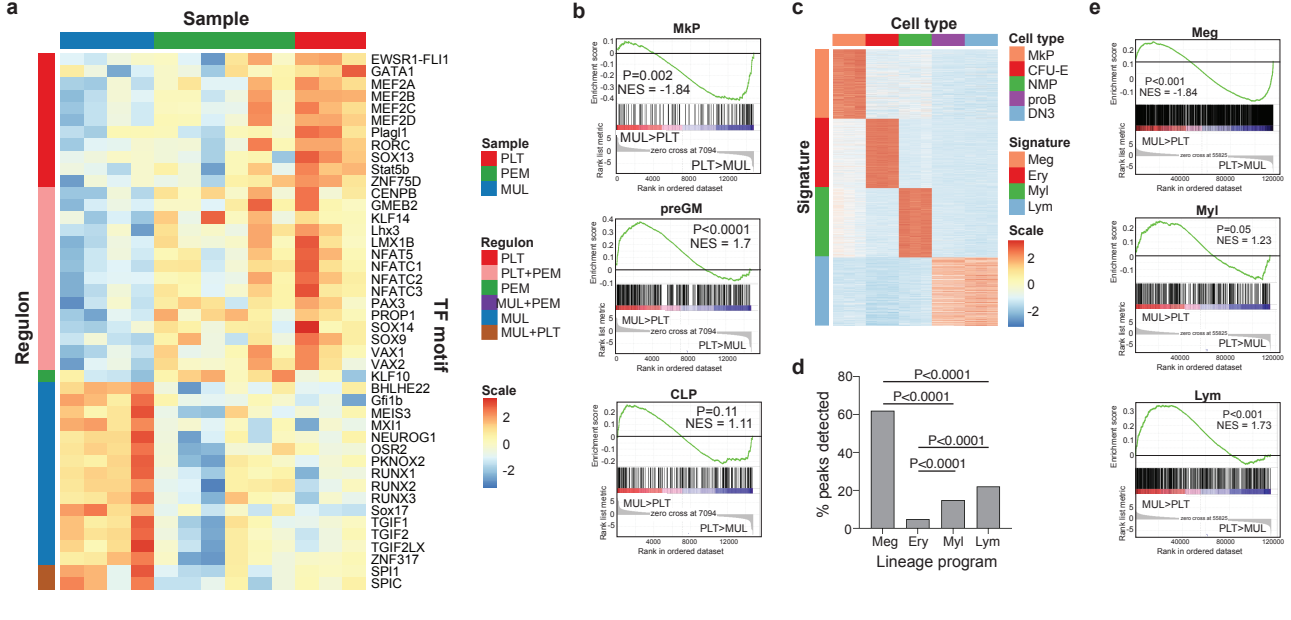
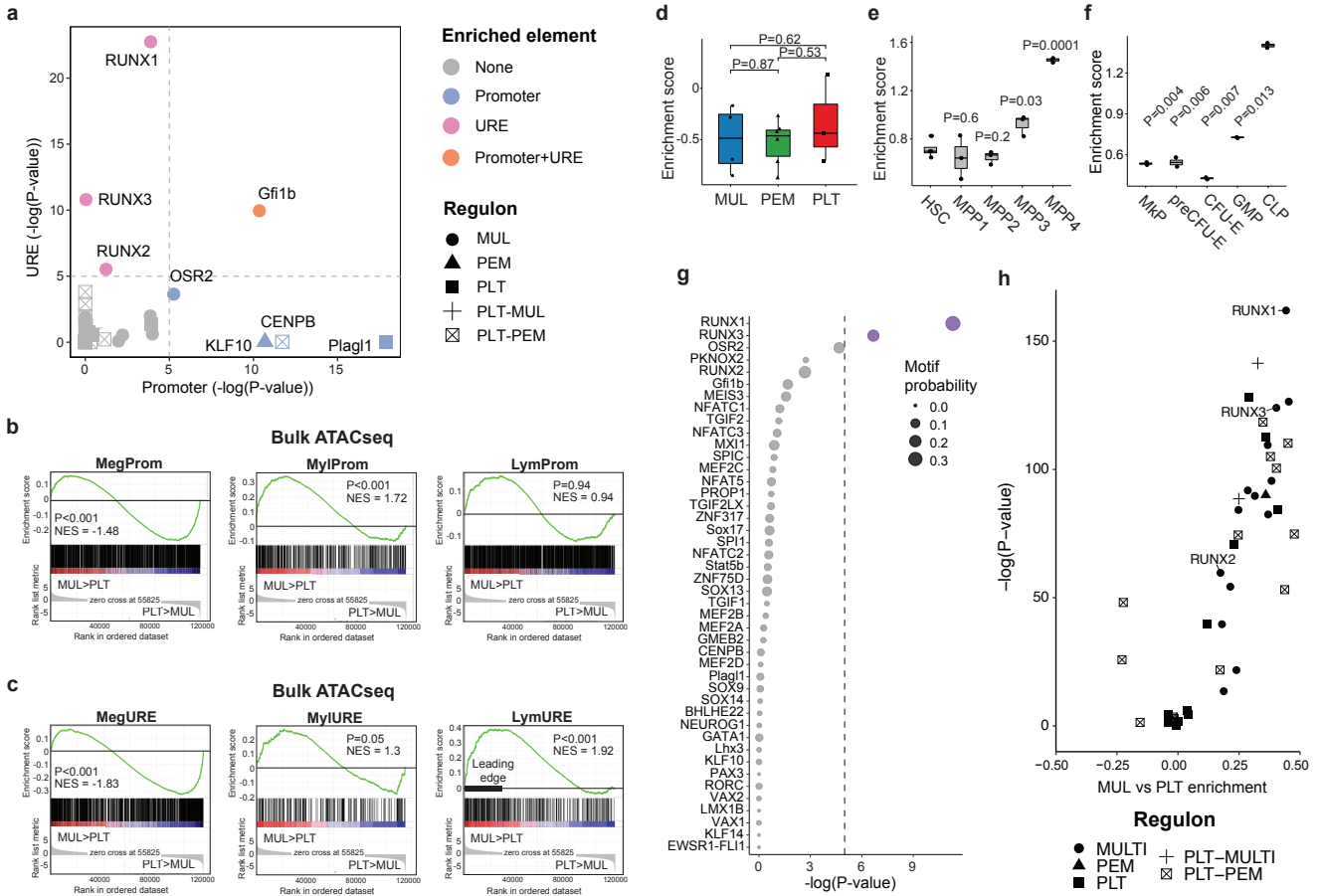
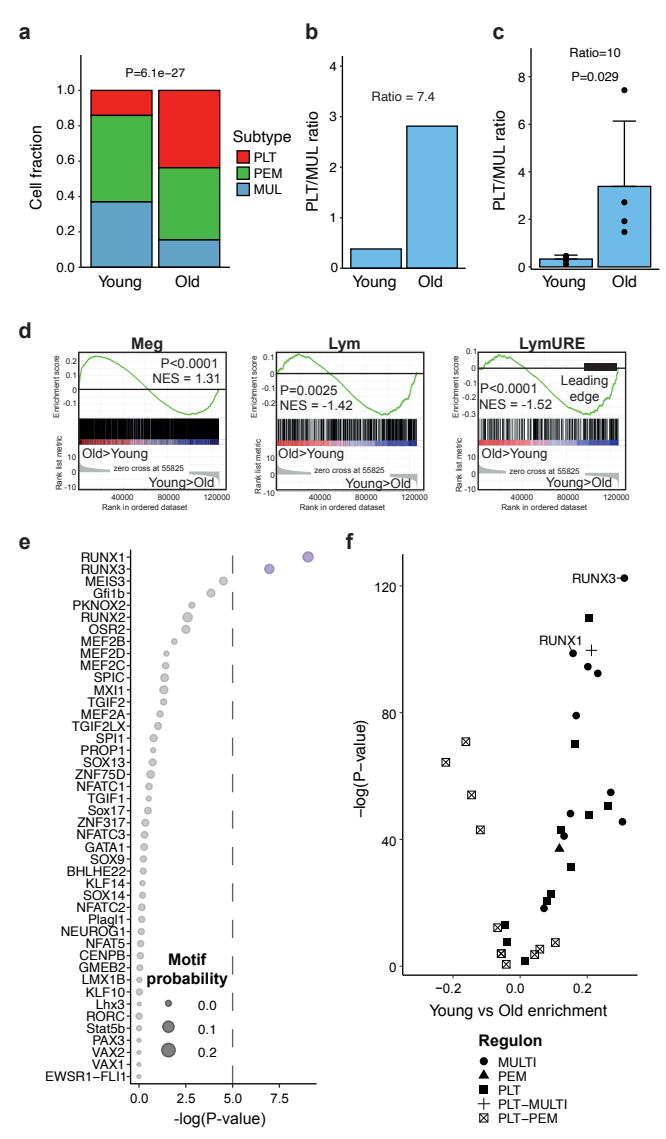

a
Sort single Vwf+ BM HSCs (CD45.2)
200,000 WT BM cells (CD45.1)
Lethally-irradiated recipients (CD45.1)
FACS blood readout: Platelets, erythrocytes, myeloid, B and T cells
Single cell RNA-seq of donor-derived HSCs
AAAAAAA
AAAAAAA
AAAAAAA
b
PLT
PEM
MUL
% PB cells
Week
8-9
12-13
16-17
21-27
8
12
16
17-23
17
Plt
Ery
My
B-cell
CD4 T
CD8 T
c
Importance Score
Gene rank
Erdr1
Cd27
Mt1
Pam
Rp45s
Actb
Tsc22d3
Ctla2a
Ctla2b
Phactr2
d
UMAP-1
UMAP-2
Cluster
Cluster 1
Cluster 2
Cluster 3
no Cluster
Clone
MUL 1
MUL 2
PEM 1
PEM 2
PLT 1
PLT 2

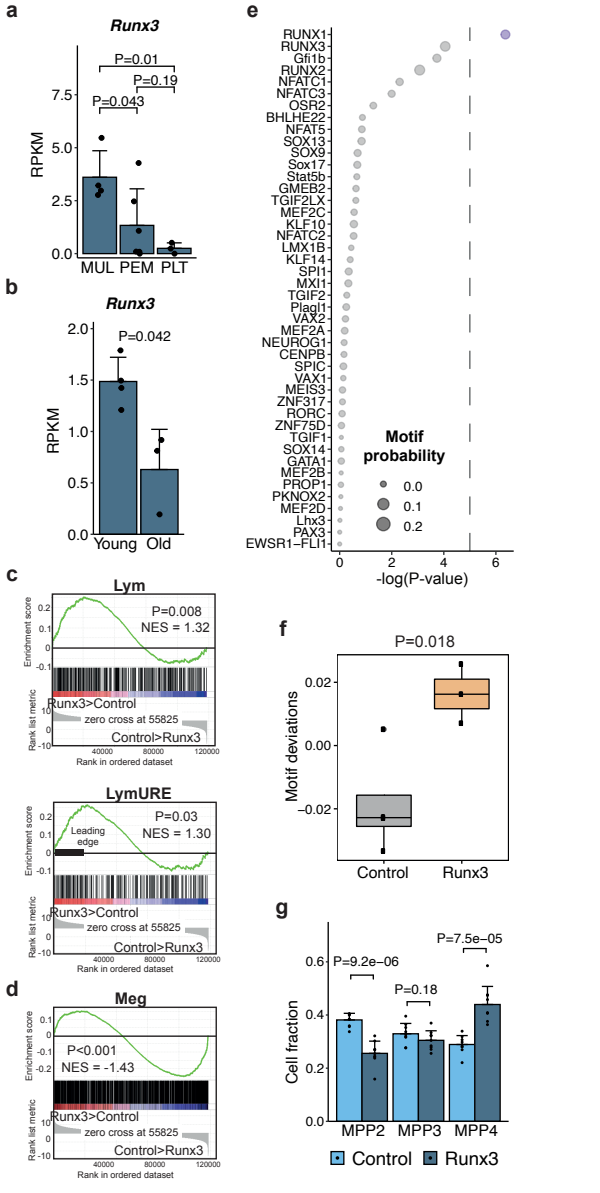
Figure 1

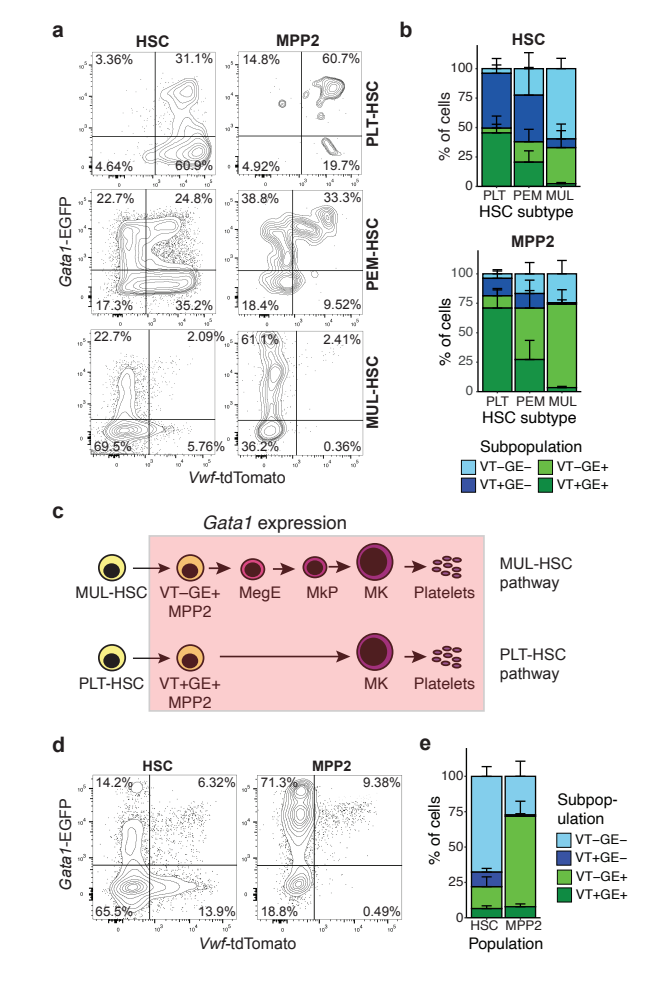

a
Sample
Regulon
RNAseq
b
Sample
Regulon
ATACseq
Sample
PLT
PEM
MUL
Regulon
PLT
PLT+PEM
PEM
MUL+PEM
MUL
MUL+PLT
Scale
3
0
-3
c
MUL3 MUL1 MUL2 MUL4 PEM1 PEM4 PEM5 PEM2 PEM6 PEM3 PLT3 PLT1 PLT2
MUL1
MUL3
MUL4
MUL2
PEM4
PEM1
PEM2
PEM5
PEM6
PEM3
PLT3
PLT1
PLT2
Means
MUL
PEM
PLT
Spearman correlation
1
0
-1
d
MUL3 MUL2 MUL1 MUL4 PLT3 PLT1 PLT2 PEM2 PEM1 PEM5 PEM3 PEM4 PEM6
MUL3
MUL2
MUL1
MUL4
PLT3
PLT1
PLT2
PEM2
PEM1
PEM5
PEM3
PEM4
PEM6
Means
MUL
PEM
PLT
Spearman correlation
1
0
-1
e
HSC/MPP gene sets
HSC3 CLUSTER
MULTILINEAGE_RODRIGUEZ
HIGH-OUTPUT_RODRIGUEZ
HSC4 CLUSTER
HSC2 CLUSTER
STEMSCORE_GILADI
RACFP_DIM_LAURIDSEN
DHSC_VS_AHSC_CABEZAS
HSC_PIETRAS
SERIAL-ENGR_RODRIGUEZ
SUMO_WILSON
MOLO_WILSON
HSC_VS_MPP4_CABEZAS
LOW-OUTPUT_RODRIGUEZ
MK-BIAS_RODRIGUEZ
HSC1 CLUSTER
PLT-HSC up
MUL-HSC up
-log10(P)
2.0
2.5
3.0
3.5
4.0
NES
-2
0
2
f
Hallmark gene sets
E2F_TARGETS
MYC_TARGETS_V1
MYC_TARGETS_V2
OXIDATIVE_PHOSPHORYLATION
G2M_CHECKPOINT
DNA_REPAIR
HYPOXIA
APOPTOSIS
P53_PATHWAY
CHOLESTEROL_HOMEOST
UV_RESPONSE_DN
EPITH_MESENCH_TRANSITION
UV_RESPONSE_UP
TNFA_SIGNALING_VIA_NFKB
NOTCH_SIGNALING
INFLAMMATORY_RESPONSE
INF_ALPHA_RESPONSE
INF_GAMMA_RESPONSE
PLT-HSC up
MUL-HSC up
-log10(P)
1.5
2.0
2.5
3.0
3.5
4.0
NES
-2
0
2

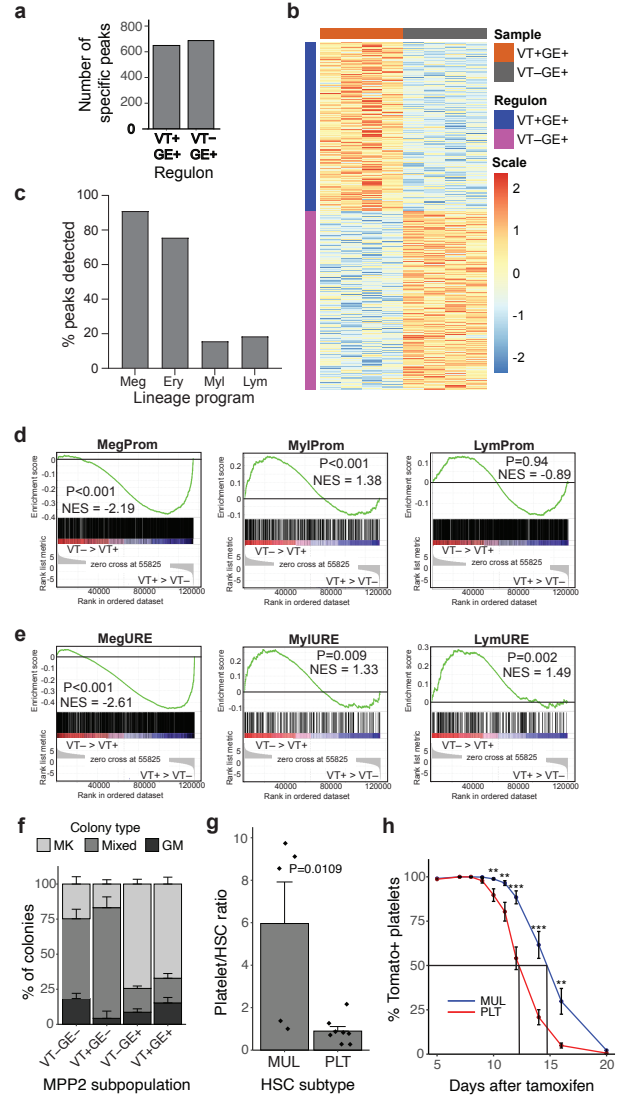

a
Sample
Regulon
TF motif
EWSR1-FLI1
GATA1
MEF2A
MEF2B
MEF2C
MEF2D
Plagl1
RORC
SOX13
Stat5b
ZNF75D
CENPB
GMEB2
KLF14
Lhx3
LMX1B
NFAT5
NFATC1
NFATC2
NFATC3
PAX3
PROP1
SOX14
SOX9
VAX1
VAX2
KLF10
BHLHE22
Gfi1b
MEIS3
MXI1
NEUROG1
OSR2
PKNOX2
RUNX1
RUNX2
RUNX3
Sox17
TGIF1
TGIF2
TGIF2LX
ZNF317
SPI1
SPIC
Sample
PLT
PEM
MUL
Regulon
PLT
PLT+PEM
PEM
MUL+PEM
MUL
MUL+PLT
Scale
2
0
-2
b
MkP
P=0.002
NES = -1.84
MUL>PLT
PLT>MUL
zero cross at 7094
Enrichment score
Rank list metric
Rank in ordered dataset
preGM
P<0.0001
NES = 1.7
CLP
P=0.11
NES = 1.11
c
Cell type
Signature
Cell type
MkP
CFU-E
NMP
proB
DN3
Signature
Meg
Ery
Myl
Lym
Scale
2
0
-2
d
% peaks detected
P<0.0001
P<0.0001
P<0.0001
P<0.0001
Meg
Ery
Myl
Lym
Lineage program
e
Meg
P<0.001
NES = -1.84
Myl
P=0.05
NES = 1.23
Lym
P<0.001
NES = 1.73

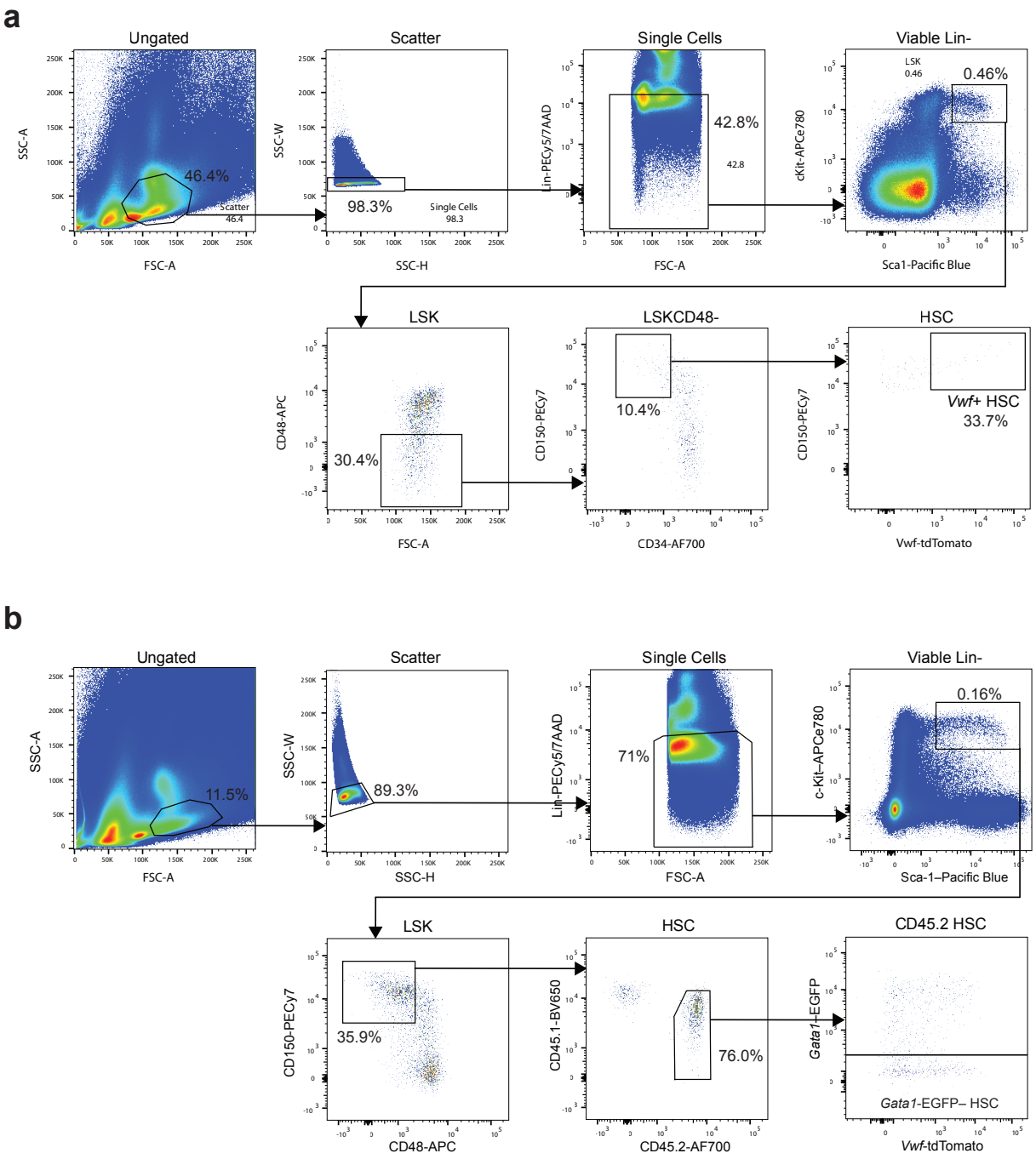

a
RUNX1
RUNX3
RUNX2
Gfi1b
OSR2
CENPB
KLF10
Plagl1
URE (-log(P-value))
Promoter (-log(P-value))
Enriched element
None
Promoter
URE
Promoter+URE
Regulon
MUL
PEM
PLT
PLT-MUL
PLT-PEM
b
Bulk ATACseq
MegProm
P<0.001
NES = -1.48
MylProm
P<0.001
NES = 1.72
LymProm
P=0.94
NES = 0.94
MUL>PLT
PLT>MUL
Rank in ordered dataset
c
Bulk ATACseq
MegURE
P<0.001
NES = -1.83
MylURE
P=0.05
NES = 1.3
LymURE
P<0.001
NES = 1.92
Leading edge
d
P=0.62
P=0.53
P=0.87
Enrichment score
MUL
PEM
PLT
e
P=0.0001
P=0.03
P=0.6
P=0.2
HSC
MPP1
MPP2
MPP3
MPP4
f
P=0.004
P=0.006
P=0.007
P=0.013
MkP
preCFU-E
CFU-E
GMP
CLP
g
RUNX1
RUNX3
OSR2
PKNOX2
RUNX2
Gfi1b
MEIS3
NFATC1
TGIF2
NFATC3
MXI1
SPIC
MEF2C
NFAT5
PROP1
TGIF2LX
ZNF317
Sox17
SPI1
NFATC2
Stat5b
ZNF75D
SOX13
TGIF1
MEF2B
MEF2A
GMEB2
CENPB
MEF2D
Plagl1
SOX9
SOX14
BHLHE22
NEUROG1
GATA1
Lhx3
KLF10
PAX3
RORC
VAX2
LMX1B
VAX1
KLF14
EWSR1-FLI1
Motif probability
0.0
0.1
0.2
0.3
-log(P-value)
h
RUNX1
RUNX3
RUNX2
−log(P−value)
MUL vs PLT enrichment
Regulon
MULTI
PEM
PLT
PLT−MULTI
PLT−PEM


a
P=6.1e−27
Cell fraction
Subtype
PLT
PEM
MUL
Young
Old
b
Ratio = 7.4
PLT/MUL ratio
Young
Old
c
Ratio=10
P=0.029
PLT/MUL ratio
Young
Old
d
Meg
P<0.0001
NES = 1.31
Lym
P=0.0025
NES = -1.42
LymURE
P<0.0001
NES = -1.52
Leading edge
Enrichment score
Rank list metric
Old>Young
zero cross at 55825
Young>Old
Rank in ordered dataset
e
RUNX1
RUNX3
MEIS3
Gfi1b
PKNOX2
RUNX2
OSR2
MEF2B
MEF2D
MEF2C
SPIC
MXI1
TGIF2
MEF2A
TGIF2LX
SPI1
PROP1
SOX13
ZNF75D
NFATC1
TGIF1
Sox17
ZNF317
NFATC3
GATA1
SOX9
BHLHE22
KLF14
SOX14
NFATC2
Plagl1
NEUROG1
NFAT5
CENPB
GMEB2
LMX1B
KLF10
Lhx3
RORC
Stat5b
PAX3
VAX2
VAX1
EWSR1−FLI1
Motif probability
0.0
0.1
0.2
-log(P-value)
f
RUNX3
RUNX1
−log(P−value)
Young vs Old enrichment
Regulon
MULTI
PEM
PLT
PLT−MULTI
PLT−PEM

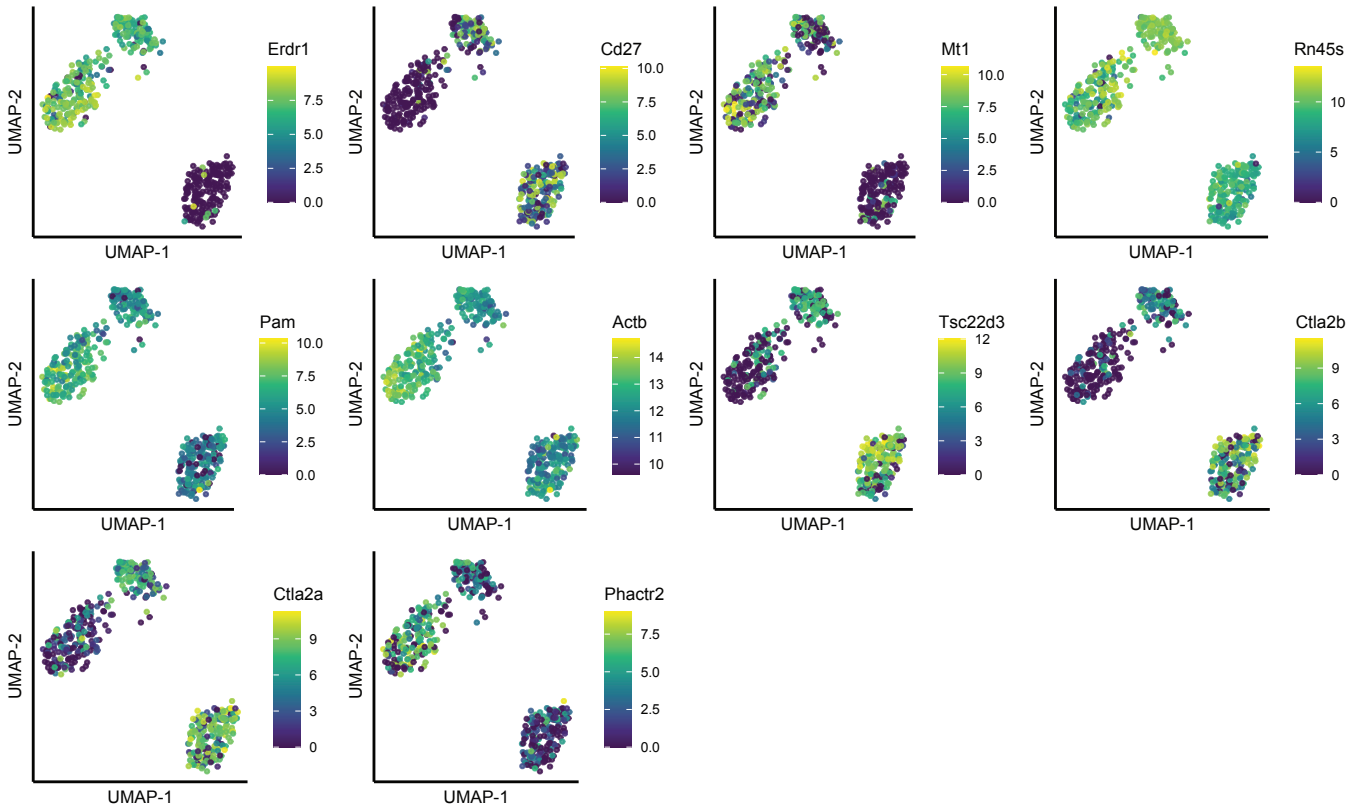

a
Runx3
P=0.01
P=0.19
P=0.043
RPKM
7.5
5.0
2.5
0.0
MUL PEM PLT
b
Runx3
P=0.042
RPKM
2.0
1.5
1.0
0.5
0.0
Young Old
e
RUNX1
RUNX3
Gfi1b
RUNX2
NFATC1
NFATC3
OSR2
BHLHE22
NFAT5
SOX13
SOX9
Sox17
Stat5b
GMEB2
TGIF2LX
MEF2C
KLF10
NFATC2
LMX1B
KLF14
SPI1
MXI1
TGIF2
Plagl1
VAX2
MEF2A
NEUROG1
CENPB
SPIC
VAX1
MEIS3
ZNF317
RORC
ZNF75D
TGIF1
SOX14
GATA1
MEF2B
PROP1
PKNOX2
MEF2D
Lhx3
PAX3
EWSR1−FLI1
Motif probability
0.0
0.1
0.2
0
2
4
6
-log(P-value)
c
Lym
Enrichment score
P=0.008
NES = 1.32
Runx3>Control
zero cross at 55825
Control>Runx3
Rank list metric
Rank in ordered dataset
LymURE
Leading edge
P=0.03
NES = 1.30
d
Meg
P<0.001
NES = -1.43
f
P=0.018
Motif deviations
0.02
0.00
−0.02
Control
Runx3
g
P=9.2e−06
P=0.18
P=7.5e−05
Cell fraction
0.6
0.4
0.2
0.0
MPP2
MPP3
MPP4
Control
Runx3


a
HSC
MPP2
3.36%
31.1%
14.8%
60.7%
PLT-HSC
4.64%
60.9%
4.92%
19.7%
22.7%
24.8%
38.8%
33.3%
PEM-HSC
17.3%
35.2%
18.4%
9.52%
22.7%
2.09%
61.1%
2.41%
MUL-HSC
69.5%
5.76%
36.2%
0.36%
Gata1-EGFP
Vwf-tdTomato
b
HSC
% of cells
PLT PEM MUL
HSC subtype
MPP2
% of cells
PLT PEM MUL
HSC subtype
Subpopulation
VT−GE−
VT+GE−
VT−GE+
VT+GE+
c
Gata1 expression
MUL-HSC
VT−GE+ MPP2
MegE
MkP
MK
Platelets
MUL-HSC pathway
PLT-HSC
VT+GE+ MPP2
MK
Platelets
PLT-HSC pathway
d
HSC
MPP2
14.2%
6.32%
71.3%
9.38%
65.5%
13.9%
18.8%
0.49%
Gata1-EGFP
Vwf-tdTomato
e
% of cells
HSC MPP2
Population
Subpop-ulation
VT−GE−
VT+GE−
VT−GE+
VT+GE+

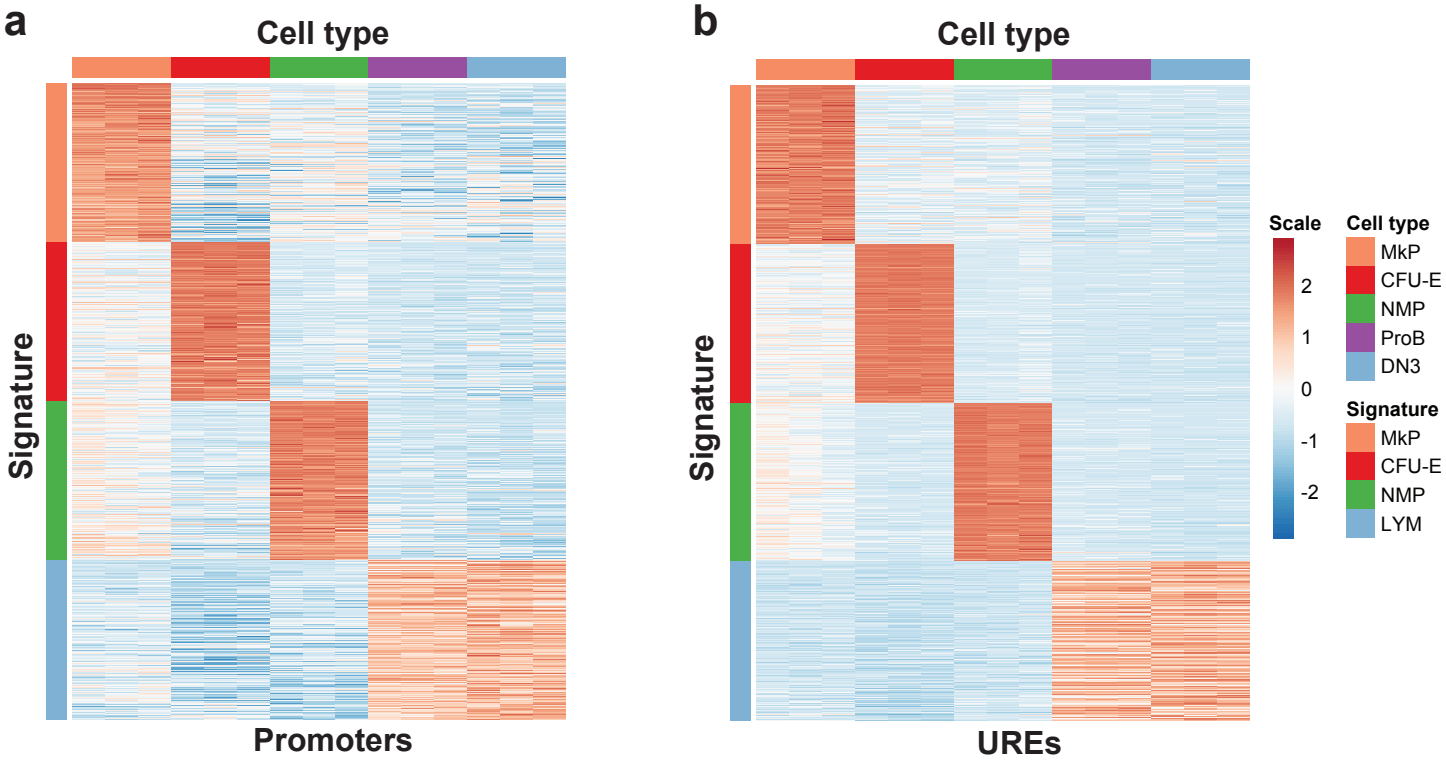

a
Number of specific peaks
800
600
400
200
0
VT+ GE+
VT– GE+
Regulon
b
Sample
VT+GE+
VT–GE+
Regulon
VT+GE+
VT–GE+
Scale
2
1
0
-1
-2
c
% peaks detected
100
80
60
40
20
0
Meg
Ery
Myl
Lym
Lineage program
d
MegProm
P<0.001
NES = -2.19
MylProm
P<0.001
NES = 1.38
LymProm
P=0.94
NES = -0.89
Enrichment score
Rank list metric
VT– > VT+
zero cross at 55825
VT+ > VT–
Rank in ordered dataset
e
MegURE
P<0.001
NES = -2.61
MylURE
P=0.009
NES = 1.33
LymURE
P=0.002
NES = 1.49
f
Colony type
MK
Mixed
GM
% of colonies
VT–GE–
VT+GE–
VT–GE+
VT+GE+
MPP2 subpopulation
g
Platelet/HSC ratio
P=0.0109
MUL
PLT
HSC subtype
h
% Tomato+ platelets
MUL
PLT
Days after tamoxifen



Extended Data Figure 1



Extended Data Figure 2



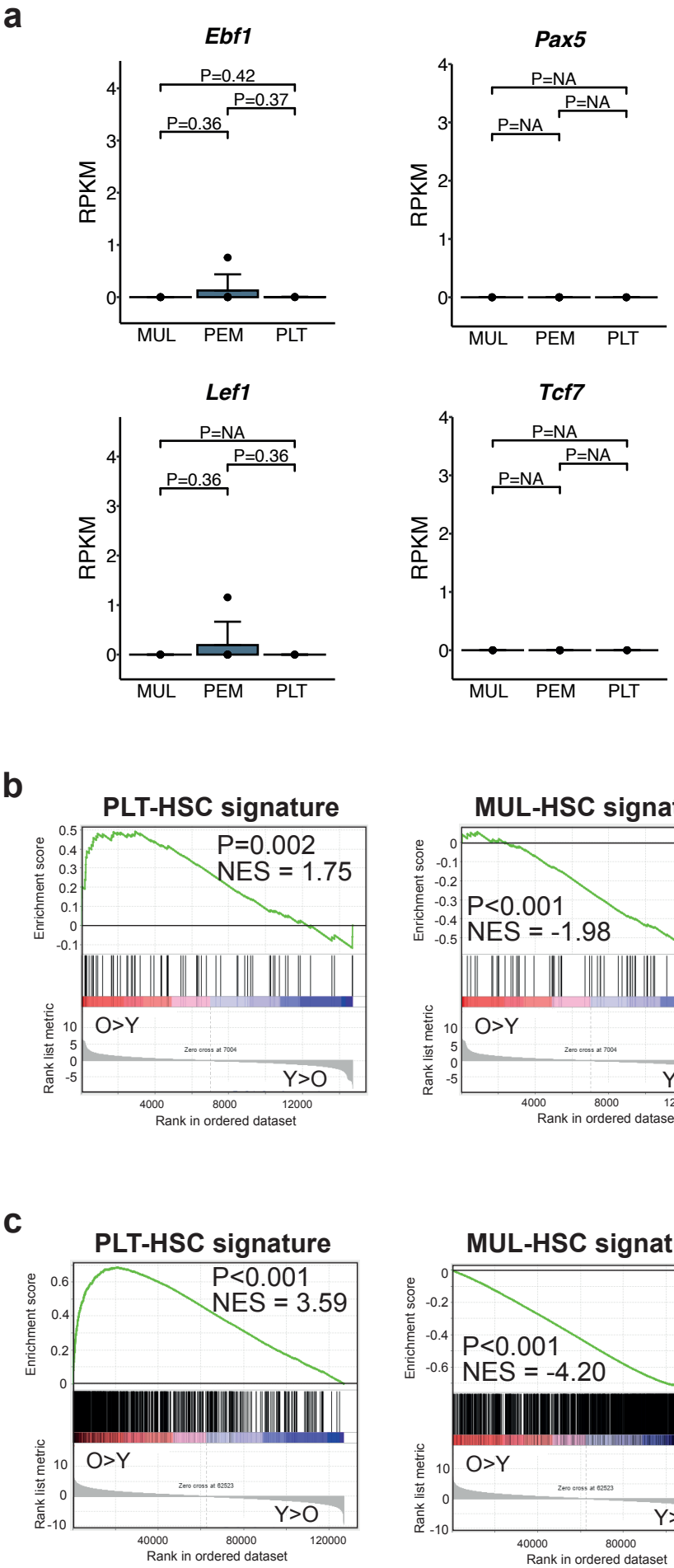
c

TF	PWM	Motif ID	Raw score	P–value	Hit %
Plagl1		MA1615.1	5.22	1.4e–18	11 %
CENPB		MA0637.1	0.346	1.94e–12	11 %
KLF10		MA1511.1	10.4	2.07e–11	10 %
Gfi1b		MA0483.1	2.12	4.5e–11	8 %
OSR2		MA1646.1	2.43	5.5e–06	8 %
TGIF2LX		MA1571.1	2	0.000102	6 %
MEF2B		MA0660.1	2.03	0.000126	1 %
RUNX1		MA0002.2	2.39	0.00013	9 %
PKNOX2		MA0783.1	0.609	0.000137	2 %
TGIF2		MA0797.1	1.23	0.00607	4 %
TGIF1		MA0796.1	0.163	0.0107	2 %
RUNX2		MA0511.2	0.898	0.0588	8 %
GMEB2		MA0862.1	0.239	0.0737	5 %
ZNF75D		MA1601.1	2.5	0.349	7 %
SPIC		MA0687.1	2.48	0.37	6 %
KLF14		MA0740.1	15.4	0.499	4 %
MEF2D		MA0773.1	1.87	0.607	1 %
SPI1		MA0080.5	9.75	0.665	5 %
RUNX3		MA0684.2	1.14	0.912	8 %
Stat5b		MA1625.1	2.33	0.999	4 %
SOX14		MA1562.1	0.391	1	2 %
EWSR1–FLI1		MA0149.1	154000	1	0 %
MEF2C		MA0497.1	1.14	1	2 %
ZNF317		MA1593.1	1.75	1	4 %
MEIS3		MA0775.1	0.973	1	4 %
Lhx3		MA0135.1	0.952	1	0 %
PROP1		MA0715.1	0.545	1	1 %
MEF2A		MA0052.4	1.02	1	2 %
PAX3		MA0780.1	0.235	1	0 %
BHLHE22		MA0818.1	0.207	1	1 %
GATA1		MA0035.4	0.435	1	4 %
LMX1B		MA0703.2	0.399	1	1 %
MXI1		MA1108.2	0.858	1	5 %
NEUROG1		MA0623.2	0.189	1	1 %
NFAT5		MA0606.1	0.995	1	2 %
NFATC1		MA0624.1	0.942	1	4 %
NFATC2		MA0152.1	1.3	1	4 %
NFATC3		MA0625.1	0.912	1	3 %
RORC		MA1151.1	0.574	1	3 %
SOX13		MA1120.1	0.595	1	5 %
Sox17		MA0078.1	0.553	1	4 %
SOX9		MA0077.1	0.506	1	3 %
VAX1		MA0722.1	0.371	1	0 %
VAX2		MA0723.1	0.374	1	0 %

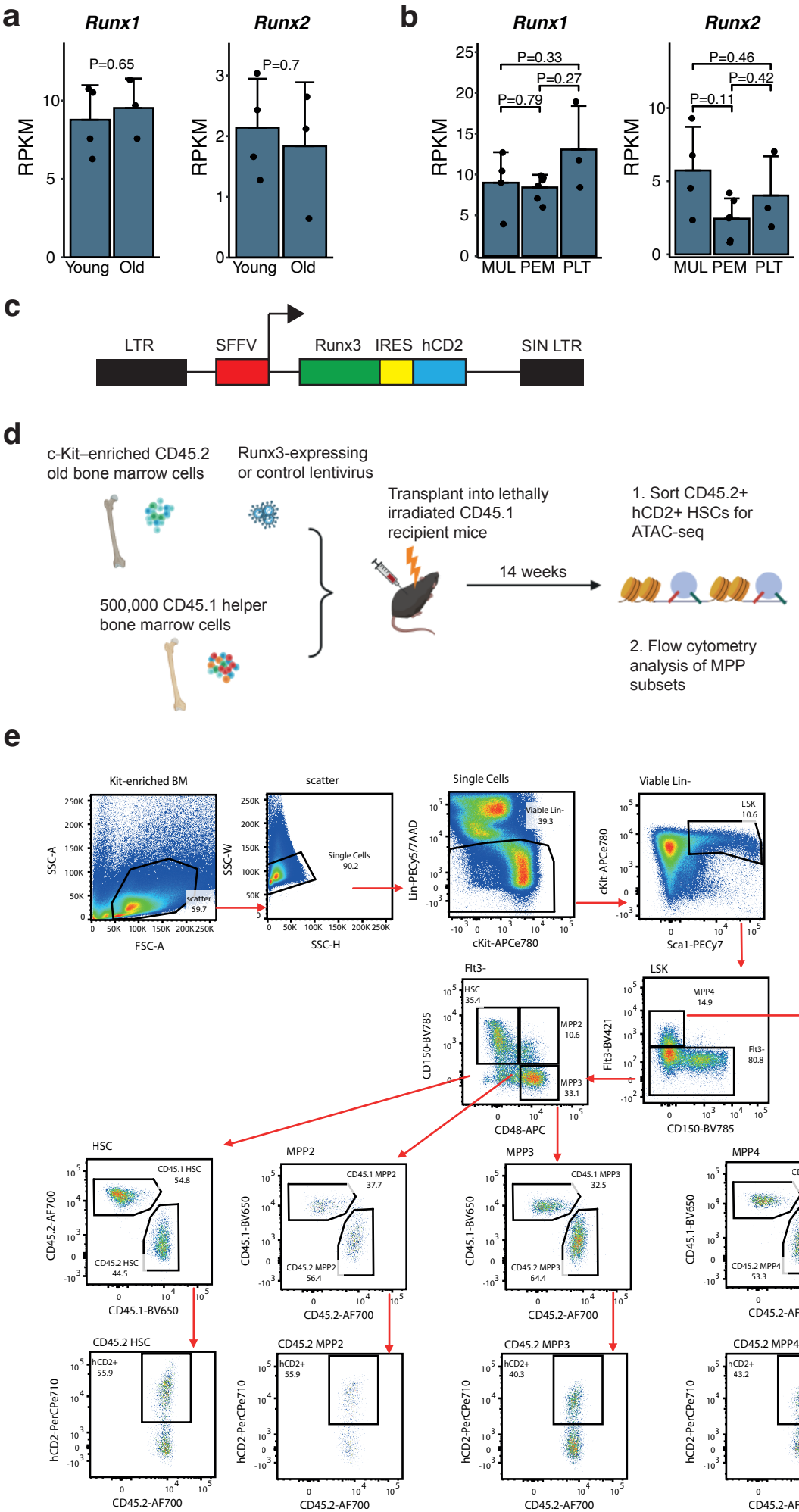
d

TF	PWM	Motif ID	Raw score	P–value	Hit %
RUNX1		MA0002.2	6.38	1.8e–23	26 %
RUNX3		MA0684.2	2.52	1.58e–11	15 %
Gfi1b		MA0483.1	3.97	1.1e–10	13 %
RUNX2		MA0511.2	1.93	3e–06	19 %
NFATC1		MA0624.1	1.37	0.000157	10 %
OSR2		MA1646.1	3.03	0.000229	9 %
NFATC3		MA0625.1	1.5	0.00127	10 %
PKNOX2		MA0783.1	0.905	0.0102	2 %
NFATC2		MA0152.1	1.92	0.0107	10 %
MXI1		MA1108.2	1.84	0.0123	14 %
MEIS3		MA0775.1	1.33	0.0391	8 %
NFAT5		MA0606.1	2	0.0449	8 %
MEF2B		MA0660.1	1.99	0.0458	2 %
SPIC		MA0687.1	3.77	0.0858	7 %
Sox17		MA0078.1	1.01	0.136	8 %
SOX13		MA1120.1	1.03	0.136	9 %
MEF2D		MA0773.1	3.41	0.16	1 %
MEF2C		MA0497.1	4.73	0.186	4 %
TGIF2LX		MA1571.1	1.27	0.238	5 %
SPI1		MA0080.5	14.5	0.242	7 %
BHLHE22		MA0818.1	0.891	0.247	2 %
TGIF2		MA0797.1	0.825	0.286	3 %
ZNF75D		MA1601.1	2.63	0.315	9 %
MEF2A		MA0052.4	3.59	0.393	3 %
GMEB2		MA0862.1	0.183	0.611	4 %
ZNF317		MA1593.1	2.47	0.611	6 %
NEUROG1		MA0623.2	0.758	0.613	2 %
Stat5b		MA1625.1	2.77	0.621	5 %
SOX9		MA0077.1	0.796	0.7	3 %
PROP1		MA0715.1	1.03	0.793	2 %
TGIF1		MA0796.1	0.0248	0.817	0 %
SOX14		MA1562.1	0.386	0.882	2 %
CENPB		MA0637.1	0.0724	0.933	4 %
Plagl1		MA1615.1	1.97	0.935	4 %
KLF10		MA1511.1	3.77	0.941	4 %
LMX1B		MA0703.2	0.734	0.96	1 %
RORC		MA1151.1	0.634	0.989	4 %
GATA1		MA0035.4	0.675	0.99	3 %
KLF14		MA0740.1	2.85	0.998	0 %
Lhx3		MA0135.1	0.628	0.999	1 %
VAX2		MA0723.1	0.598	1	1 %
PAX3		MA0780.1	0.128	1	1 %
VAX1		MA0722.1	0.613	1	0 %
EWSR1–FLI1		MA0149.1	0.926	1	0 %

Extended Data Figure 3

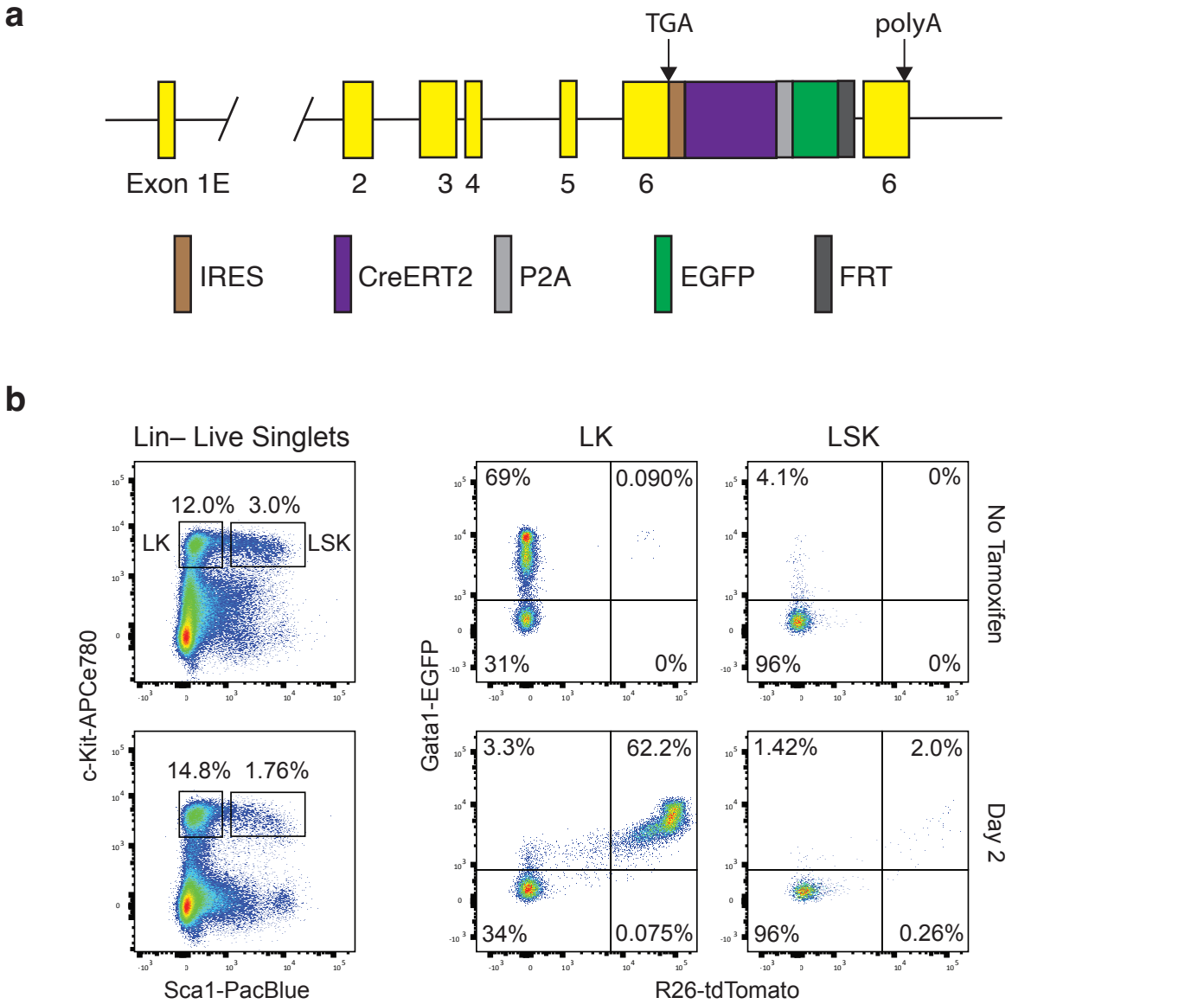


Extended Data Figure 4



Lorem

Extended Data Figure 5



Extended Data Figure 6