

Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios

Philipp S. Hoppe^{1,2}, Michael Schwarzfischer³, Dirk Loeffler^{1,2}, Konstantinos D. Kokkaliaris^{1,2}, Oliver Hilsenbeck^{1,2,3}, Nadine Moritz², Max Endele^{1,2}, Adam Filipczyk², Adriana Gambardella⁴, Nouraiz Ahmed¹, Martin Etzrodt¹, Daniel L. Couturier¹, Michael A. Rieger², Carsten Marr³, Michael K. Strasser³, Bernhard Schaubberger², Ingo Burtscher⁵, Olga Ermakova⁶, Antje Bürger⁷, Heiko Lickert^{5,8}, Claus Nerlov^{4,9}, Fabian J. Theis^{3,10} & Timm Schroeder^{1,2}

The mechanisms underlying haematopoietic lineage decisions remain disputed. Lineage-affiliated transcription factors^{1,2} with the capacity for lineage reprogramming³, positive auto-regulation^{4,5} and mutual inhibition^{6,7} have been described as being expressed in uncommitted cell populations⁸. This led to the assumption that lineage choice is cell-intrinsically initiated and determined by stochastic switches of randomly fluctuating cross-antagonistic transcription factors³. However, this hypothesis was developed on the basis of RNA expression data from snapshot and/or population-averaged analyses^{9–12}. Alternative models of lineage choice therefore cannot be excluded. Here we use novel reporter mouse lines and live imaging for continuous single-cell long-term quantification of the transcription factors GATA1 and PU.1 (also known as SPI1). We analyse individual haematopoietic stem cells throughout differentiation into megakaryocytic–erythroid and granulocytic–monocytic lineages. The observed expression dynamics are incompatible with the assumption that stochastic switching between PU.1 and GATA1 precedes and initiates megakaryocytic–erythroid versus granulocytic–monocytic lineage decision-making. Rather, our findings suggest that these transcription factors are only executing and reinforcing lineage choice once made. These results challenge the current prevailing model of early myeloid lineage choice.

Multipotent haematopoietic stem and progenitor cells (HSPCs) are thought to differentiate into all blood cell types through a series of progenitor cell types with increasingly restricted lineage potential—for example, the common myeloid progenitor (CMP), which then further differentiates into megakaryocytic–erythroid (MegE) and granulocytic–monocytic (GM) progenitors (MEPs and GMPs)¹³. The molecular mechanisms controlling lineage choice remain controversial. The prevailing model assumes that lineage choice is initiated and determined by stochastic fluctuations of cross-antagonistic transcription factor (TF) pairs^{3,14}, developed around the haematopoietic TFs PU.1 and GATA1. PU.1 and GATA1 are expressed in GM and MegE cells¹³, respectively; they are required for the production of mature cells of these lineages^{1,2} and can reprogram cells towards their lineages upon overexpression³. PU.1 and GATA1 can cross-inhibit each other's activity^{6,7} and activate their own transcription^{4,5}. This wiring can generate bi-stable switches, with random higher expression of one TF, leading to increased expression of this TF and repression of competing TFs. PU.1 and *Gata1* mRNAs were described as being co-expressed before HSPC lineage choice⁸. Therefore, lineage decisions may be initiated by random fluctuations of TF levels, breaking the TF equilibrium of a cell. The specific wiring of the TF network would lead to specific probabilities

for individual TFs to ‘win’ and thus to stable frequencies of lineage choices. In this model, TFs would not only execute and reinforce, but also initiate and make lineage decisions¹⁴.

However, this model was based on low-resolution expression data that cannot be used to exclude alternative models. Most studies only analysed HSPC population averages^{9–11}, thus masking cellular heterogeneity^{12,15} and dead cells to non-quantitatively measure mRNA expression, ignoring relevant protein expression dynamics and future lineage choice^{12,16}. We therefore developed approaches for continuous long-term single-cell quantification of PU.1 and GATA1 protein expression in individual HSPCs from before until after their lineage choice. We created knock-in mouse lines with reading frames for yellow (enhanced yellow fluorescent protein; eYFP) and red (mCherry) fluorescent proteins knocked into the gene loci for PU.1 and *Gata1*, respectively (Fig. 1a, Extended Data Fig. 1a). The fluorescent proteins are fused to the C terminus of each TF, allowing quantification by fluorescence intensity. We mated the resulting PU.1^{eYFP} (ref. 17) and GATA1^{mCherry} mice to create PU.1^{eYFP}GATA1^{mCherry} mice. These are homozygous for both PU.1^{eYFP} and *Gata1*^{mCherry} alleles, and males are hemizygous for X-chromosomal *Gata1*^{mCherry}.

As described previously, no² or reduced¹⁸ PU.1 expression or altered PU.1 function¹⁹, and no¹ or altered GATA1 expression²⁰ have drastic phenotypes, causing leukaemia or absence of mature GM or MegE cells, for example. In contrast, PU.1^{eYFP}GATA1^{mCherry} mice show no aberrant phenotypes, are born at normal Mendelian ratios (Extended Data Table 1), and did not show increased mortality throughout adulthood (data not shown). The cellular composition of GATA1-mCherry-expressing fetal livers (Extended Data Fig. 1b), peripheral blood (Extended Data Fig. 1c) and bone marrow (Fig. 1b–f) from adult PU.1^{eYFP}GATA1^{mCherry} mice was unchanged. Colony formation *in vitro* was unaltered for PU.1^{eYFP}GATA1^{mCherry} cells (Fig. 1g, h). We could also not observe a difference in GM to MegE lineage output of PU.1^{eYFP}GATA1^{mCherry} versus wild-type HSPCs in competitive repopulation assays (Fig. 1i, j, Extended Data Fig. 2). Finally, reprogramming of cells to MegE or GM lineages by GATA1-mCherry or PU.1-eYFP, respectively, was as efficient as with wild-type TFs, both in wild-type and in PU.1^{eYFP}GATA1^{mCherry} cells (Extended Data Fig. 3). In conclusion, PU.1-eYFP and GATA1-mCherry fusion proteins function normally.

We showed normal expression and stability of the fusion proteins by quantitative immunofluorescence staining for PU.1 and GATA1. Simultaneous staining for the TFs and the fluorescent proteins (Fig. 1k–n) showed high expression correlation and co-localization in HSPC nuclei. Distributions of PU.1 or GATA1 protein expression were

¹Department of Biosystems Science and Engineering, ETH Zurich, 4058 Basel, Switzerland. ²Research Unit Stem Cell Dynamics, Helmholtz Zentrum München, 85764 Neuherberg, Germany.

³Institute of Computational Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany. ⁴MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK. ⁵Institute of Diabetes and Regeneration Research, Institute of Stem Cell Research, Helmholtz Zentrum München, Business Campus Garching, 85748 Garching, Germany. ⁶Istituto di Biologia Cellulare e Neurobiologia, CNR, 00015 Monterotondo, Italy. ⁷Institute of Developmental Genetics, Helmholtz Zentrum München, 85764 Neuherberg, Germany.

⁸Beta Cell Biology, Medical Faculty, Technical University Munich (TUM), 80333 Munich, Germany. ⁹EMBL Mouse Biology Unit, 00015 Monterotondo, Italy. ¹⁰Department of Mathematics, Technical University Munich, 85748 Garching, Germany.

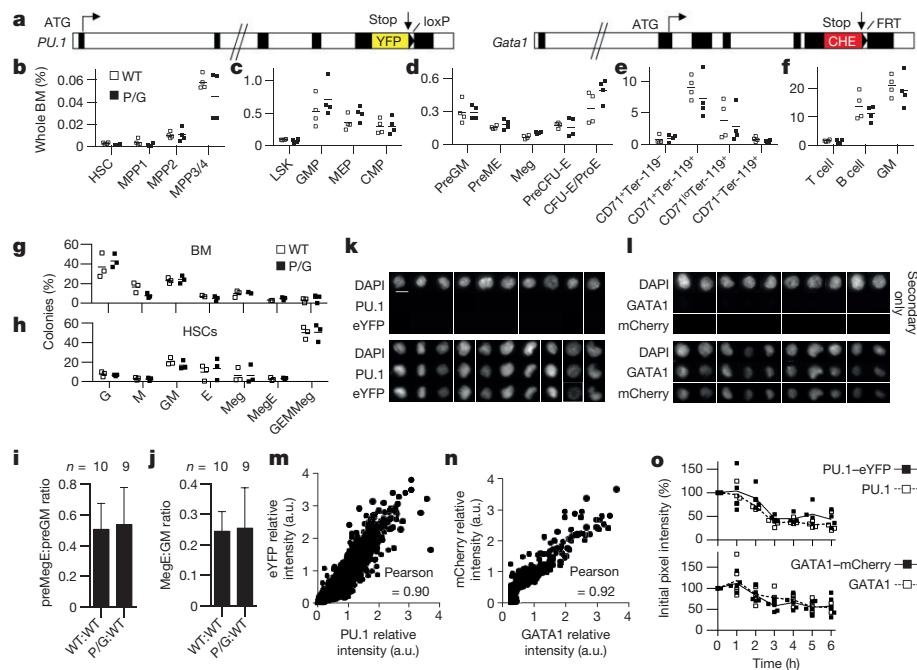


Figure 1 | Normal function and stability of PU.1-eYFP and GATA1-mCherry. **a**, Endogenous gene loci after knock-in. Black boxes, exons. **b–f**, Normal haematopoiesis in PU.1^{eYFP}GATA1^{mCherry} (P/G) double homozygous mice. Composition of adult bone marrow (4 independent experiments). One-way MANOVA for **b–f**, $P = 0.35, 0.35, <0.03, 0.31, 0.16$, respectively. E, erythroid. **g–j**, Normal lineage choice of PU.1^{eYFP}GATA1^{mCherry} HSPCs. **g, h**, Colony-forming assay from whole bone marrow (BM) (**g**) and HSCs (**h**) (mean, 3 independent experiments each). No significant difference for each population (Wilcoxon rank sum test, P values >0.2 and >0.4 , respectively).

not changed in PU.1^{eYFP}GATA1^{mCherry} HSPC populations (Extended Data Figs 4, 5a–e), and resembled those read out by fluorescence of the TF fusion proteins (see below, Fig. 2a). Fusion of fluorescent proteins to PU.1 and GATA1 therefore did not alter their expression. In addition, we could not detect changes in the stability of PU.1-eYFP and GATA1-mCherry by quantitative immunostaining (Fig. 1o). Thus, although only surrogate reporters, fluorescent protein fusions can be used²¹ to quantify expression of PU.1 and GATA1 proteins in living HSPCs.

As expected, MEPs express high levels of GATA1-mCherry, but only low levels of PU.1-eYFP. GMPs express high levels of PU.1-eYFP, whereas most are negative for GATA1-mCherry (Fig. 2a). These expression patterns are identical to those previously described for endogenous TFs^{13,22}, and are similar, but not the same, as for other reporters²³. GATA1 is known to have a role downstream of GMPs²⁴, and a small GMP subpopulation co-expressed PU.1 and GATA1 (Fig. 2a).

i, j, MegE or GM lineage choice is not changed in competitive transplantation assays (Wilcoxon rank sum test, $P = 0.97$ and 0.84 , respectively. Compare Extended Data Fig. 2). **k–n**, Immunostaining for PU.1 and eYFP (**k, m**) and GATA1 and mCherry (**l, n**), day 7 of HSC differentiation. Representative examples from three independent experiments. DAPI, nuclei stain; a.u., arbitrary units. Scale bar, 10 μ m. **o**, Normal stability of TF fusion proteins. PU.1-eYFP (4 independent experiments) or GATA1-mCherry (5 independent experiments) protein decay after 50 μ M Cycloheximide treatment of GMPs or preMegE cells, respectively. Data from quantitative immunostaining against PU.1 or GATA1.

These cells do not have CMP (GEMMeg) lineage potential, and their strong GM bias suggests that GATA1 does not have a role in the GM versus MegE lineage decision of these cells (data not shown). PU.1^{mid}GATA1^{mid} progenitors mostly had only MegE potential (Fig. 2b). PU.1⁻GATA1^{hi} cells are more mature and no longer have colony potential (Fig. 2b).

HSPCs express both PU.1 and Gata1 mRNA before MegE versus GM lineage choice^{8,10,11}. We therefore expected CD34⁺CD16/32⁻c-Kit^{+/-}Sca-1⁻lineage⁻ CMPs¹³ to co-express both TFs. However, the vast majority express only high PU.1 levels, or GATA1 with low or no PU.1 expression. These are already committed to the GM and MegE lineage, respectively (Fig. 2c). Thus, this ‘CMP’ population is in fact a mixture of already committed GMPs and MEPs²². Haematopoietic stem cells (HSCs) already express intermediate levels of PU.1-eYFP, but no GATA1-mCherry (Fig. 2a). To identify the expected HSPC population

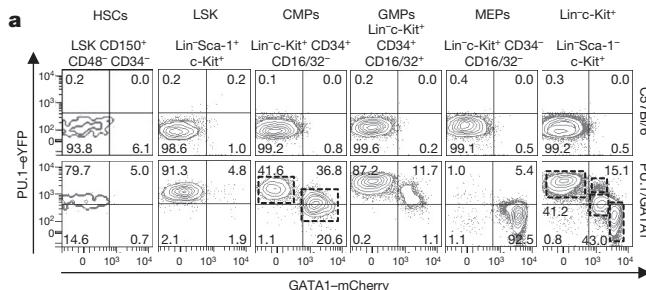


Figure 2 | PU.1-eYFP and GATA1-mCherry expression in different HSPCs. **a**, Flow cytometry analysis of adult bone marrow HSPCs (representative example of 4 independent experiments). **b, c**, Colony formation (mean \pm s.d.) of PU.1-eYFP^{hi}GATA1-mCherry⁻ (clonogenicity 39.6% \pm 7.1%; 4 independent experiments), PU.1-eYFP^{hi}GATA1-

mCherry^{mid} (32.2% \pm 11.4%; 4), PU.1-eYFP^{mid}GATA1-mCherry^{mid} (37.1% \pm 11.9%; 4) and PU.1-eYFP⁻GATA1-mCherry^{hi} (0%; 3) Lin⁻c-Kit⁺ cells (**b**), and PU.1-eYFP⁺GATA1-mCherry⁻ (58.4% \pm 14.4%; 3) and GATA1-mCherry⁺ CMPs (46.3% \pm 10.4%; 3) (**c**). Populations from dashed boxes in **a**.

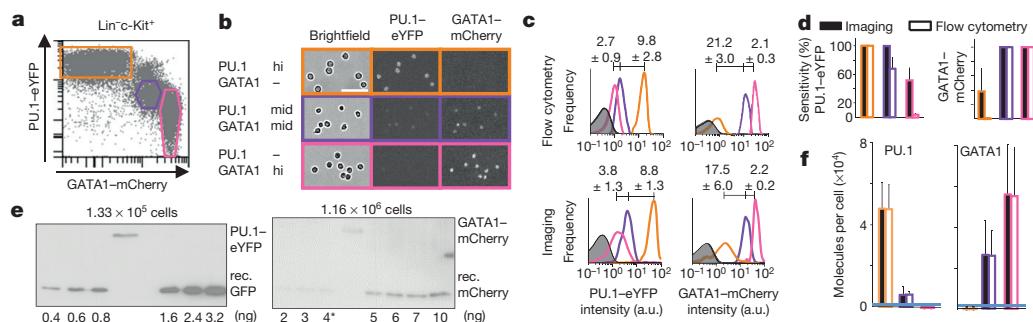


Figure 3 | Sensitive quantification of PU.1-eYFP and GATA1-mCherry protein numbers by live-cell imaging. **a, b,** HSPCs were sorted and imaged. Scale bar, 50 μ m. **c,** Relative fluorescence quantification by imaging or flow cytometry yields comparable results (mean \pm s.d., representative examples from 3 independent experiments). Filled histograms: non-fluorescent control cells. **d,** Better sensitivity of imaging over flow cytometry (mean percentages \pm s.d. of cells gated in **a** above the detection threshold). **e,** Quantification of molecule numbers in sorted PU.1-eYFP⁺

with GEMMeg potential downstream of PU.1^{mid}GATA1⁻ HSCs and upstream of already committed CMPs, GMPs or MEPs, we analysed the whole lineage⁻Sca-1⁺c-Kit⁺ (LSK) progenitor population (Fig. 2a). As in a previously described GATA1 reporter mouse²³, only a small number of LSK cells expressed GATA1-mCherry. The earliest HSPC population in which a small but robust GATA1-mCherry⁺ subpopulation could be detected was the MPP2 (LSK CD34⁺CD48⁺CD150⁺CD135⁻) population^{25,26} (Extended Data Fig. 5f). However, this GATA1⁺ MPP2 subpopulation is already MegE committed (see below).

The PU.1-GATA1 switch model is based on near-stoichiometric PU.1 and GATA1 co-expression, and thus mutual functional inhibition, before GM versus MegE lineage decision^{6,7}. In contrast, we could not identify a PU.1-eYFP⁺GATA1-mCherry⁺ HSPC population with robust GEMMeg potential. However, this data from snapshot cytometry analysis cannot exclude that differentiating PU.1^{mid}GATA1⁻ HSCs may quickly pass through a state with similar PU.1 and GATA1 expression. We therefore extended approaches for long-term imaging and single-cell tracking^{21,27,28} to allow continuous live quantification of PU.1-eYFP and GATA1-mCherry in differentiating HSPCs²⁹. PU.1-eYFP and GATA1-mCherry fluorescence was better detected by imaging than by flow cytometry, with a greater dynamic range and higher sensitivity (Fig. 3a-d). TF protein numbers in individual cells were estimated by comparison to defined amounts of recombinant fluorescent proteins in western blot analyses (Fig. 3e, f and Extended Data Table 2; for gel source data, see Supplementary Fig. 1).

We cultured HSCs under conditions allowing both MegE and GM differentiation, which were detected by expression of GATA1-mCherry or CD16/32 (ref. 13), respectively (Extended Data Fig. 6). Due to its overlapping expression, PU.1-eYFP expression alone does not allow early lineage detection (Extended Data Fig. 6a). Although these culture conditions do not resemble all possible *in vivo* conditions,

bone marrow Lin⁻c-Kit⁺ cells and E14.5 GATA1-mCherry⁺ fetal liver cells by comparison to defined amounts of recombinant (rec.) eGFP and mCherry, respectively. Representative examples from 3 independent experiments are shown. Asterisk (*) indicates incomplete loading of sample. **f,** Estimation of molecule numbers in the populations from **a**. Mean protein abundance per cell is shown. Error bars include uncertainty from western blot quantification and fold changes from flow cytometry or imaging (mean \pm s.d.). Blue lines, imaging detection thresholds.

they do allow differentiation into all relevant lineages, thus enabling analysis of the core mechanisms expected to underlie MegE versus GM lineage choice.

We quantified absolute PU.1 and GATA1 protein levels in single differentiating HSCs and their progeny, throughout up to 11 generations (Fig. 4a, Extended Data Fig. 7). About 6.5×10^6 total measurements, including 3.7×10^5 (1.8×10^5 manually curated) fluorescence measurements, in four experiments with 1,080 CD16/32 and 681 GATA1-mCherry onsets from 256 different HSC colonies were analysed (Extended Data Table 3). As expected, cells differentiating into the GM lineage had increased PU.1-eYFP levels over time and later expressed CD16/32 (Fig. 4b, left panels, Supplementary Video 1). Unexpectedly, we did not detect GATA1-mCherry expression at any point during GM differentiation. This is in contrast to expectations from the previous model in which the PU.1-GATA1 switch acts as the initiator of this lineage choice. Our detection limit for GATA1-mCherry is about 1,900 molecules per cell. In about half of all GM differentiations, PU.1-eYFP levels steadily increased from the starting HSC until the onset of CD16/32 expression. In the other half, PU.1-eYFP levels transiently dropped, then steadily increased. However, only $25 \pm 5\%$ of all GM time-course analyses showed PU.1 numbers dropping below 8,100 molecules (the average expression in HSCs). Moreover, only about $1 \pm 1\%$ of GM-differentiating cells transiently dropped to below 2,000 PU.1 molecules, and thus to similar levels of potentially maximally expressed GATA1 molecules. GATA1 levels thus do not have a relevant role during GM differentiation.

Cells that expressed detectable GATA1-mCherry, during HSC differentiation or in freshly sorted GATA1⁺ MPP2 cells, always further differentiated into PU.1-eYFP⁻GATA1-mCherry⁺ MegE cells (Fig. 4b, right panels, Supplementary Video 2). This confirms that GATA1 expression onset is a marker of MegE lineage commitment (Extended

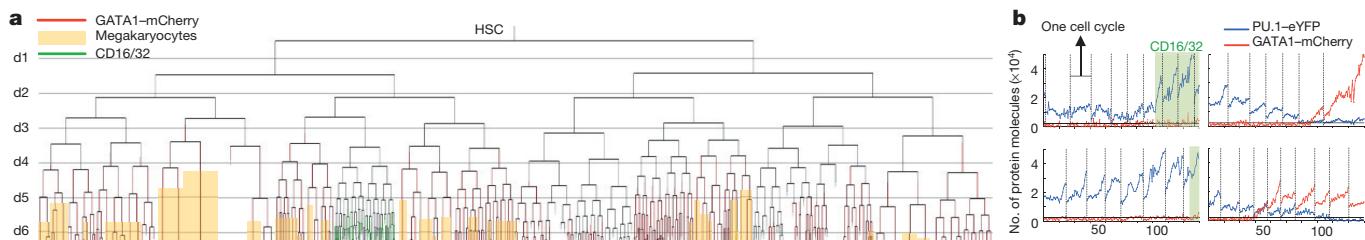


Figure 4 | Single-cell dynamics refute random PU.1:GATA1 ratios as initiators of early myeloid lineage choice. **a,** Single-cell genealogy of a differentiating HSC. CD16/32 detection by live antibody staining, megakaryocytes determined by cell morphology. **b,** Typical expression

dynamics of PU.1 and GATA1 expression of GM (left panels) or MegE differentiating cells (right panels). Black horizontal lines represent detection threshold for GATA1-mCherry.

Data Fig. 6). Importantly, this was independent of PU.1 levels during GATA1 expression onset, demonstrating that PU.1:GATA1 ratios do not regulate MegE commitment. In addition, PU.1-eYFP levels already decreased before detectable GATA1-mCherry expression in $63 \pm 19\%$ of cases (compare Fig. 4b, upper right panel). Thus, GATA1 expression is not the cause of PU.1 downregulation. In all remaining cases, GATA1 was upregulated, while PU.1 was still expressed at different levels. However, in these cases, cells later always differentiated into a PU.1⁻GATA1^{high} MegE state, showing that different PU.1 levels during lineage decision-making are irrelevant for the onset of GATA1 expression.

In conclusion, we did not observe a reproducible PU.1–GATA1 double-positive stage through which all differentiating HSCs pass (Supplementary Video 3). PU.1 and GATA1 are independently regulated at the start of GM or MegE differentiation. These observed protein dynamics are incompatible with random and cross-regulatory PU.1–GATA1 co-expression acting as the central mechanism that initiates MegE versus GM lineage choice³. This conclusion is in line with observations that lineage choice is still possible after deletion of PU.1 or GATA1, where only further maturation of committed cells is impaired^{1,2,30}. Our data are compatible with that from other reporter mice²³. However, our results also demonstrate discrepancies between protein and mRNA expression in uncommitted HSPCs^{8,31}, the latter of which had originally led to the development of currently accepted models.

Although we demonstrate that ratios of total PU.1:GATA1 protein numbers are not the central mechanism that initiates HSPC lineage decisions, we cannot exclude the possibility that only a very small subset of expressed PU.1 proteins may be actively competing with potentially existing GATA1 protein expressed below our detection limit. The PU.1–GATA1 switch may be involved in the lineage choice of other cell types²⁴ not analysed here. Likewise, other TF switches could be involved in GM versus MegE or other lineage choices. We conclude that physical PU.1–GATA1 interaction and antagonism^{6,7} could serve as an execution and/or reinforcing mechanism making terminal differentiation irreversible, but not as a decision-making mechanism inducing it. We expect other TFs and signalling pathways activated by extracellular signals to be upstream regulators of lineage-specific TFs, and their complex interplay will be of interest for future analyses.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 June 2015; accepted 13 May 2016.

- Pevny, L. et al. Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* **349**, 257–260 (1991).
- Scott, E. W., Simon, M. C., Anastasi, J. & Singh, H. Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science* **265**, 1573–1577 (1994).
- Graf, T. & Enver, T. Forcing cells to change lineages. *Nature* **462**, 587–594 (2009).
- Tsai, S. F., Strauss, E. & Orkin, S. H. Functional analysis and *in vivo* footprinting implicate the erythroid transcription factor GATA-1 as a positive regulator of its own promoter. *Genes Dev.* **5**, 919–931 (1991).
- Chen, H. et al. PU.1 (Spi-1) autoregulates its expression in myeloid cells. *Oncogene* **11**, 1549–1560 (1995).
- Zhang, P. et al. PU.1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding. *Blood* **96**, 2641–2648 (2000).
- Nerlov, C., Querfurth, E., Kulessa, H. & Graf, T. GATA-1 interacts with the myeloid PU.1 transcription factor and represses PU.1-dependent transcription. *Blood* **95**, 2543–2551 (2000).
- Miyamoto, T. et al. Myeloid or lymphoid promiscuity as a critical step in hematopoietic lineage commitment. *Dev. Cell* **3**, 137–147 (2002).
- Huang, S., Guo, Y.-P., May, G. & Enver, T. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev. Biol.* **305**, 695–713 (2007).
- Akashi, K. et al. Transcriptional accessibility for genes of multiple tissues and hematopoietic lineages is hierarchically controlled during early hematopoiesis. *Blood* **101**, 383–389 (2003).

- Månsson, R. et al. Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors. *Immunity* **26**, 407–419 (2007).
- Hoppe, P. S., Couto, D. L. & Schroeder, T. Single-cell technologies sharpen up mammalian stem cell research. *Nat. Cell Biol.* **16**, 919–927 (2014).
- Akashi, K., Traver, D., Miyamoto, T. & Weissman, I. L. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**, 193–197 (2000).
- Orkin, S. H. & Zon, L. I. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631–644 (2008).
- Etzrodt, M., Ende, M. & Schroeder, T. Quantitative single-cell approaches to stem cell research. *Cell Stem Cell* **15**, 546–558 (2014).
- Kueh, H. Y., Chamhakear, A., Nutt, S. L., Elowitz, M. B. & Rothenberg, E. V. Positive feedback between PU.1 and the cell cycle controls myeloid differentiation. *Science* **341**, 670–673 (2013).
- Kirstetter, P., Anderson, K., Porse, B. T., Jacobsen, S. E. W. & Nerlov, C. Activation of the canonical Wnt pathway leads to loss of hematopoietic stem cell repopulation and multilineage differentiation block. *Nat. Immunol.* **7**, 1048–1056 (2006).
- Rosenbauer, F. et al. Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. *Nat. Genet.* **36**, 624–630 (2004).
- Moreau-Gachelin, F. et al. Spi-1/PU.1 transgenic mice develop multistep erythroleukemias. *Mol. Cell. Biol.* **16**, 2453–2463 (1996).
- Heyworth, C., Pearson, S., May, G. & Enver, T. Transcription factor-mediated lineage switching reveals plasticity in primary committed progenitor cells. *EMBO J.* **21**, 3770–3781 (2002).
- Filipczyk, A. et al. Network plasticity of pluripotency transcription factors in embryonic stem cells. *Nat. Cell Biol.* **17**, 1235–1246 (2015).
- Pronk, C. J. H. et al. Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. *Cell Stem Cell* **1**, 428–442 (2007).
- Arinobu, Y. et al. Reciprocal activation of GATA-1 and PU.1 marks initial specification of hematopoietic stem cells into myeloerythroid and myelolymphoid lineages. *Cell Stem Cell* **1**, 416–427 (2007).
- Iwasaki, H. et al. Identification of eosinophil lineage-committed progenitors in the murine bone marrow. *J. Exp. Med.* **201**, 1891–1897 (2005).
- Miyawaki, K. et al. CD41 marks the initial myelo-erythroid lineage specification in adult mouse hematopoiesis: redefinition of murine common myeloid progenitor. *Stem Cells* **33**, 976–987 (2015).
- 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).
- Rieger, M. A., Hoppe, P. S., Smejkal, B. M., Eitelhuber, A. C. & Schroeder, T. Hematopoietic cytokines can instruct lineage choice. *Science* **325**, 217–218 (2009).
- Eilken, H. M., Nishikawa, S. & Schroeder, T. Continuous single-cell imaging of blood generation from haemogenic endothelium. *Nature* **457**, 896–900 (2009).
- Hilsenbeck, O. et al. Software tools for single-cell tracking and quantification of cellular and molecular properties. *Nat. Biotechnol.* <http://dx.doi.org/10.1038/nbt.3626> (2016).
- Mancini, E. et al. FOG-1 and GATA-1 act sequentially to specify definitive megakaryocytic and erythroid progenitors. *EMBO J.* **31**, 351–365 (2012).
- Hu, M. et al. Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev.* **11**, 774–785 (1997).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to S. Ammersdoerfer, H. Oller, C. Raithel, B. Vogel and A. Ziegler for technical support. This work was supported by an EMBO long-term fellowship to M.E.T., the ‘EUCOMM: Tools for Functional Annotation of the Mouse Genome’ (EUCOMMTOOLS) project (FP7-HEALTH-F4-2010-261492) to A.B., the ERC (starting grant Latent Causes), the German Federal Ministry of Education and Research (BMBF), the German Research Foundation (DFG) within the SPPs 1395 (InKoMBio) and 1356 to F.J.T., and DFG SFB 684 to T.S. and the SNF to T.S. T.S. and O.H. acknowledge financial support from SystemsX.ch.

Author Contributions P.S.H. planned and performed experiments and analysed data; M.Sc. programmed and applied quantitative imaging software and performed protein quantification and statistical analysis with M.St., C.M. and F.J.T. M.Sc., D.L., K.D.K., M.En., N.M., M.A.R., N.A., M.Et., and A.F. provided support for time-lapse imaging, flow cytometry and software development. O.H. and B.S. programmed single-cell tracking software with T.S. D.L.C. contributed to immunofluorescence staining. I.B., H.L. and A.B. contributed to generation of GATA1^{mCherry} mice. O.E., A.G. and C.N. provided the PU.1^{eYFP} mouse and competitive transplantations. F.J.T. designed and supervised the data analysis and modelling part. T.S. designed the study, programmed software, analysed data and wrote the paper with P.S.H. All authors read and commented on the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.S. (timm.schroeder@bsse.ethz.ch).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Generation of *Gata1-mCherry* knock-in mice. The knock-in construct was cloned by using conventional restriction-enzyme-mediated cloning and recombinering³² using the BAC RPCIB731C02198Q (Source BioScience) that contained the *Gata1* locus. The final knock-in construct consisted of a 5.0 kbp 5'-homology arm lasting until the last codon of *Gata1* (skipping the endogenous stop-codon) followed by a short linker sequence (5'-AGAGCATCAGGTACCAAGTGGAGCT-3'), the coding sequence for mCherry³³ and a FRT (FLP recognition target)-flanked phospho-glycine kinase (PGK) promoter-driven neomycin (neo) resistance gene and the 4.6 kbp 3'-homology arm. After removal of the neo selection marker, the *Gata1-mCherry* fusion mRNA transcript utilizes the endogenous 3' UTR.

JM8 mouse embryonic stem (ES) cell lines derived from the C57Bl/6N strain were grown on gelatinized tissue culture plates. Cells were maintained in knockout DMEM media (Gibco) supplemented with 2 mM glutamine, 1% β-mercaptoethanol (360 µl in 500 ml PBS, sterile-filtrated), 10–15% fetal calf serum (Invitrogen) and 500 U ml⁻¹ ESGRO leukaemia-inhibitory factor (Millipore). Electroporation of ES cells was carried out in a 25-well cuvette using the ECM 630 96-well electroporator / HT-200 automatic plate handler (BTX Harvard Apparatus; set at 700 V, 400 Ω, 25 µF). Immediately before electroporation, cell suspensions of ~10⁷ cells and ~2.5 µg of linearized targeting vector DNA were mixed in a final volume of 120 µl PBS. Cells were seeded onto a gelatinized 10-cm dish and colonies were picked after 8–9 days of puromycin (3 µg ml⁻¹) selection. The colonies were expanded in four copies of 96-well plates for archiving and characterization. Cells were frozen in supplemented knockout DMEM with 10% DMSO and stored in vapour over liquid nitrogen. After identification of positive clones, cells were thawed and expanded for aggregation.

Correctly targeted ES cell clones were identified by Southern blot using probes at the designated locations (Extended Data Fig. 1a) after digestion of genomic DNA with the restriction enzymes BamHI and XbaI, respectively. PCR primers for generating the Southern probes from BAC DNA were 5'-CAGCCACTG CCCAAATAGGTGGAG-3' and 5'-CTCCACCTATTGGCAGTGGCTG-3' (5'-probe) and 5'-CTGAAGTGGTCTGGACTTTAC-3' and 5'-TGAGGAAGA GGGAAAGGATGTGAAG-3' (3'-probe).

From one ES cell clone, germline chimaeras were generated by ES-cell aggregation with CD1 morulae and the FRT-flanked neo selection cassette was deleted *in vivo* by a Flp-e deleter strain by recombinase-mediated excision³⁴.

Animal experiments were approved by veterinary office of Canton Basel-Stadt, Switzerland and Regierung von Oberbayern.

Genotyping. PCR primers for checking presence or absence of the NEO cassette were 5'-GCATGGACGAGCTGTACAAG-3', 5'-CTGCACGAGAC TAGTGAGAC-3' and 5'-GCAGGAGAATGGAAATGTG-3' leading to a 223 bp band after successful removal. Unsuccessful removal would have led to a 387 bp band. Primers for checking the presence or absence of Flp-recombinase were 5'-GTTCTATATGCTGCCACTCC-3' and 5'-GAGCGATAAGCGTGTCTG-3' leading to 176 bp band at its presence. *GATA1-mCherry* mice were genotyped using the primers 5'-GCATGGACGAGCTGTACAAG-3', 5'-AGGACTGCCC ACCTCTATC-3' and 5'-GCAGGAGAATGGAAATGTG-3' leading to a 297 bp band in the case of wild-type *Gata1* and a 223 bp band in the case of *Gata1-mCherry*.

Isolation and staining of primary HSPCs and blood. Male and female mice for blood counts, bone marrow analysis, and time-lapse movies were killed at the age of 12–16 weeks. Blood counts were quantified on an Abc Animal Blood Counter (scil animal care company). Isolation of primary cells and flow cytometry sorting was performed as described^{13,22,35,36}. All flow cytometry was performed on a FACS Aria I or III (BD Bioscience). In brief, pelvis, femurs, tibiae, humeri and vertebrae of adult mice were isolated, crushed and incubated with anti-CD16/32 antibody (clone 2.4G2, BD Pharmingen, or clone 93, eBioscience) before staining with the desired antibody cocktail. Cells prepared for sorting HSCs were subjected to ACK Lysing Buffer (Lonza) after crushing, followed by lineage depletion using biotinylated antibodies against CD3e (clone 145-2C11), CD11b (clone M1/70), CD19 (clone 1D3), CD41 (clone MWReg30), B220 (clone RA3-6B2), Gr-1 (clone RB6-8C5) and TER-119 (clone TER-119, all eBioscience) and streptavidin-conjugated beads Roti-MagBeads (Carl Roth). The following antibodies were used for staining: anti-CD34 (RAM34), anti-CD48 (HM48-1), anti-CD105 (MJ7/18), anti-CD117 (2B8), anti-CD117 (ACK2), anti-CD135 (A2F10), anti-Sca-1 (D7, all eBioscience) and anti-CD150 (TC15-12F12.2, BioLegend). Different HSPC types within the lineage⁻Sca-1⁺c-Kit⁺ (LSK) population (Fig. 1b) were identified as:

HSC, LSK CD150⁺CD34⁻CD48⁻; MPP1, LSK CD150⁺CD34⁺CD48⁻; MPP2, LSK CD150⁺CD34⁺CD48⁺; MPP3/4, LSK CD150⁻CD34⁺CD48⁺. When cells were prepared for sorting myeloid progenitors, CD41-biotin was omitted. CD41 was also omitted for the analysis of the MPP subpopulations (Extended Data Fig. 5f).

Single-cells were sorted into 384-well plates (Greiner Bio-One), according to manufacturer's instructions, on a FACS AriaIII (BD Bioscience).

E14.5 fetal livers were isolated, individualized, and stained with antibodies for analysis. For GATA1 protein numbers quantification, GATA1-mCherry⁺ cells were sorted.

Colony assays. All colony assays were performed in Methocult GF M3434 (STEMCELL Technologies) according to manufacturer's instructions.

Competitive transplants. After ACK lysis (Life Technologies), freshly isolated bone marrow from CD45.1 homozygous C57Bl/6 wild-type mice was mixed 1:1 with bone marrow from either CD45.2 homozygous C57Bl/6 wild-type mice or CD45.2 homozygous PU.1^{eYFP}GATA1^{mCherry} (P/G) mice. Cells were frozen in 90% IMDM (Life Technologies) and 10% DMSO (Sigma Aldrich) and stored above liquid nitrogen until further usage. For transplants, cells were thawed, counted and 10⁶ living cells were transplanted into the tail vein of lethally irradiated CD45.1/CD45.2 heterozygous wild-type mice. Lineage contribution of donor cell mixtures was determined by bone marrow collection and staining²² after 6–7 weeks and plotted as the ratio of preMegE/preGM cells (Fig. 1i) and whole MegE/whole GM (Fig. 1l) lineage cells of wild-type:wild-type and P/G:wild-type donor mixes.

Immunostaining. Immunostaining was performed after permeabilization with 0.2% Triton-X (AppliChem) with 8 µg ml⁻¹ anti-PU.1 (T-21), 8 µg ml⁻¹ anti-GATA1 (N6) (both Santa Cruz), 10 µg ml⁻¹ anti-GFP (Aves Labs) and 5 µg ml⁻¹ anti-mCHERRY (ab167453) (Abcam) primary antibodies in 10% donkey serum in TBS-T (Tris-buffered saline, 0.1% Tween 20) overnight at 4°C, three washing steps of each 5 min and 10 µg ml⁻¹ Alexa Fluor dyes conjugated donkey secondary antibodies (Jackson ImmunoResearch) for 1 h at room temperature in 10% donkey serum in TBS-T. Images were acquired on a Nikon Eclipse Ti-E microscope. Fluorescent signals were quantified by segmentation of nuclear DAPI staining and background subtraction (without 'gain') as described³⁷. In order to compare transcription factors and their fusions regarding their stability, cells were kept in medium additionally supplied with 50 µM Cycloheximide and split into several vials. For immunostainings, cells were transferred to poly-L-lysine (Sigma Aldrich) coated object slides, stored for 10 min at 4°C, fixed with paraformaldehyde (Sigma Aldrich), stored at 4°C and stained (see above).

Western blot analysis. For western blotting, cells were spun down at designated time points and directly lysed in Laemmli-buffer, boiled at 100 °C and frozen at –20°C until further usage in SDS-PAGE analysis. For protein number quantification, designated amounts of recombinant GFP (Clontech) and mCherry (ChromoTek) were used. Gels were run on a 10% SDS-gel, blotted onto a PVDF-membrane (BioRad) and incubated with either anti-GFP (7.1, 13.1; Roche) or anti-mCherry antibody (1C51; abcam). Chemiluminescence was detected using horseradish peroxidase (HRP)-conjugated secondary antibodies (Jackson), ECL Plus Western Blotting Detection Reagents (GE Healthcare) and medical X-ray films (Fujifilm). Signal intensities were quantified using ImageJ software.

Cytospins. Cells were spun on object slides as per the manufacturer's instructions (Hettich), dried and stained with May-Grünwald (Carl Roth) and Giemsa solution (Sigma Aldrich).

Time-lapse image acquisition and tracking. Time-lapse imaging was performed at 37 °C in fibronectin (Takara Bio) coated channel slides μ-slide VI^{0,4} (ibidi), in StemSpan SFEM (STEMCELL Technologies) medium supplemented with 100 ng ml⁻¹ SCF, 100 ng ml⁻¹ TPO, 10 ng ml⁻¹ IL-3, 10 ng ml⁻¹ IL-6 (all mouse; PeproTech), 5 U ml⁻¹ EPO (human; PromoKine), 50 U ml⁻¹ penicillin, 50 µg ml⁻¹ Streptomycin (Invitrogen), self-labelled Alexa Fluor 647 (Invitrogen) anti-CD16/32 antibody (2.4G2) and 5% CO₂ using an Axio Oberserver Z1 microscope (Zeiss). A HXP 120 (Zeiss) was used as fluorescent light source. 46HE, 43HE (both Zeiss) and Cy5 (AHF) filter sets were used to detect eYFP, mCherry and Alexa Fluor 647, respectively, at exposure times between 400–1,500 ms using an AxioCam HRm (Zeiss). Brightfield pictures were acquired every 60–120 s, fluorescent pictures for the quantification of PU.1-eYFP and GATA1-mCherry were acquired every 30 min and every 3–4 h for the detection of CD16/32-Alexa-Fluor-647. Pictures used for quantifications were saved in lossless TIF or PNG format. Single-cell tracking and image quantification were performed using self-written software as described^{21,27–29}.

Inference of absolute protein numbers. *Western blot dilution assays.* For protein number quantification, known cell numbers of PU.1-eYFP^{hi} progenitors and GATA1-mCherry⁺ E14.5 fetal liver cells were resolved by western blotting on 10% polyacrylamide gels and compared with different levels of recombinant GFP protein (Clontech) or mCherry (Antibodies Online). PU.1-eYFP, GATA1-mCherry, GFP and mCherry proteins were detected using an anti-GFP primary antibody (Roche) or an anti-mCherry antibody (Abcam). All quantifications of band intensities were performed using ImageJ software by manually drawing a gate around the bands and subtracting the mean of the same area above and below the bands for primary HSPCs.

Model. After comparing several models to describe the data, we found a sigmoidal function to best describe the relationship between the dilution of the standard x and the observed intensity y :

$$y(x) = \left(\frac{\lambda x^n}{K^n + x^n} \right) \cdot \epsilon \quad (1)$$

Here, the exponent n determines the steepness of the sigmoidal, K sets the inflection point, λ is the maximum of the curve and ϵ is a lognormally distributed error term with expectation 1 and standard deviation σ as suggested for western blot data³⁸. This model outperformed other models, such as linear models with and without intercept according to the Bayesian information criterion and coefficient of variation between replicates. We estimate model parameters using 10,000 local optimizations initialized according to Latin-hypercube sampling.

We determine the number of proteins, P_j , per cell from the sample intensity, y_j , of replicate j (that is, western blot) as

$$P_j = \frac{K_j}{\left(\frac{\lambda_j}{y_j} - 1 \right)^{\frac{1}{n_j}}} \cdot \frac{1}{c_j \cdot w} \quad (2)$$

where c_j is number of loaded cells and w is the molecular weight for the protein of interest. The first term is obtained by solving equation (1) for x , whereas the second term relates dilution in nanogram to absolute protein numbers per cell. The parameters λ_j , K_j and n_j have been obtained by the local optimization.

Error propagation. As P_j is a combination of uncertain variables, we obtained error bars for each P_j individually by applying standard error propagation to account for uncertainties in the number of cells c_j (we assume a standard deviation of 10%) and uncertainties in the model (estimated via the standard deviation σ of our noise model ϵ_j in equation (1)). However, we find that the uncertainties for each individual replicate P_j are always with a factor <0.3 smaller than the inter-replicate standard deviations. Therefore, we only consider the standard deviation across replicates, as this is the dominant source of uncertainty in our procedure.

Mapping protein numbers to different HSPC populations. Inference. From reference cell types used for western blot dilution assays, we map mean protein numbers to all other cell types using the mean fluorescence intensities from flow cytometry. As we observe fluorescence intensities from fusion protein levels, we assume a linear relation between flow cytometry intensities and molecule numbers. We obtain the average protein amount per flow cytometry intensity of our reference cell type r as $\psi_r = \frac{\bar{P}^r}{\bar{MFI}^r}$. Here, \bar{MFI}^r is the mean fluorescence intensity of our protein of interest in the reference population (for example, PU.1-eYFP^{hi}) used for the western blotting, averaged over N replicates: $\bar{MFI}^r = \frac{1}{N} \sum_{i=1}^N MFI_i^r$. By \bar{P}^r we denote the estimated amount of protein from western blotting via equation (2), averaged over three replicates j .

We calculate the average amount of protein in any other population of interest (for example, GMPs) as

$$\overline{P}^{\text{GMP}} = \psi_r \cdot \overline{MFI}^{\text{GMP}} = \frac{\overline{MFI}^{\text{GMP}}}{\overline{MFI}^r} \cdot \bar{P}^r,$$

where $\overline{MFI}^{\text{GMP}}$ is the mean fluorescence intensity of our protein of interest in the GMP population averaged over N replicates.

Error Propagation. To obtain error bars for the protein amount in a population of interest, we perform error propagation, taking into account the uncertainty ΔP^r in protein numbers (as described above, we only consider inter-replicate variation), as well as uncertainty in the mean fluorescence intensities (standard deviation ΔMFI^r over the MFI_i^r) (for example, GMPs):

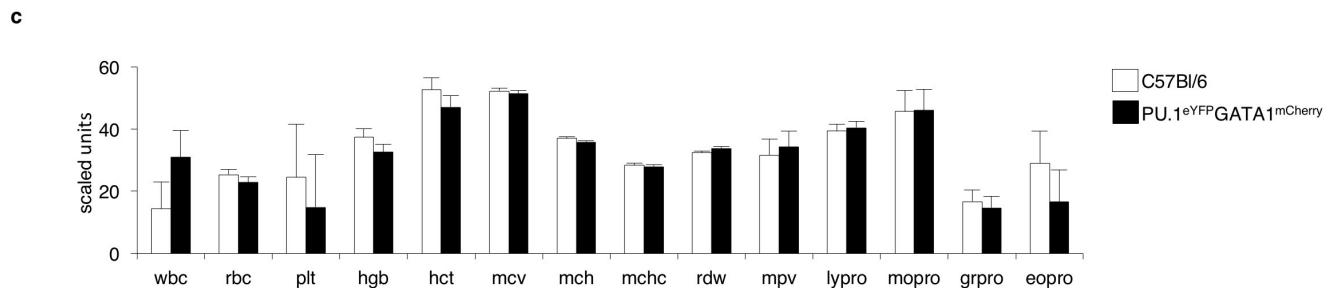
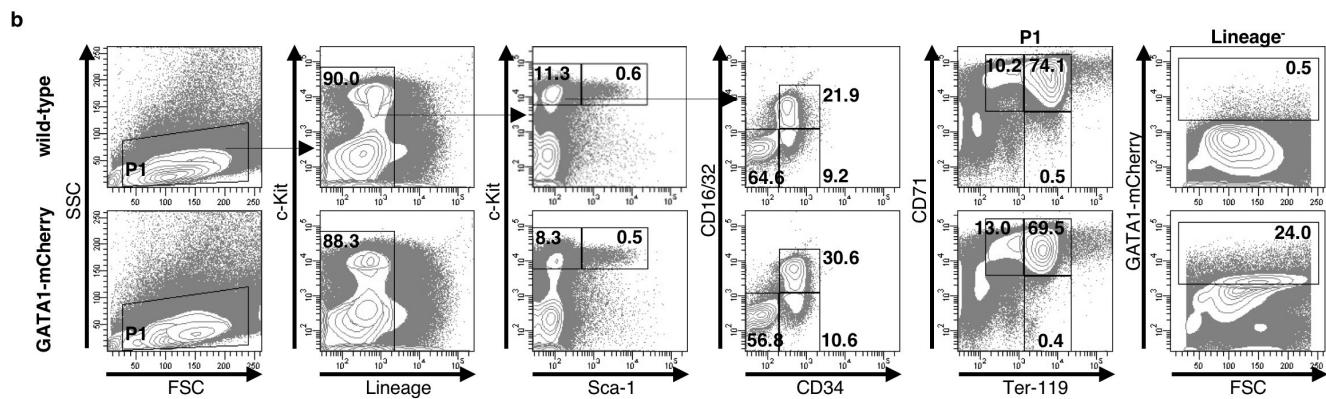
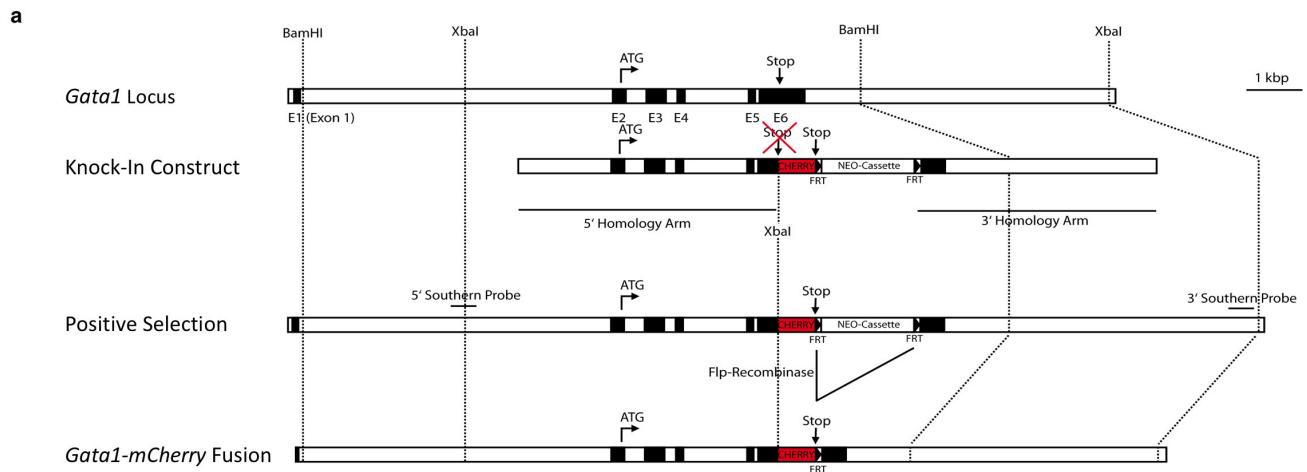
$$\Delta P^{\text{GMP}} = \sqrt{(\Delta MFI^{\text{GMP}})^2 \cdot \left(\frac{\bar{P}^r}{\bar{MFI}^r} \right)^2 + (\Delta MFI^r)^2 \cdot \left(\frac{\overline{MFI}^{\text{GMP}} \cdot \bar{P}^r}{(\bar{MFI}^r)^2} \right)^2 + (\Delta P^r)^2 \cdot \left(\frac{\overline{MFI}^{\text{GMP}}}{\bar{MFI}^r} \right)^2}$$

Mapping protein numbers to pixel intensities in imaging. The mean fluorescence intensity of the first time points was used to calibrate PU.1-eYFP protein abundance in time-lapse experiments. Whenever a movie cell exceeds twice the detection limit in the GATA1-mCherry channel for more than five consecutive time points, the cell itself and all its descendants were annotated as GATA1-mCherry positive. Mean protein abundance of GATA1 positive movie cells has been calibrated to the mean protein abundance of PU.1^{mid}-GATA1^{mid} in flow cytometry. Protein levels are then interpolated linearly³⁸.

Single-cell tracking and fluorescence quantification. Single-cell tracking was performed as described^{21,27–29}. Briefly, self-written software allows following individual cell identities over many days in order to generate genealogy trees. Fluorescence image normalization was applied as described³⁷. Custom written software semi-automatically identifies shapes of tracked cells and quantifies protein levels resulting in normalized intensity time traces independently of timing and location in time-lapse imaging. Detection thresholds were determined by simulating *in silico* background cells based on manually selected pixels containing only background signal and based on manually inspected cell areas. The 99% quartile of the resulting distribution of *in silico* background cells is referred to as the detection threshold, which is extrapolated to protein numbers for each movie as described above.

Code availability. Software used for single cell tracking and fluorescence quantification used in this study is published and open sourced²⁹.

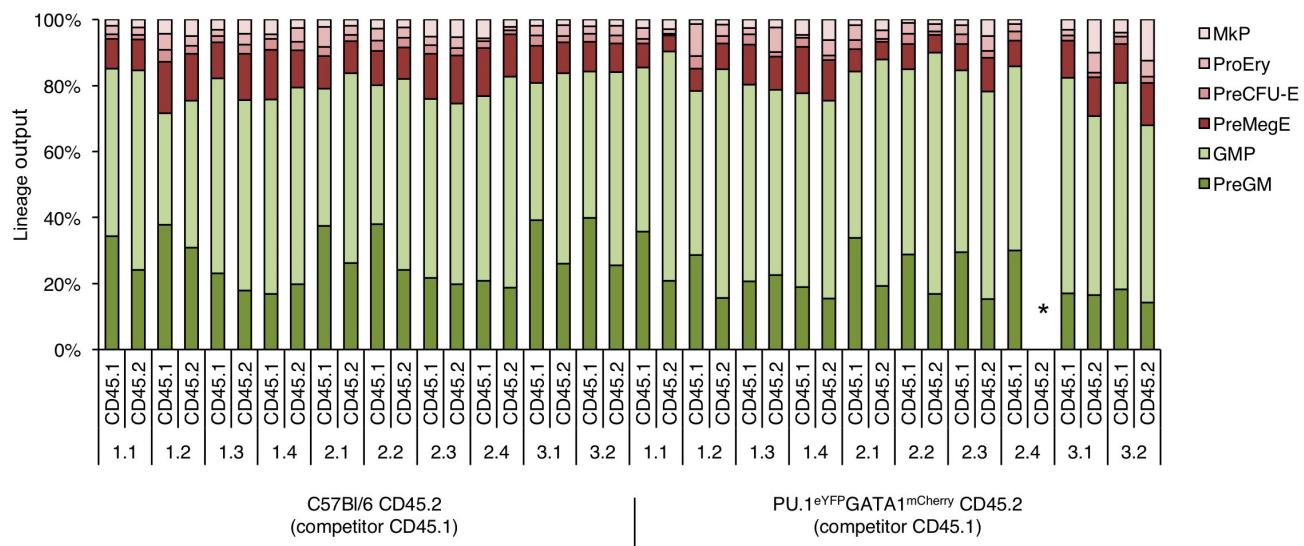
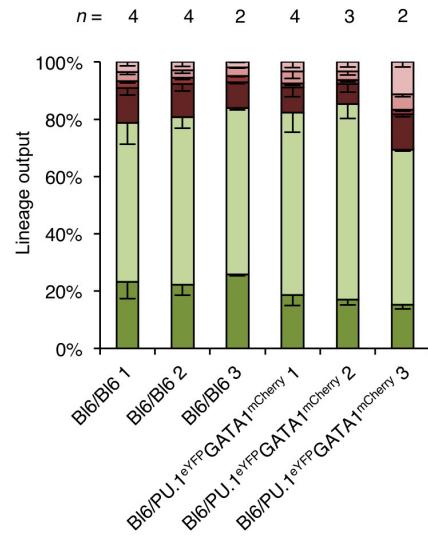
32. Liu, P., Jenkins, N. A. & Copeland, N. G. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res.* **13**, 476–484 (2003).
33. Shaner, N. C. et al. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol.* **22**, 1567–1572 (2004).
34. Dymecki, S. M. Flp recombinase promotes site-specific DNA recombination in embryonic stem cells and transgenic mice. *Proc. Natl Acad. Sci. USA* **93**, 6191–6196 (1996).
35. 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).
36. Wilson, A. et al. Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* **135**, 1118–1129 (2008).
37. Schwarzbacher, M. et al. Efficient fluorescence image normalization for time lapse movies. *Proc. Microsc. Image Anal. with Appl. Biol.* (2011).
38. Kreutz, C. et al. An error model for protein quantification. *Bioinformatics* **23**, 2747–2753 (2007).



Extended Data Figure 1 | Knock-in of fluorescent proteins does not alter the composition of fetal livers and adult peripheral blood.

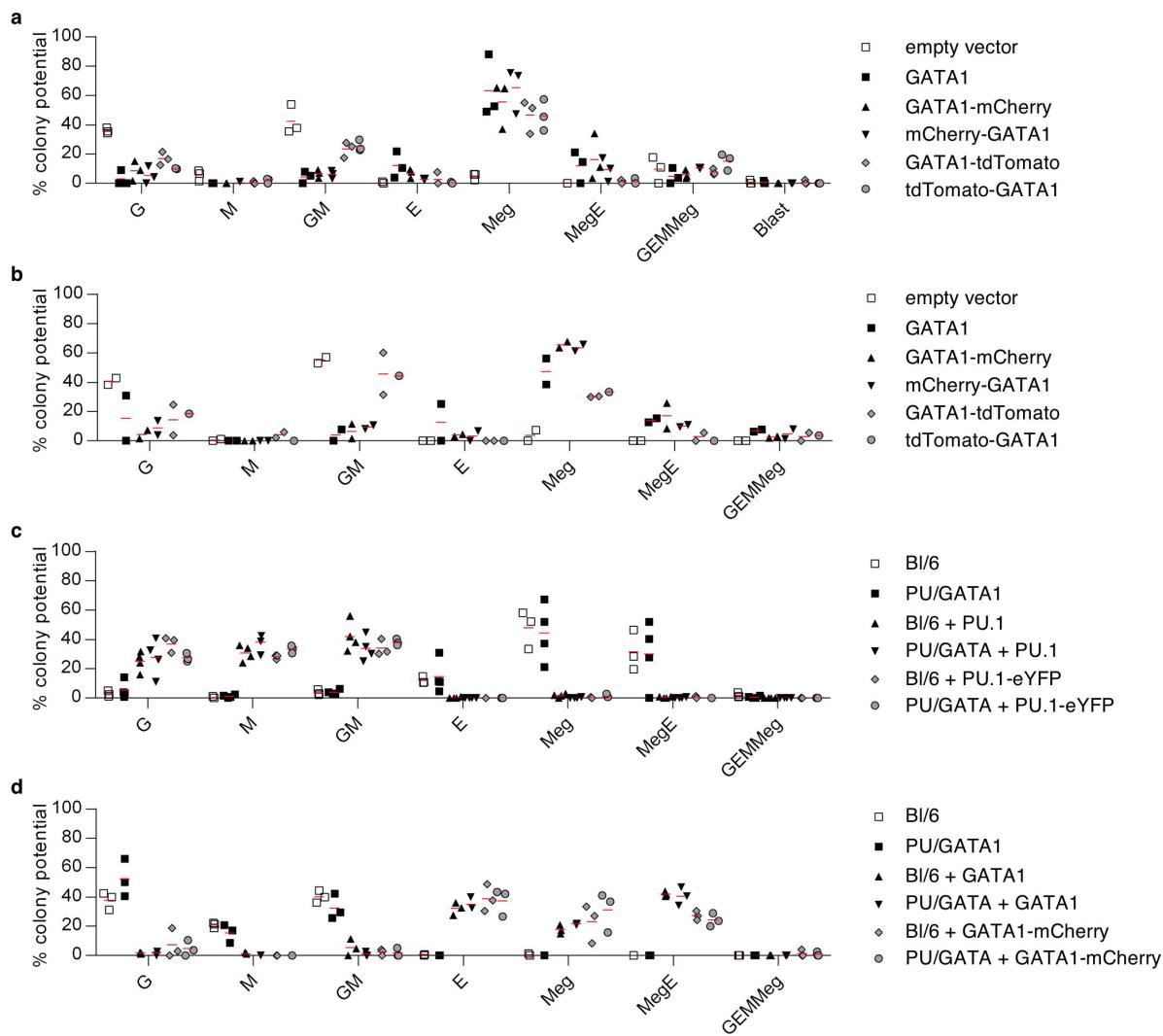
a, *Gata1-mCherry* knock-in strategy. BamHI and XbaI were chosen to generate restriction fragment length polymorphisms (from 11.1 kbp to 5.7 kbp in the case of XbaI and from 9.9 kbp to 11.1 kbp in the case of BamHI) to screen for successful knock-ins. Genomic sequences for Southern probes were identified at indicated positions. The final knock-in construct contained a 5.0 kilo base pairs (kbp) long 5' homology arm until the last codon of *Gata1*, a short linker sequence (5'-AGAGCATCAGGTACCACTGGAGCT-3'), the open reading frame (ORF) of mCherry, a FRT-flanked Neomycin-resistance cassette (including a eukaryotic and a prokaryotic promoter and a polyadenylation signal) and a 4.6 kbp long 3' homology arm.

embryos were collected, subjected to Ficoll-density centrifugation, pooled (C57Bl/6, 7 fetal livers; GATA1-mCherry, 6 fetal livers) and analysed by flow cytometry. Shown are percentages of the parental gate. **c**, Peripheral blood counts of adult mice (C57Bl/6 $n = 6$ biological replicates; PU.1^{eYFP} GATA1^{mCherry} $n = 9$; error bars, mean + s.d.). wbc, white blood cells (200 cells per μL); rbc, red blood cells (4×10^5 cells per μL); plt, platelets (20 cells per μL); hgb, haemoglobin (0.4 g dL^{-1}); hct, haematocrit (%); mcv, mean corpuscular volume (μm^3); mch, mean corpuscular haemoglobin (0.4 pg); mchc, mean corpuscular haemoglobin concentration (g dL^{-1}); rdw, red cell distribution width (0.4%); mpv, mean platelet volume ($0.2 \mu\text{m}^3$); lypro, % lymphocytes of wbc (2%); mopro, percentage monocytes of wbc (0.1%); grpro, percentage of granulocytes of wbc (%); eopro, percentage eosinophils of wbc (0.2%). No significant difference, one-way MANOVA ($P > 0.09$).

a**b**

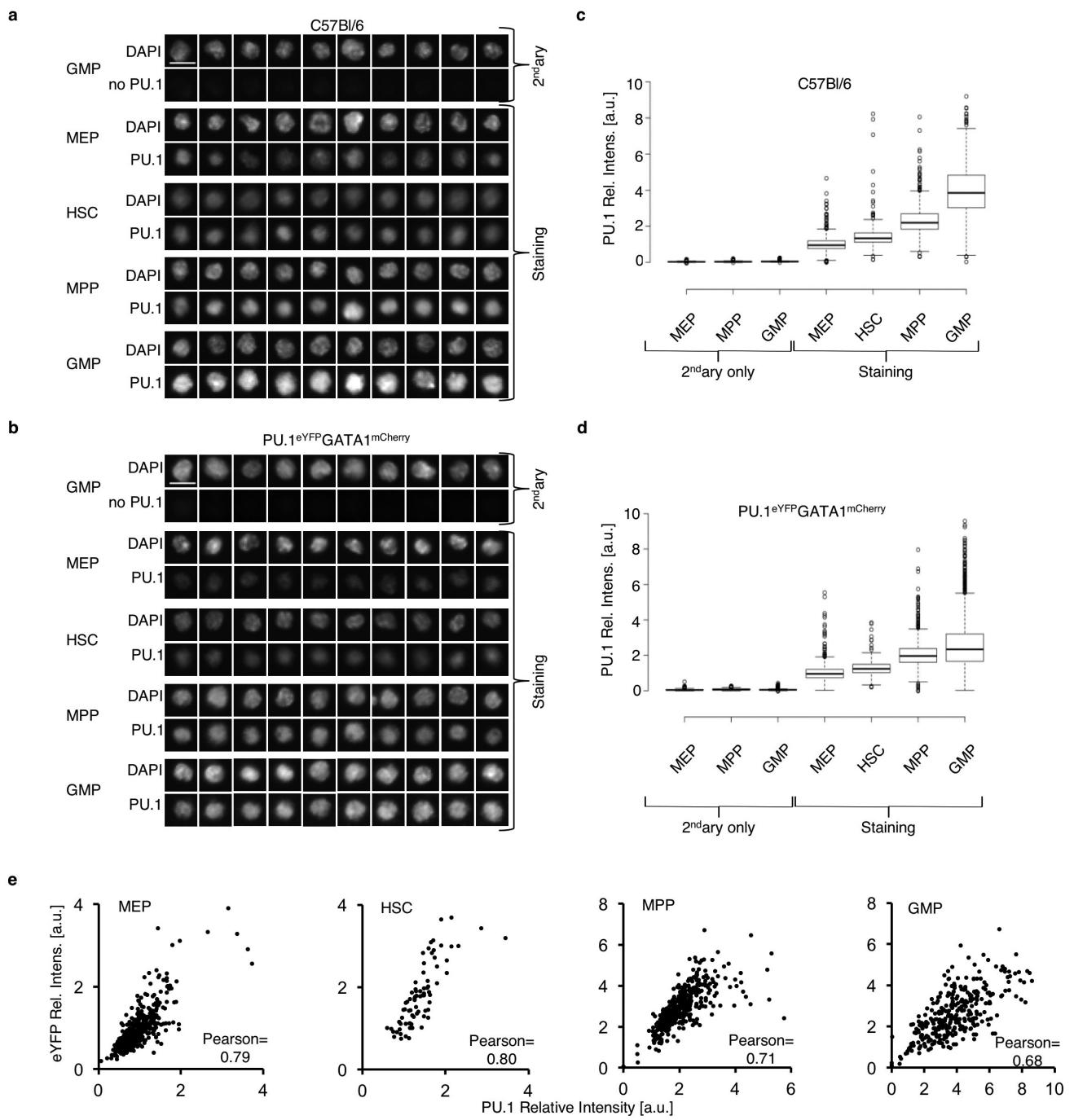
Extended Data Figure 2 | MegE versus GM-lineage differentiation of GATA1-mCherry/PU.1-eYFP cells *in vivo* is unaltered upon competitive transplantation. **a,** 10^6 bone marrow cells each from CD45.1 C57Bl/6 and CD45.2 C57Bl/6 or from CD45.1 C57Bl/6 and CD45.2 PU.1^{eYFP}GATA1^{mCherry} mice were transplanted into lethally irradiated recipient mice and bone marrow progenitor cell composition was analysed after 6–7 weeks. X and Y (X.Y, e.g. 1.1)

denote donor pair (X) and recipient mouse (Y). MkP, megakaryocyte progenitor; proEry, proerythroblast; preCFU-E, pre-colony-forming unit erythrocyte; preMegE, premegakaryocyte-erythrocyte progenitor; preGM, pregranulocyte-macrophage progenitor. Asterisk (*) indicates data excluded owing to low donor contribution. **b,** Summarized bone marrow lineage contribution per donor pair (from a).



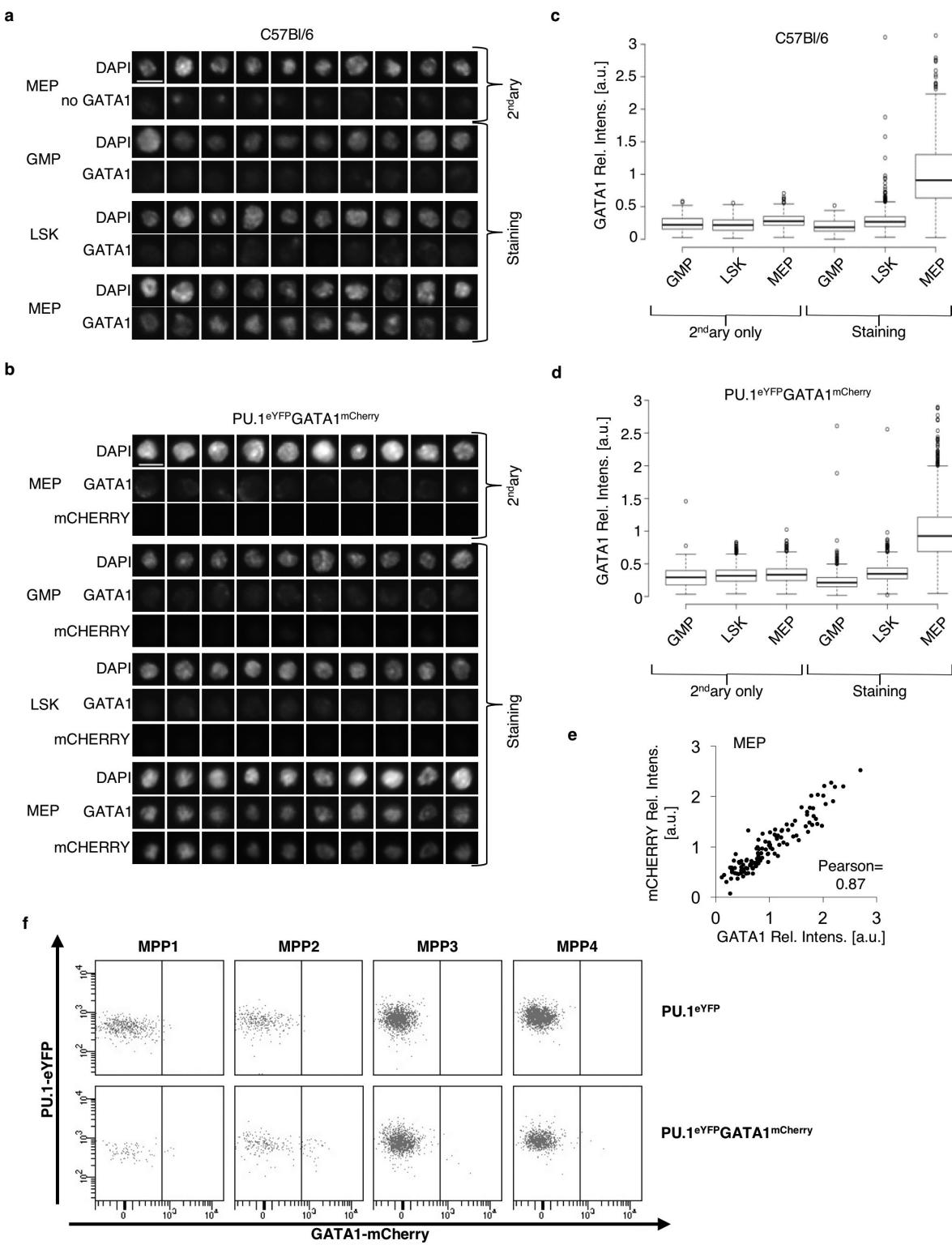
Extended Data Figure 3 | Normal reprogramming capacity of PU.1-eYFP and GATA1-mCherry in both wild-type and PU.1^{eYFP} GATA1^{mCherry} cells. **a, b**, LSK CD34⁺Flt3⁺ were sorted, transduced with lentivirus expressing the indicated proteins and plated in methylcellulose under permissive conditions. Cells from C57Bl/6 mice (**a**) (3 independent experiments) and cells from PU.1-eYFP knock-in mice ($n = 2$; except tdTomato-GATA1 $n = 1$) (**b**) were used. **c, d**, PreMegE cells (**c**) or preGM cells (**d**) from both C57Bl/6 wild-type and PU.1^{eYFP} GATA1^{mCherry} (PU/GATA) knock-in mice were sorted and transduced with mock, PU.1 or PU.1-eYFP expressing lentivirus (**c**) or with mock, GATA1 or GATA1-mCherry expressing lentivirus (**d**), respectively. After 24 h, cells were seeded in methylcellulose under permissive conditions. Colonies were scored after 8–10 days of culture

(3 independent experiments; except for C57Bl/6, PU/GATA, C57Bl/6 + PU.1 and PU/GATA + PU.1 $n = 4$). There was no significant difference between PU.1 and PU.1-eYFP overexpression in either C57Bl/6 or PU/GATA cells ($P > 0.23$; Kruskal-Wallis test). In contrast, C57Bl/6 and PU/GATA colonies without PU.1-eYFP overexpression were significantly different ($P < 0.05$ for M, Meg and E) (**c**). There was no significant difference between GATA1 and GATA1-mCherry overexpression in either C57Bl/6 or PU/GATA cells ($P > 0.77$; Kruskal-Wallis test). In contrast, C57Bl/6 and PU/GATA colonies without GATA1-mCherry overexpression were significantly different ($P < 0.007$, MegE) (**d**). Data are mean values. GEMMeg, granulocytic, erythroid, monocytic, megakaryocytic; MegE, megakaryocytic-erythroid; Meg, megakaryocytic; E, erythroid; GM, granulocytic-monocytic; M, monocytic; G, granulocytic.



Extended Data Figure 4 | PU.1 is expressed in nuclei of all HSCs, MPPs, GMPs and MEPs of wild-type and PU.1^{eYFP}GATA1^{mCherry} mice, and overlaps with eYFP expression. **a, b**, Indicated cell populations from wild-type C57Bl/6 (**a**) and PU.1^{eYFP}GATA1^{mCherry} (**b**) mice were sorted, fixed and probed with PU.1 antibody followed by staining with secondary antibody. Representative examples from controls ('secondary', without primary antibody) and staining of GMPs, MPPs, HSCs, and MEPs (CD150⁺ progenitors) are shown. DAPI stains nuclei. Scale bar, 10 µm. **c, d**, Quantifications of relative PU.1 expression levels determined by pixel intensities (rel. intens.). Data includes samples from three independent experiments (biological replicates), each of which was normalized to the mean expression levels of the respective MEP population. Individual data points for **c** are 884 MEP secondary only (475, 148 and 261 data points from the individual experiments), 1,218 MPP secondary only (553, 260

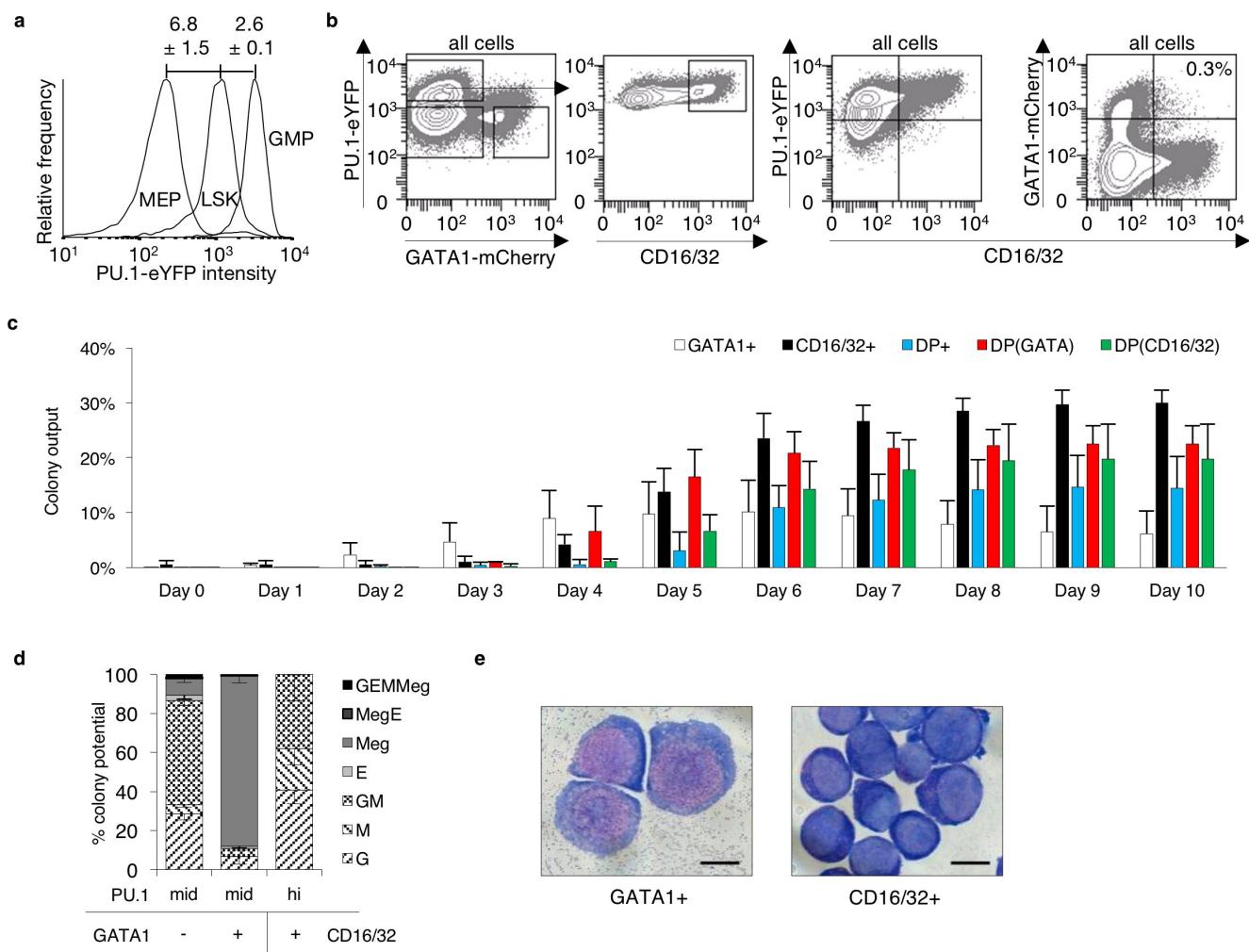
and 405), 755 GMP secondary only (599, 122 and 34), 1,213 MEP (659, 371 and 183), 603 HSCs (360, 194 and 49), 1,458 MPP (749, 283 and 426) and 819 GMP (571, 183 and 65). Individual data for **d** points are 1,530 MEP secondary only (739, 449 and 342 data points from the individual experiments), 1,194 MPP secondary only (547, 394 and 253), 1,866 GMP secondary only (1,616, 126 and 124), 1,521 MEP (518, 581 and 422), 273 HSCs (116, 79 and 78), 1,531 MPP (616, 463 and 452) and 2,339 GMP (1,351, 673 and 315). **e**, Correlation plot of PU.1 and eYFP staining in the indicated cell populations from one experiment. Pixel intensities were normalized to the mean expression in MEPs. Pearson correlation coefficients are displayed. Mean Pearson correlation values (\pm s.d.) of all three independent experiments were 0.82 ± 0.04 (MEP), 0.69 ± 0.10 (HSC), 0.71 ± 0.06 (MPP) and 0.63 ± 0.08 (GMP).



Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | GATA1 is expressed in nuclei of MegE committed cells of wild-type and PU.1^{eYFP}GATA1^{mCherry} mice, and overlaps with mCherry expression. **a, b**, Indicated cell populations from wild-type C57Bl/6 (**a**) and PU.1^{eYFP}GATA1^{mCherry} (**b**) mice were sorted, fixed and probed with GATA1 (**a, b**) and mCherry (**b**) antibody followed by staining with secondary antibodies. Representative examples from controls ('secondary', without primary antibody) and staining of GMP, LSK and MEP (CD150⁺ progenitors) are shown. DAPI stains nuclei. Scale bar, 10 μm. **c, d**, Quantifications of relative GATA1 expression levels determined by pixel intensities. Data includes samples from three independent experiments (biological replicates), each of which was normalized to the mean expression levels of the respective MEP population. Individual data points for **c** are 292 GMP secondary only (56, 188 and 48 data points from the individual experiments), 698 LSK secondary only (155, 287 and 256), 563 MEP secondary only (308, 216

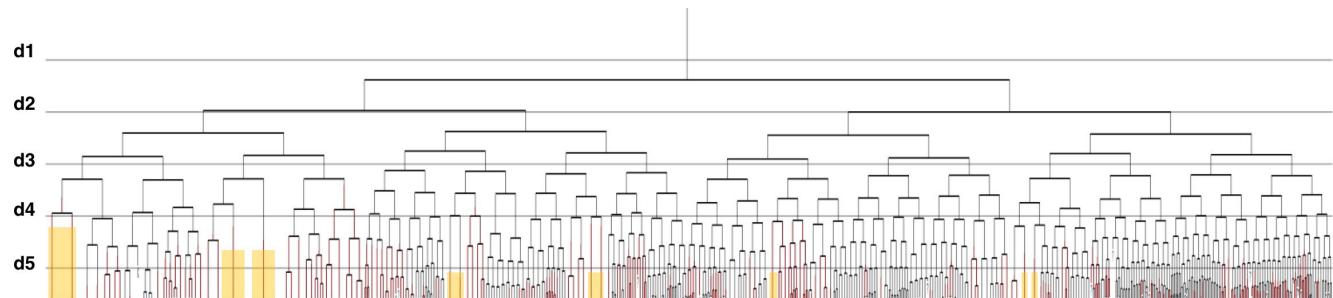
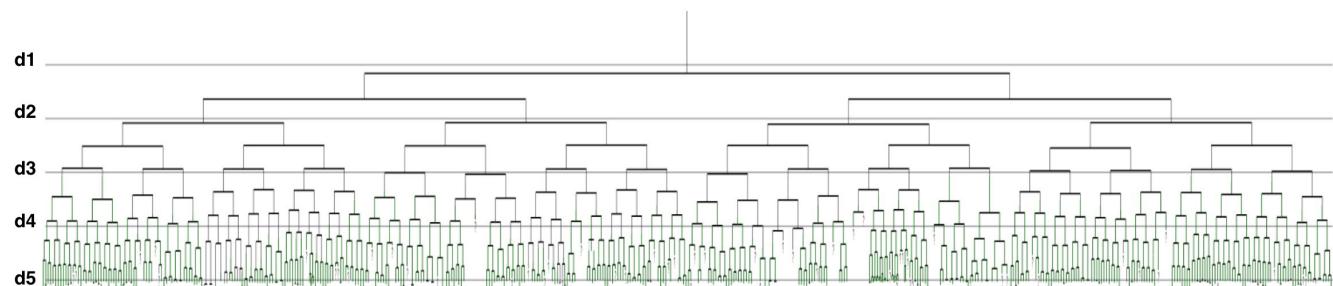
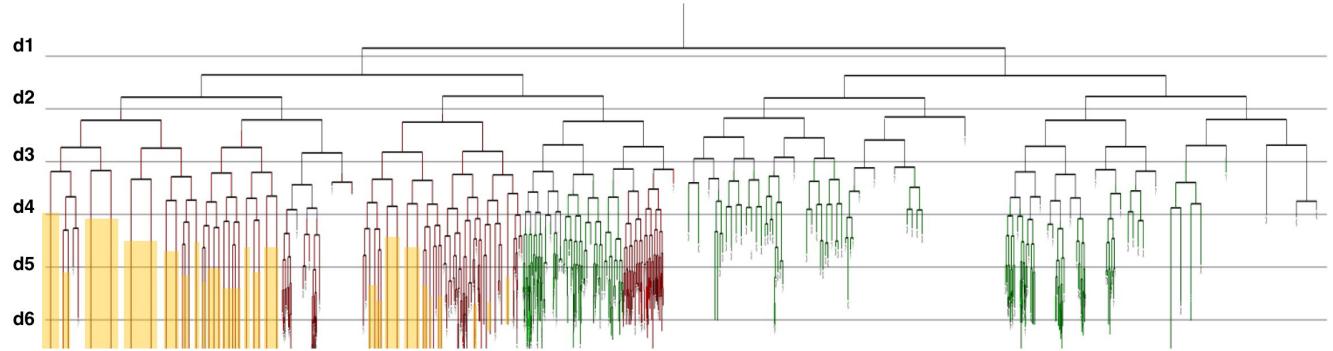
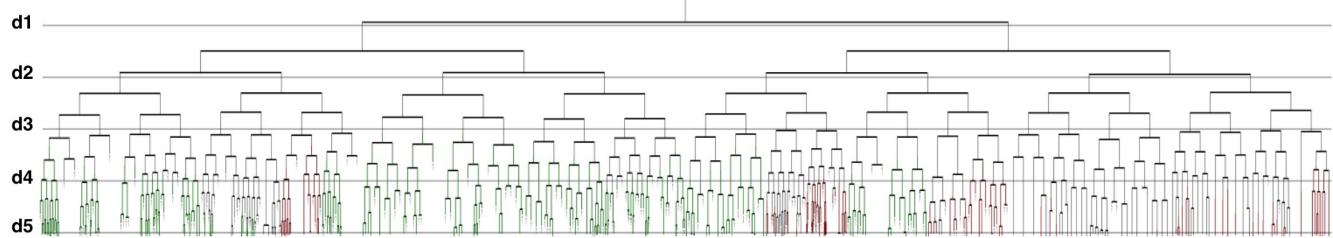
and 39), 344 GMP (64, 158 and 122), 1167 LSK (394, 294 and 479) and 590 MEP (252, 155 and 183). Individual data points for **d** are 485 GMP secondary only (171, 173 and 141 data points from the individual experiments), 1,295 LSK secondary only (360, 552 and 383), 886 MEP secondary only (561, 203 and 122), 462 GMP (73, 114 and 275), 1,184 LSK (252, 441 and 491) and 865 MEP (632, 115 and 118). **e**, Correlation plot of GATA1 and mCherry staining in the indicated cell populations from one experiment. Pixel intensities were normalized to the mean expression in MEPs. The mean Pearson correlation value of all three independent experiments was 0.87 ± 0.06 (s.d.). **f**, Representative example of GATA1-mCherry expression in MPP1–4 from three independent experiments (biological replicates). GATA1-mCherry is also expressed in some MPP2 cells³⁶. These are already committed to the MegE lineage (data not shown). Upper panels PU.1^{eYFP} mouse (negative control), lower panels PU.1^{eYFP}GATA1^{mCherry} mouse.



Extended Data Figure 6 | CD16/32 and GATA1-mCherry expression can be used as a lineage-marker for the GM and MegE lineage, respectively. **a**, Overlapping PU.1-eYFP expression histograms of MEPs, LSKs and GMPs show that PU.1-eYFP levels alone are not sufficient to attribute individual cells to a specific HSPC population. Mean ± s.d.; $n = 4$ (biological replicates); one representative example shown. **b**, Flow cytometry of day 4 HSC culture. Representative example from three independent experiments (biological replicates). GATA1-mCherry and CD16/32 are mutually exclusive. **c**, Single HSCs were sorted into single wells of a 384-well plate. Colonies were observed for 10 days by one brightfield image per day. Expression of PU.1-eYFP, GATA1-mCherry, and CD16/32 was qualitatively assessed on each day. Colonies were scored into exclusive GATA1-mCherry⁺ (white bars), exclusive CD16/32⁺, or

GATA1-mCherry⁺ and CD16/32 double positive (DP) colonies (57%). DP colonies were further subdivided into colonies that started to express GATA1-mCherry and CD16/32 on the same day (blue bars), expressed GATA1-mCherry at least one day before CD16/32 (red bars), or expressed CD16/32 before GATA1-mCherry (green bars). Missing percentages to 100% mean that colonies have either not expressed any marker yet or that individual colonies have died during the course of 10 days. All surviving colonies have turned on at least one marker (GATA1-mCherry or CD16/32) by day 9. Mean (+s.d.) of three biological replicates ($n = 141, 185, 129$ colonies). **d**, Colony potential of sorted cells from **b**, mean ± s.d. ($n = 3$). **e**, Cytospin of cells from day 4 cultures, representative example from three independent experiments. Scale bars, 10 μ m.

— GATA1-mCherry
 Megakaryocyte
— CD16/32



Extended Data Figure 7 | Additional HSC genealogy examples. Example trees with mixed GM/MegE (1st and 2nd tree), only GM (3rd tree) and only MegE differentiation (4th tree). Compare Fig. 4.

Extended Data Table 1 | PU.1^{eYFP}GATA1^{mCherry} mice are born at normal Mendelian ratios

	Offspring	Frequency	Expected
PU.1 and GATA1			
PU.1 ^{WT/WT} GATA1 ^{WT/Y}	3	5.8%	6.3%
PU.1 ^{WT/eYFP} GATA1 ^{WT/Y}	5	9.6%	12.5%
PU.1 ^{eYFP/eYFP} GATA1 ^{WT/Y}	1	1.9%	6.3%
PU.1 ^{WT/WT} GATA1 ^{mCherry/Y}	3	5.8%	6.3%
PU.1 ^{WT/eYFP} GATA1 ^{mCherry/Y}	9	17.3%	12.5%
PU.1 ^{eYFP/eYFP} GATA1 ^{mCherry/Y}	3	5.8%	6.3%
PU.1 ^{WT/WT} GATA1 ^{mCherry/WT}	2	3.8%	6.3%
PU.1 ^{eYFP/WT} GATA1 ^{mCherry/WT}	11	21.2%	12.5%
PU.1 ^{eYFP/eYFP} GATA1 ^{mCherry/WT}	5	9.6%	6.3%
PU.1 ^{WT/WT} GATA1 ^{mCherry/mCherry}	2	3.8%	6.3%
PU.1 ^{eYFP/WT} GATA1 ^{mCherry/mCherry}	5	9.6%	12.5%
PU.1 ^{eYFP/eYFP} GATA1 ^{mCherry/mCherry}	3	5.8%	6.3%
PU.1			
PU.1 ^{WT/WT}	10	19.2%	25.0%
PU.1 ^{eYFP/WT}	30	57.7%	50.0%
PU.1 ^{eYFP/eYFP}	12	23.1%	25.0%
GATA1			
GATA1 ^{WT/Y}	9	17.3%	25.0%
GATA1 ^{mCherry/Y}	15	28.8%	25.0%
GATA1 ^{mCherry/WT}	18	34.6%	25.0%
GATA1 ^{mCherry/mCherry}	10	19.2%	25.0%

Offspring of the mating male *PU.1^{eYFP/WT} Gata1^{mCherry/Y}* mice with female *PU.1^{eYFP/WT} Gata1^{mCherry/WT}* mice. Shown are real and expected frequencies for the respective genotypes.

Extended Data Table 2 | Protein abundance in different cell populations in flow cytometry and imaging

Flow Cytometry		
Cell population	PU.1eYFP	GATA1mCHERRY
PU.1 ⁺ GATA1- LK progenitor	43.1 ± 10.6 × 10 ³	NA
E14.5 GATA1 ⁺ fetal liver	NA	23.0 ± 9.8 × 10 ³
PU.1 ^{high} GATA1 ⁻	47.3 ± 12.9 × 10 ³	-- / < negative gate
PU.1 ^{mid} GATA1 ^{mid}	5.9 ± 1.8 × 10 ³	25.5 ± 12.3 × 10 ³
PU.1·GATA1 ^{high}	-- / < negative gate	54.6 ± 23.8 × 10 ³
HSC	8.1 ± 2.1 × 10 ³	-- / < negative gate
LSK	16.4 ± 4.3 × 10 ³	-- / < negative gate
GMP	42.7 ± 11.7 × 10 ³	-- / < negative gate
MEP	-- / < negative gate	49.4 ± 21.4 × 10 ³
Negative gate (GMP)	4.4 ± 1.2 × 10 ³	6.5 ± 3.8 × 10 ³
Negative gate (MEP)	4.7 ± 1.5 × 10 ³	8.4 ± 4.6 × 10 ³

Imaging		
Cell population	PU.1eYFP	GATA1mCHERRY
PU.1 ^{high} GATA1 ⁻	47.3 ± 12.9 × 10 ³	-- / < negative gate
PU.1 ^{mid} GATA1 ^{mid}	5.5 ± 4.4 × 10 ³	25.6 ± 16.3 × 10 ³
PU.1·GATA1 ^{high}	-- / < negative gate	54.6 ± 23.8 × 10 ³
HSC	8.1 ± 2.1 × 10 ³	-- / < negative gate
GMP	40.1 ± 4.7 × 10 ³	-- / < negative gate
Negative gate (in silico)	1.1 ± 2.0 × 10 ³	1.9 ± 4.4 × 10 ³

Calculated PU.1-eYFP and GATA1-mCherry protein molecule numbers both, for flow cytometry and imaging for the respective cell populations, as well as the negative gates (that is, the detection thresholds).

Extended Data Table 3 | Data overview from time-lapse movies

	Movie 1	Movie 2	Movie 3	Movie 4	All Movies
Starting cells	63	61	62	70	256
Early Apoptosis (<48h)	15	16	8	7	46
Early Apoptosis (<48h) %	23,8%	26,2%	12,9%	10,0%	18,2%
Lost without onset	5	1	0	3	9
Lost without onset %	7,9%	1,6%	0,0%	4,3%	3,5%
Trees with onsets	34	32	31	51	148
Trees with onsets %	54,0%	52,5%	50,0%	72,9%	57,4%
Trees without onsets	9	12	23	9	53
Trees without onsets %	14,3%	19,7%	37,1%	12,9%	21,0%
GM onsets	227	146	163	544	1080
ME onsets	89	230	93	269	681
Starting cells with >0 GM onset	31	20	19	31	101
Starting cells with >0 ME onset	8	16	17	29	70
Double positive trees	6	4	5	8	23
Trees with >1 GM (and no ME) onset	25	16	14	23	78
Trees with >1 ME (and no GM) onset	2	12	12	21	47
Deepest tracked division per tree (Mean)	6,6	6,0	7,0	7,8	6,9
Standard Deviation	4,5	4,4	3,2	3,7	4,0
Trees with max. 0 divisions	12	11	3	4	30
Trees with max. 1 divisions	2	3	1	3	9
Trees with max. 2 divisions	0	1	3	1	5
Trees with max. 3 divisions	0	0	1	2	3
Trees with max. 4 divisions	1	2	4	0	7
Trees with max. 5 divisions	1	1	4	4	10
Trees with max. 6 divisions	1	0	6	6	13
Trees with max. 7 divisions	2	2	4	4	12
Trees with max. 8 divisions	6	3	4	3	16
Trees with max. 9 divisions	8	8	12	6	34
Trees with max. 10 divisions	5	9	7	14	35
Trees with max. 11 divisions	6	2	3	5	16
Trees with max. 12 divisions	3	2	2	7	14
Trees with max. 13 divisions	1	0	0	1	2
Mean Onset per Tree (if >0 onsets)	6,0	4,1	5,1	5,4	5,2
Mean Standard Deviation	2,6	3,5	2,8	3,1	3,0
Mean Onset Generation 0	4	10	4	10	28
Mean Onset Generation 1	0	3	2	0	5
Mean Onset Generation 2	0	1	1	0	2
Mean Onset Generation 3	1	0	1	1	3
Mean Onset Generation 4	0	2	2	2	6
Mean Onset Generation 5	4	1	4	7	16
Mean Onset Generation 6	9	2	4	8	23
Mean Onset Generation 7	7	6	7	8	28
Mean Onset Generation 8	5	5	5	9	24
Mean Onset Generation 9	3	2	1	4	10
Mean Onset Generation 10	1	0	0	1	2
Mean Onset Generation 11	0	0	0	0	0
Mean Onset Generation 12	0	0	0	0	0
Mean Onset Generation 13	0	0	0	0	0

Information on tracked colonies per independent movie and their fate outcome. Number of colonies regarding their latest tracked division and mean marker onset (GATA1-mCherry and/or CD16/32). General tree fates include 'apoptosis' (<48 h after movie start), 'lost' (no information because cell identity lost), 'onset' (at least one marker onset per tracked tree) and 'no onset' (no marker onset in tracked tree).