nature cell biology

LETTERS

Inferring rules of lineage commitment in haematopoiesis

Cristina Pina[1], Cristina Fugazza[1], Alex J. Tipping[1,2], John Brown[1,2], Shamit Soneji[1,2], Jose Teles[3], Carsten Peterson[3] and Tariq Enver[1,4]

How the molecular programs of differentiated cells develop as cells transit from multipotency through lineage commitment remains unexplored. This reflects the inability to access cells undergoing commitment or located in the immediate vicinity of commitment boundaries. It remains unclear whether commitment constitutes a gradual process, or else represents a discrete transition. Analyses of *in vitro* self-renewing multipotent systems have revealed cellular heterogeneity with individual cells transiently exhibiting distinct biases for lineage commitment[1–6]. Such systems can be used to molecularly interrogate early stages of lineage affiliation and infer rules of lineage commitment. In haematopoiesis, population-based studies have indicated that lineage choice is governed by global transcriptional noise, with self-renewing multipotent cells reversibly activating transcriptome-wide lineage-affiliated programs[7]. We examine this hypothesis through functional and molecular analysis of individual blood cells captured from self-renewal cultures, during cytokine-driven differentiation and from primary stem and progenitor bone marrow compartments. We show dissociation between self-renewal potential and transcriptome-wide activation of lineage programs, and instead suggest that multipotent cells experience independent activation of individual regulators resulting in a low probability of transition to the committed state.

Transcription factors among other regulatory molecules have been shown to control cell differentiation, and in some cases play instructive roles in lineage identity, with the ability to reprogram unilineage-restricted cells to alternative lineage fates[8–14]. Elucidating how the regulators are activated in the multipotent state to participate in lineage commitment is central to understanding fate choices, and can be approached through characterization of their ground-state expression pattern in individual self-renewing and newly committed cells.

We conducted our study in the well-characterized multipotent haematopoietic cell line EML (ref. 15), as the use of a growth-factor-dependent cell line allows the capture of self-renewing and early committed states otherwise too transient to be identified in *in vivo* or *ex vivo* analyses of haematopoiesis. Population-based analyses of the EML cell line have indicated that cells can fluctuate between erythroid- and myeloid-lineage-biased states while retaining self-renewal potential, operationally defined as culture-reconstituting capacity[7]. These lineage-biased states are characterized by global expression of lineage-associated transcriptomes qualitatively and quantitatively comparable to differentiated cells (Supplementary Fig. S1a). Experimentally, the erythroid- and myeloid-biased states can be captured on the basis of surface expression of stem cell antigen 1 (Sca1; Supplementary Fig. S1b), with Sca1lo and Sca1hi cells representing the erythroid-biased and myeloid-biased compartment, respectively. When sorted, the Sca1lo compartment expresses higher levels of genes affiliated with the erythroid lineage (Supplementary Fig. S1c,d) and when cultured as a population, it reconstitutes cultures, regenerating the initial bell-shaped Sca1 distribution including Sca1hi cells (Fig. 1a and Supplementary Fig. S1e); this capacity is operationally equated with self-renewal in the context of the EML model cell line. Although population-based data such as these are consistent with a model of fluctuation between global lineage-biased states within multipotent cells (Supplementary Fig. S1a), such a model necessarily requires that self-renewal capacity and lineage-biased transcriptomes coexist in individual cells. By definition, this is not ascertainable with population-based studies and we therefore interrogated the functional and molecular properties of individual cells focusing our analysis primarily on specification of the erythroid lineage.

Examination of the Sca1lo population at the single-cell level revealed that Sca1lo cells are heterogeneous for expression of the erythroid-affiliated transcription factor encoded by *Gata1* (Fig. 1b and Supplementary Fig. S1f). Individual Sca1lo cells also differ in their culture-reconstituting capacity (Fig. 1c). Taken together, these results

[1]Stem Cell Laboratory, UCL Cancer Institute, University College London, Paul O'Gorman Building, 72 Huntley Street, London WC1E 6BT, UK. [2]Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK. [3]Computational Biology and Biological Physics, Department of Theoretical Physics, Lund University, Lund SE-223 62, Sweden.
[4]Correspondence should be addressed to T.E. (e-mail: t.enver@ucl.ac.uk)

Received 2 November 2011; accepted 12 January 2012; published online 19 February 2012; DOI: 10.1038/ncb2442

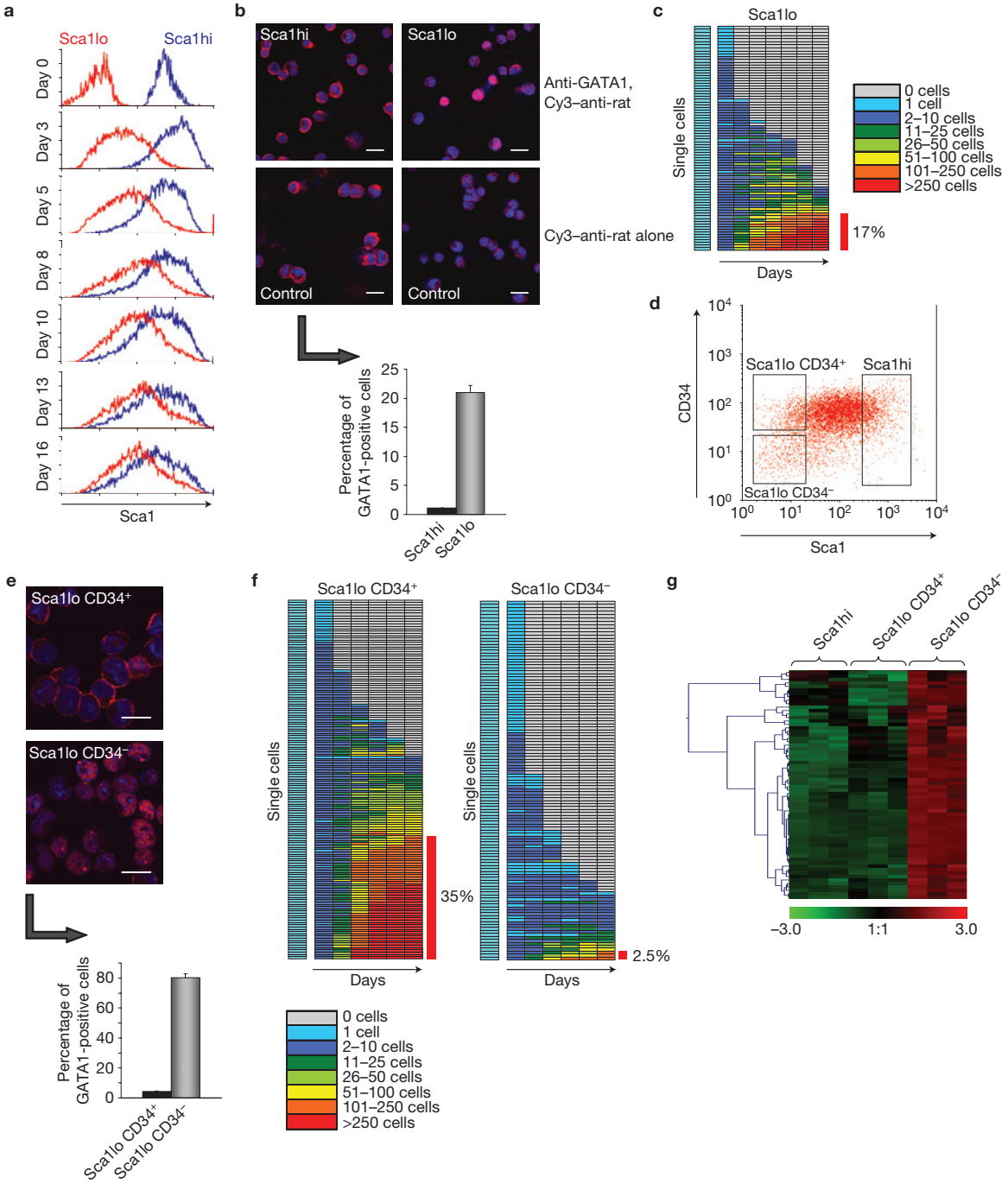

Figure 1 Individual self-renewing cells do not express transcriptome-wide lineage programs. (**a**) Reconstitution of the bulk Sca1 profile of EML cells from Sca1lo- (red) and Sca1hi- (blue) initiated cultures as analysed by flow cytometry. Corresponding growth plots are shown in Supplementary Fig. S1e; data are representative of two independent experiments. (**b**) The micrographs show immunostaining of EML Sca1hi (left) and Sca1lo (right) cells for GATA1 protein; the GATA1 signal is nuclear and present in a fraction of Sca1lo cells (top); control panels (bottom) are secondary antibody alone and show surface detection of the Sca1 antibody used for sorting; scale bars, 15 μm. The histogram summarizes two independent experiments scored by three independent observers (mean + s.e.m., $n = 3$). (**c**) Clonal analysis of the culture-reconstituting potential of Sca1lo EML cells. The detached column represents sorting of 1 cell per well; the subsequent columns represent scoring time points; rows represent progeny of single cells deposited at day 0. The plot summarizes two independent experiments. Large clones (>100 cells) reconstitute a broad Sca1 distribution. (**d**) Flow cytometry plot highlighting $CD34^+$ and $CD34^-$ populations within the Sca1lo compartment. Sca1lo encompasses the 15% lowest-staining cells; the 15% highest-staining Sca1hi cells are almost exclusively $CD34^+$. (**e**) GATA1 immunostaining of Sca1lo $CD34^+$ (top micrograph) and Sca1lo $CD34^-$ (bottom micrograph) cells and their respective quantification (histogram), plotted as in **b**; scale bars, 15 μm. The histogram summarizes two independent experiments scored by three independent observers (mean + s.e.m., $n = 3$). (**f**) Clonal analysis of the culture-reconstituting potential of Sca1lo $CD34^+$ (left) and Sca1lo $CD34^-$ (right) cells; set-up as in **c**. The data summarize two independent experiments. The three colonies of >100 cells resulting from Sca1lo $CD34^-$ cells were obtained in only one of the two experiments and did not reconstitute the original culture profile along the CD34 axis (data not shown). (**g**) Heat-map representation of the expression of genes in the erythroid-associated Sca1lo signature as per previously published microarray data[7] in Sca1hi, Sca1lo $CD34^+$ and Sca1lo $CD34^-$ EML cells. Differentials are defined as twofold up in Sca1lo versus Sca1hi at a *B* value >2; non-differential probes of genes differential in ref. 7 were excluded; the full gene list is given in Supplementary File S1.

raise the possibility that self-renewal capacity and lineage-affiliated gene expression are present in different cells within the Sca1lo population (Supplementary Fig. S1g). Subdividing the Sca1lo population on the basis of CD34 expression[16] supports this view (Fig. 1d). Sca1lo CD34$^-$ cells contain virtually all of the GATA1 protein expression (Fig. 1e) and are devoid of culture-reconstitution capacity (Fig. 1f); they have accelerated erythroid differentiation in response to erythropoietin (Epo; Supplementary Fig. S1h) and lack neutrophil potential (Supplementary Fig. S1i). Conversely, the Sca1lo CD34$^+$ compartment contains all of the Sca1lo culture-reconstitution capacity (Fig. 1f) and is multipotent (Supplementary Fig. S1h,i); GATA1 protein expression is minimal or absent in these cells (Fig. 1e). We thus show that the Sca1lo population contains two separate compartments: erythroid-committed Sca1lo CD34$^-$ cells (EryCP, erythroid-committed population) with no culture-reconstitution capacity; and a culture-reconstituting Sca1lo CD34$^+$ population (CRSca1lo) with erythroid and neutrophil differentiation potential. We then interrogated the transcriptional signatures of these two compartments for evidence of transcriptome-wide erythroid lineage bias. We used the erythroid signature defined by microarray analysis of unfractionated Sca1lo and Sca1hi EML cells[7]. This signature was present in the erythroid-committed EryCP and totally absent from culture-reconstituting CRSca1lo cells, which were indistinguishable from CRSca1hi (Fig. 1g).

Thus, interrogation of the co-residency of self-renewal capacity and lineage-biased transcriptomes in single multipotent cells did not support the notion that mutipotential cells sample global lineage programs before lineage commitment. Rather, previous observations of reversible molecular lineage bias[7] in the multipotent EML model system also used in this study could be accounted for by coexistence of separable culture-reconstituting and committed populations (Supplementary Fig. S1g).

The ability to capture these related culture-reconstituting and early committed populations from the same dynamical culture system allows exploration of the transcriptional composition of cells either side of a commitment boundary, and this may provide insight into the process of transition.

We first explored the early committed state represented by the EryCP compartment. Global gene expression profiling analysis demonstrates that EryCP cells are transcriptionally closer to both CRSca1lo and CRSca1hi compartments than to erythroid-differentiated EML cells (Fig. 2a). Consistent with being functionally committed to the erythroid lineage, EryCP cells show evidence of erythroid-affiliated gene expression programs, but these are significantly less developed, both qualitatively and quantitatively, than those present in erythroid-differentiated EML cells (Fig. 2b). This, together with their transcriptional proximity to culture-reconstituting cells, indicates that EryCP cells are at an early stage of erythroid commitment, as might be expected from their spontaneous generation in self-renewing culture conditions.

Qualitative differences in the erythroid programs of EryCP and erythroid-differentiated EML cells indicate that commitment can occur in the absence of a complete erythroid transcriptome. The quantitatively lower expression of components of the erythroid program could be the result of uniformly lower expression of those genes in all cells in the EryCP compartment, or may instead reflect cell-to-cell variation in gene expression.

We quantitatively analysed gene expression in multiple arrays of single cells using the Fluidigm dynamic array platform. The panel of genes analysed was representative of the EML-derived erythroid signature as well as being determined by genetics to be important in erythroid lineage development[17,18]. Specifically, *Hbb-y*, *Epb4.2*, *Klf1*, *Epor* and *Gata1* are erythroid-affiliated genes, *Sfpi1*, *Mpo* and *Gfi1* are myeloid-affiliated and the remaining genes are broadly expressed in multipotent cells, with most also active in the erythroid series.

This analysis reveals significant cell-to-cell variation in gene expression in early erythroid-committed cells (Fig. 2c). This is evident both in respect of erythroid-affiliated genes and of components of alternative lineage programs. The results indicate that commitment can occur in the absence of key lineage regulators such as *Gata1* and also without silencing of genes associated with alternative lineage programs such as *Mpo*. We investigated whether cells that did express *Gata1* exhibited less variable gene expression. These cells could be isolated on the basis of integrin αE (Itgae) expression (Fig. 2d). Although the program seemed more complete, as exemplified by *Klf1* expression, significant cell-to-cell variation was still apparent (Fig. 2e, left). Only a small number of cells with higher levels of *Klf1* seemed more homogeneous in their profiles and had downregulated myeloid components of the program; compatible with these observations, analysis of individual erythroid-differentiated EML cells revealed a more complete and homogeneous expression of erythroid-affiliated genes together with a more widespread silencing of myeloid genes (Fig. 2e, right).

We next considered the possibility that the heterogeneity in erythroid gene expression and the incomplete resolution of alternative lineage programs observed in EryCP cells reflected their erythroid commitment in the absence of erythropoietin. We therefore analysed gene expression in individual CD34$^-$ Itgae$^-$ cells generated in the presence of erythropoietin (Supplementary Fig. S2a,b) and found them to be similarly heterogeneous (Supplementary Fig. S2c). Using our present panel of genes, these cells were indistinguishable from CD34$^-$ Itgae$^-$ cells captured from self-renewal cultures (Fig. 3a,b and Supplementary Fig. S2c). Similarity in gene expression between these two cell types extended to the global scale as evidenced by transcriptome-wide microarray profiling (Fig. 3c). These *in vitro* data generated from cell lines indicate that commitment occurs in the presence of limited expression of erythroid lineage components that may also vary between individual cells.

We then investigated whether these findings extended to bone-marrow-derived haematopoietic components that have undergone commitment in a physiological setting. We sorted cells of the earliest defined steps of erythroid lineage restriction (pre-megakaryocytic/erythroid progenitors, preMegE) as well as the later colony-forming unit erythroid (CFUe) stage[19] (Fig. 3d).

Individual preMegE cells show considerable heterogeneity of erythroid-affiliated gene expression, which largely resolves in more differentiated CFUe (Fig. 3e). Notably, preMegE cells seem more advanced in their lineage progression than EML-derived EryCP cells, as evidenced by a higher expression frequency of *Klf1* and more widespread silencing of myelo-monocytic genes such as *Mpo*. This highlights the utility of the EML system in affording analysis of early post-commitment stages.

We next investigated how early erythroid-committed cells differed in molecular terms from self-renewing cells. We carried out Fluidigm

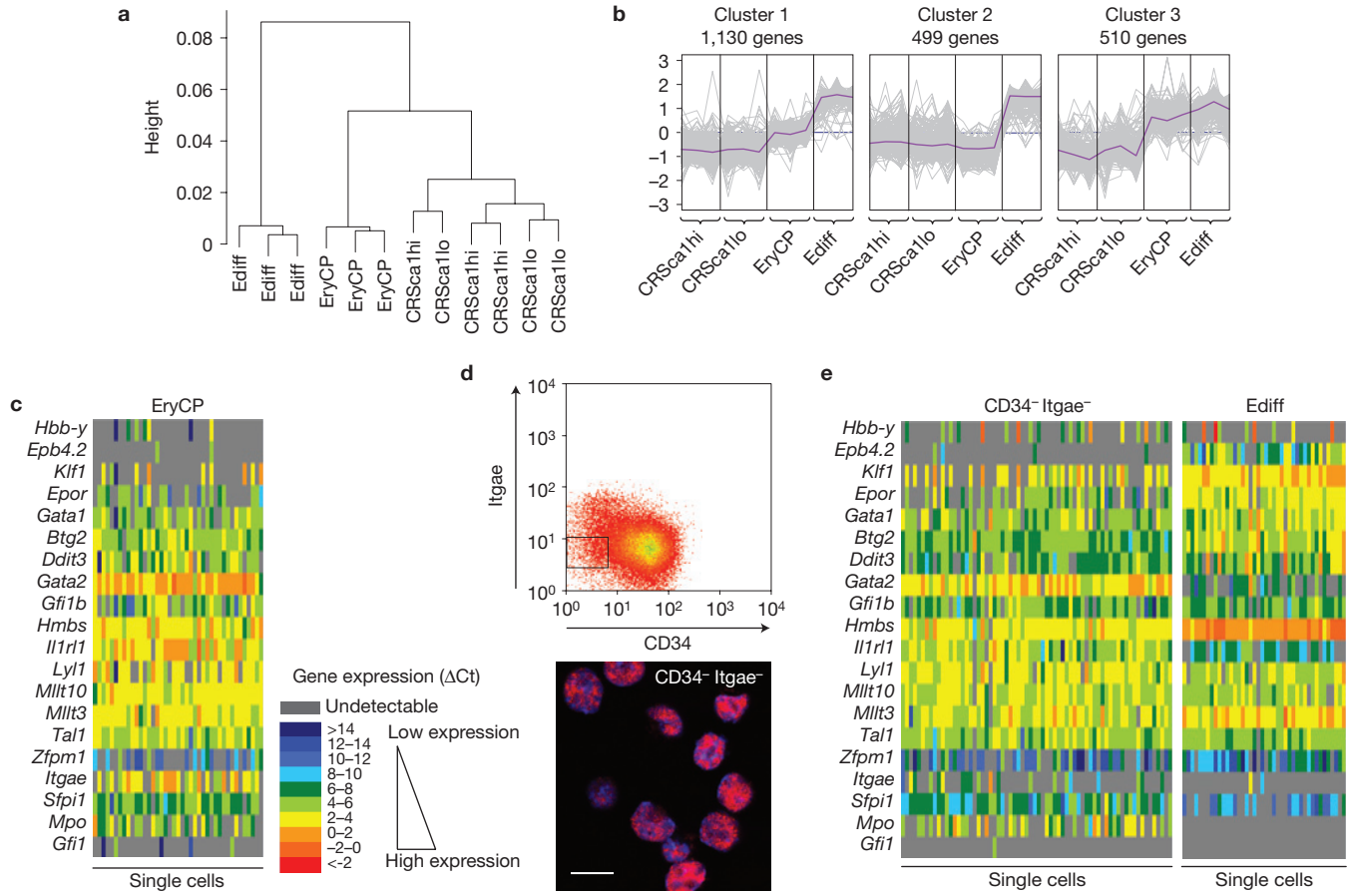

Figure 2 Post-commitment assembly of the EML erythroid program is gradual and heterogeneous. (**a**) Transcriptome dendrogram of culture-reconstituting, committed and differentiated cells. (**b**) K-means clustering of erythroid signature genes defined as Ediff versus both CRSca1hi and CRSca1lo cells (LIMMA *B* value>10, >2-fold upregulation in Ediff); see Supplementary File S2 for the full list of differentials. (**c**) Expression of lineage-relevant genes analysed in single EryCP cells ($n = 41$). The data are ΔCt to the reference gene *Atp5a1*, categorized and colour-coded as ΔCt steps; global value range: −1.24–16.41. Columns, individual cells; rows, genes. (**d**) Flow cytometry plot showing CD34$^-$ Itgae$^-$ cells in cultures under self-renewing conditions (top). These cells have nuclear GATA1 protein (bottom, immunostaining); scale bar, 15 μm. (**e**) Gene expression in individual CD34$^-$ Itgae$^-$ cells isolated from self-renewing cultures ($n = 68$, in two independent experiments; cells were kit$^+$), and in differentiated kit$^-$ cells (Ediff, $n = 42$). The data are plotted as in **c**; global ΔCt ranges: −1.93–14.85 (CD34$^-$ Itgae$^-$), −2.11–14.46 (Ediff).

analysis on the CD34$^+$ fraction of the Sca1lo compartment that we previously showed to contain all of the Sca1lo culture-reconstitution activity (see Fig. 1f). In contrast to early erythroid-committed cells (Fig. 2c,e), Sca1lo CD34$^+$ cells seem more similar to each other (Fig. 4a). These cells do express genes such as *Tal1*, *Gata2* or *Mllt3*, which have functional roles in erythropoiesis but equally are thought important in the multipotent state[20–23]. Note however that quantitative differences in gene expression between multipotent and committed cells are observable as exemplified by the general upregulation of *Mllt3* expression level in committed cells. A fifth of Sca1lo CD34$^+$ cells exhibit expression of more erythroid-restricted genes such as *Gata1* or *Epor*. Co-expression of *Gata1* and *Epor* was rarely seen and occurred at a frequency indicating independent activation. Expression of 'later' erythroid genes such as *Klf1* and *Epb4.2* was not observed.

Similar results were obtained in other prospectively isolated fractions of EML cells (Sca1hi and CD34$^+$ Itgae$^-$ encompassing the entire Sca1 axis) that exhibited an equivalent frequency of functionally definable culture-reconstituting cells (Fig. 4b). In all of these populations used for single-cell expression analysis, culture-reconstituting frequencies as judged by the ability of individual plated cells to generate clonal cultures are of the order of 40%. However, single-cell cloning is a stringent assay that probably underestimates cell potential, a notion supported by the 30–50% cloning efficiencies of murine pluripotent embryonic stem cells cultured in conditions that maintain them in a homogeneous ground-state configuration[24]. Attempts to optimize clonal assessment of EML culture-reconstituting capacity through co-culture on either MS-5 or normal bone-marrow-derived stroma, both of which are routinely used for long-term culture of primary haematopoietic stem cells[25] (HSCs), failed to increase cloning efficiency above 40% (data not shown). However, analysis of the contribution of individual Sca1lo CD34$^+$ cells to reconstitution of cultures through tracking of the cells' divisional history by carboxyfluorescein succinimidyl ester (CFSE) labelling indicates that most of these cells are capable of culture reconstitution (Fig. 4c).

We also further investigated the potential functional heterogeneity within Sca1lo CD34$^+$ cells from a molecular perspective by extending our single-cell gene expression analysis to encompass

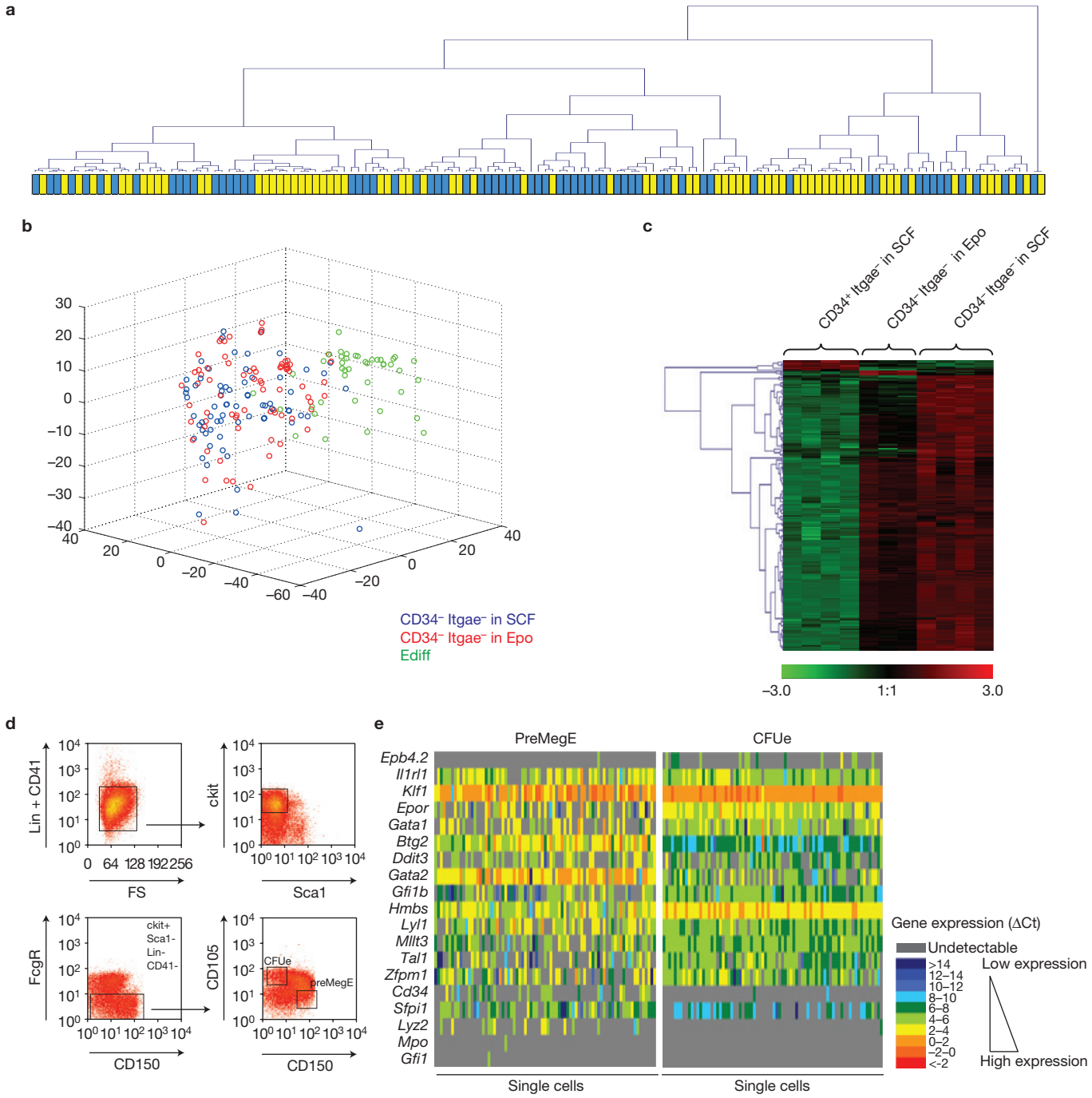

Figure 3 Molecular heterogeneity of early committed cells is observed in the presence of physiological differentation cues. (**a**) Cluster dendrogram of expression profiles of single CD34$^-$ Itgae$^-$ (kit$^+$) cells isolated in the absence (blue, $n = 68$) and in the presence of Epo (yellow, $n = 73$), two replicates per cell fraction; complete hierarchical clustering of individual cells based on Pearson correlation; see Supplementary Fig. S2c for the corresponding heat map. Similarity between expression profiles of individual cells within the same compartment measured by R^2 values (0–1): no Epo = 0.43, Epo = 0.45, both populations equally heterogeneous. R^2 Ediff–0.60 indicates more extensive homogeneity of transcriptional profiles. (**b**) Multidimensional scaling of the gene expression profiles of individual CD34$^-$ Itgae$^-$ cells isolated in the presence and in the absence of Epo, and of Ediff cells. CD34$^-$ Itgae$^-$ cells obtained under self-renewal conditions (blue) and in the presence of Epo (red) are indistinguishable and equally distant from differentiated kit$^-$ cells (green). (**c**) Heat-map representation of the EML erythroid signature defined in Fig. 2b among differentially expressed genes from any of the comparisons between CD34$^+$ Itgae$^-$, CD34$^-$ Itgae$^-$ (both in self-renewing culture conditions) and CD34$^-$ Itgae$^-$ cells after two days in Epo; the corresponding gene list is provided in Supplementary File S3. The global transcriptional profiling was carried out in quadruplicate or triplicate; differentially expressed genes are defined at a B value > 2. (**d**) Representative sorting plots of bone-marrow-derived early and late erythroid progenitor cells. preMegE sorted as lineage (lin)-CD41$^-$ Sca1$^-$ kit$^+$ FcγR$^-$ CD150$^+$ CD105$^-$; CFUe are lin-CD41$^-$ Sca1$^-$ kit$^+$ FcγR$^-$ CD150$^-$ CD105$^+$. (**e**) Expression of lineage-relevant genes tested in single preMegE ($n = 79$) and CFUe cells ($n = 78$); the results are plotted as in Fig. 2c. The data are representative of two independent experiments; global ΔCt ranges: −1.38–15.85 (preMegE) and −1.42–15.07 (CFUe). R^2 are 0.43 (preMegE) and 0.69 (CFUe); in the other independent experiment, R^2 are 0.43 (preMegE) and 0.57 (CFUe).

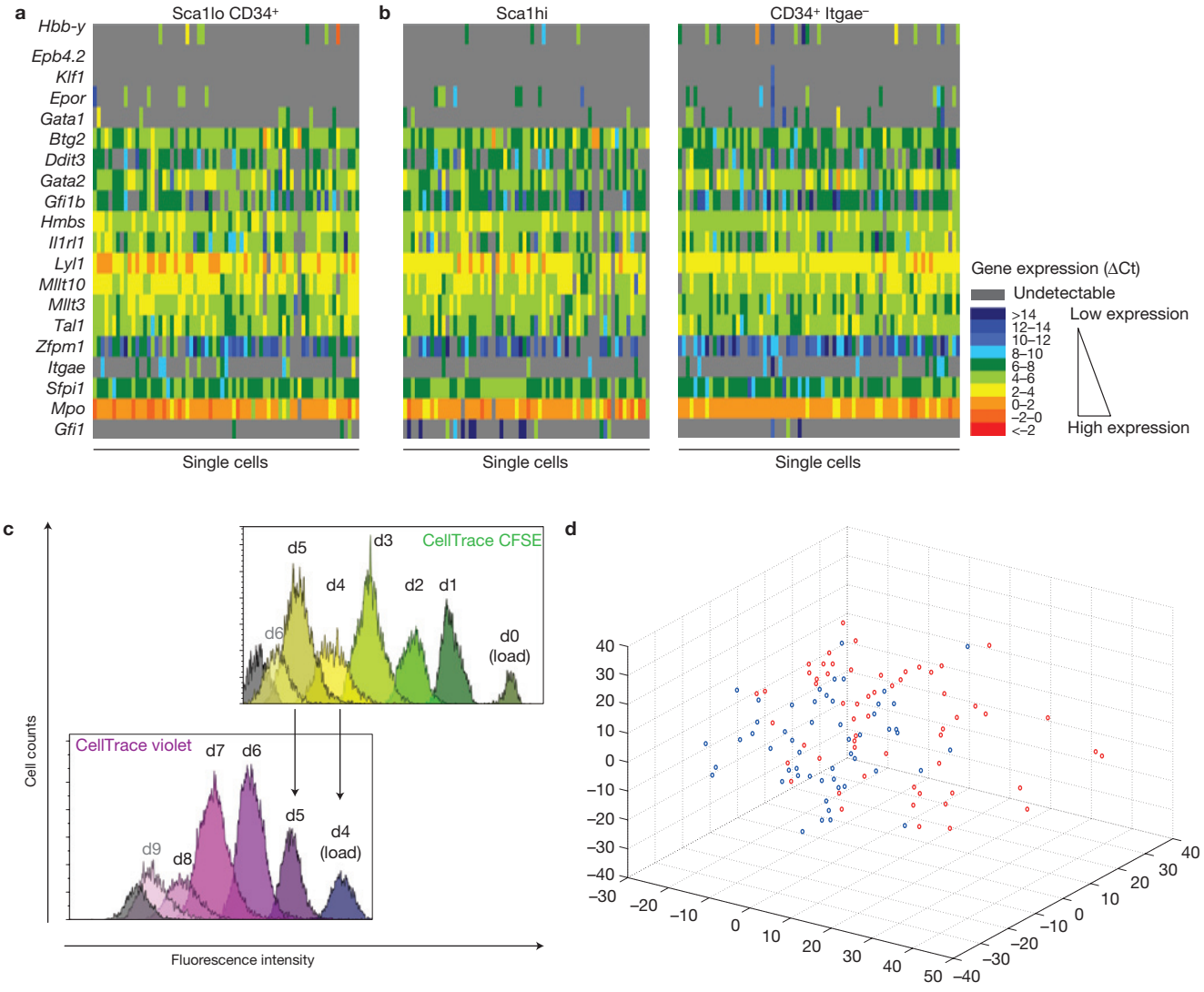

Figure 4 Self-renewing cells exhibit rare expression of individual erythroid regulators. (**a**) Expression of 20 lineage-relevant genes in individual Sca1lo CD34$^+$ cells ($n = 69$); representation as in Fig. 2c; global ΔCt range: −0.18–14.79; $R^2 = 0.67$. Data are from two independent samples with single-cell culture reconstitution capacities (percentage of clones >100 cells) of 44% and 42%. (**b**) Expression of the same genes in individual Sca1hi ($n = 64$) and CD34$^+$ Itgae$^-$ cells ($n = 73$). Representation as in Fig. 2c; global ΔCt ranges Sca1hi: −0.54–17.58; CD34$^+$ Itgae$^-$: −0.43–17.30; R^2 Sca1hi = 0.60, R^2 CD34$^+$ Itgae$^-$ = 0.70. The data for each compartment are from two independent samples with single-cell culture-reconstitution capacities of 53% and 38% (Sca1hi), and 44% (CD34$^+$ Itgae$^-$, only 1 sample functionally tested). (**c**) Divisional history of Sca1lo CD34$^+$ EML cells in bulk self-renewing cultures (representative experiment). CFSE-labelled Sca1lo CD34$^+$ cells were cultured in bulk over a period of 9 days; at day 4, all cells in culture were also labelled with CellTrace Violet. Fluorescence profiles of one or both dyes were determined by flow cytometry analysis on a daily basis (day = d1–9). Reconstitution of bulk Sca1 and CD34 profiles checked at regular intervals confirmed culture reconstitution (data not shown). (**d**) Multidimensional scaling of the expression profiles of individual Sca1lo CD34$^+$ (blue) and Sca1hi (red) cells failed to reveal subpopulations within culture-reconstituting compartments. The gene panel tested used the previous 20 lineage-affiliated genes and 24 other genes (see Supplementary Table S2) selected from an EML self-renewal signature defined as genes upregulated in CD34$^+$ Itgae$^-$ versus CD34$^-$ Itgae$^-$ cells (both cell types in self-renewing culture conditions), using a fold-change cutoff of 1.5 at a B value > 2 (full gene signature in Supplementary File S4). The genes selected code for transcription factors, chromatin remodellers, growth factor receptors and elements of signalling cascades and reflect the full range of gene expression enrichments. Each compartment was analysed in two replicate experiments for a total of 54 Sca1lo CD34$^+$ and 59 Sca1hi cells. Global ΔCt ranges: −0.79–16.76 (Sca1lo CD34$^+$); −1.39–15.76 (Sca1hi); expression patterns of the 20 lineage-affiliated genes are similar to experiments in **a** and **b**.

genes selected from an EML self-renewal signature. We reasoned that this should identify cells with different functional culture-reconstituting potential. However, this analysis failed to reveal any subpopulations within the Sca1lo CD34$^+$ compartment (Fig. 4d). Taken together, these data raise the possibility that the sporadic expression of particular lineage regulators may indeed arise within culture-reconstituting cells. This notwithstanding it remains difficult to unequivocally map gene expression patterns resident in a minor fraction of cells to a particular class of functional cells. One possible view is that these cells reside within the culture-reconstituting compartment. If so, the sporadic expression of lineage regulators may increase the probability of transit into the committed state.

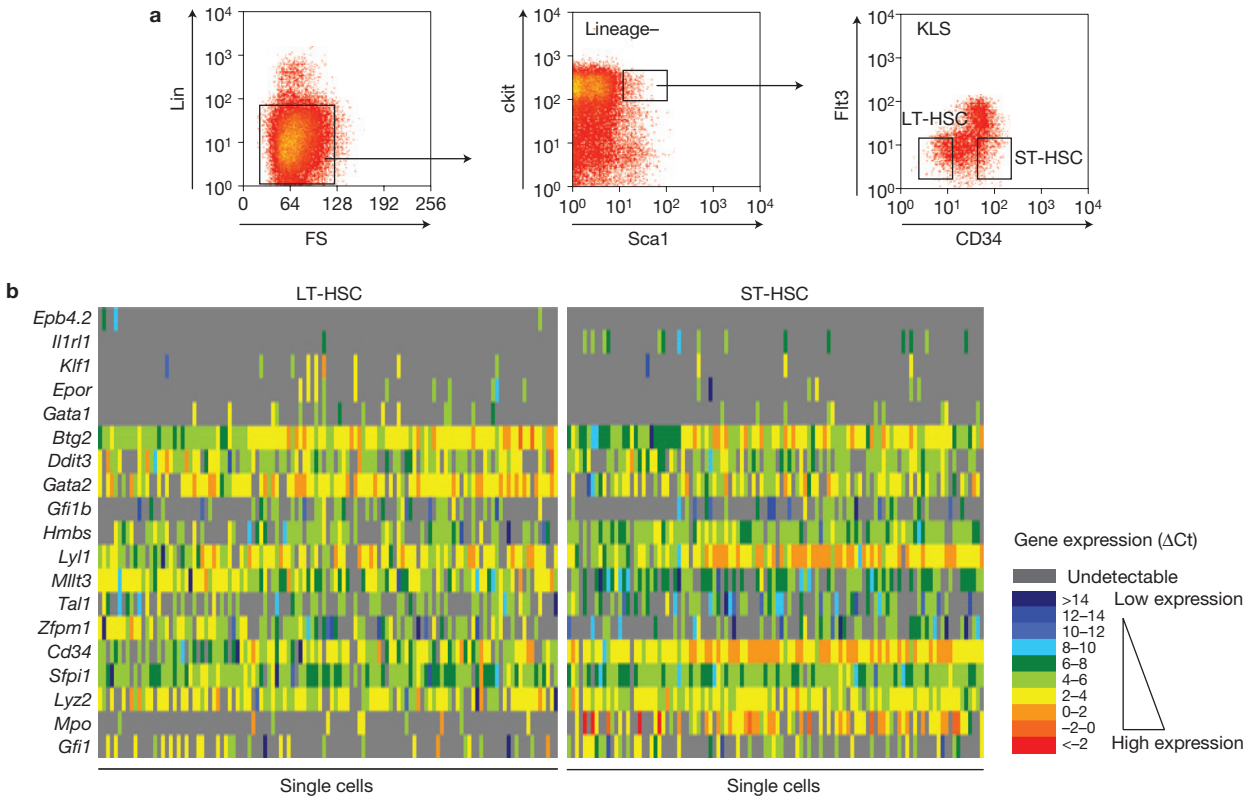

Figure 5 Mouse bone marrow stem cells exhibit infrequent and independent expression of erythroid genes. Representative sorting plots of bone-marrow-derived multipotent stem cells. Long-term repopulating HSCs (LT-HSCs) sorted as lin-kit$^+$ Sca1$^+$(KLS) CD34$^-$ Flt3$^-$; short-term repopulating HSCs (ST-HSCs) are KLS CD34$^+$ Flt3$^-$. (**b**) Expression of lineage-relevant genes tested in single LT-HSCs ($n = 117$) and ST-HSCs cells ($n = 106$). The data reflect two independent experiments. The results are plotted as in Fig. 2c; global ΔCt ranges are LT-HSCs: −0.40–16.88; ST-HSCs: −2.98–16.35.

An alternative view is that these cells are no longer functional culture-reconstituting cells. Given their gene expression patterns, such a view is still consistent with the notion that commitment occurs through independent activation of individual regulators in the absence of a coordinated lineage program. Furthermore, the heterogeneity in gene expression seen between different cells raises the possibility that there may be different potential entry points into lineage specification.

The notions that commitment occurs in the absence of a coordinated lineage program, and that lineage specification may potentially use alternative molecular routes are supported by the analysis of prospectively isolated bone-marrow-residing HSCs (ref. 26; Fig. 5a,b). Importantly, these are rare true self-renewing cells capable of multi-lineage reconstitution in an *in vivo* transplantation setting. The fact that, similarly to our observations in culture-reconstituting cells isolated from cell lines, HSCs do exhibit low-frequency expression of individual lineage-affiliated genes supports the view that sporadic expression of lineage regulators can indeed occur in the self-renewal state.

In this work we have sought to establish a rubric for lineage commitment through investigation of the molecular programs of individual cultured and primary multipotent self-renewing cells and early erythroid-committed cells. Single-cell-level analysis revealed that previous observations of reversible molecular lineage bias[7] in the multipotent EML model system also used in this study could be accounted for by coexistence of separable self-renewing and committed populations[16], with no evidence of transcriptome fluctuation in individual self-renewing cells. Furthermore, we could find no evidence of global sampling of lineage programs by individual cells before lineage commitment. Instead, our data are consistent with an alternative model of discrete and uncoordinated molecular transitions between the self-renewal and the committed states, mediated by stochastic independent modulation of the expression of key regulatory factors. Such a scenario is compatible with multiple entry routes into lineage commitment. The program of individual committed cells is then elaborated quantitatively and qualitatively downstream of these commitment events and resolves an incipient heterogeneous state to coordinated terminal differentiation. An analogous model of discordant entries into differentiation with eventual program coalescence in maturing cells was proposed at the population level in a model granulocytic cell line induced to terminally differentiate in response to dimethylsulphoxide or all-*trans* retinoic acid[27]. In this model, the HL-60 cell line showed initial genome-wide divergence in its transcriptional response to the two differentiative cues, which later resolved into a common and stable transcriptome pattern. Approached from the point of view of single cells, our results indicate that individual erythroid-committed cells may incipiently organize distinct and uncorrelated transcriptional programs around or downstream of key regulators, and that these programs eventually converge into a common transcriptional route to a complete and molecularly coherent erythroid-differentiated state. □

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturecellbiology

Note: Supplementary Information is available on the Nature Cell Biology website

ACKNOWLEDGEMENTS

The authors would like to thank R. Gupta and G. May for conceptual discussions; C. Waugh, K. Clark, A. Pizzey and T. Adejumo for cell sorting; S. McGowan for microarray data accessibility; and J. Wray for critical reading of the manuscript. J.T. is a student of the PhD Program in Computational Biology at Instituto Gulbenkian de Ciencia, Oeiras, Portugal, and was financially supported by Fundacao para a Ciencia e Tecnologia (SFRH/BD/33208/2007). C. Peterson is supported by the Swedish Foundation for Strategic Research (Senior Individual Grant). This work was financially supported by the Medical Research Council of the United Kingdom, Leukaemia and Lymphoma Research, EuroSyStem and STEMEXPAND.

AUTHOR CONTRIBUTIONS

C. Pina initiated and led the study, conducted all single-cell RT–qPCR experiments, carried out clonal reconstitution assays and divisional tracking experiments, participated in population-based analyses of EML subcompartments, analysed and interpreted experimental data, participated in figure production and wrote the paper. C.F. processed mouse bone marrow samples, carried out western blotting, participated in characterization of EML subcompartments, including clonal reconstitution assays, analysed experimental data and contributed to its interpretation and produced the figures. A.J.T. carried out immunostaining, did cell sorting, participated in population-based analyses of EML subcompartments, analysed experimental data and contributed to its interpretation and contributed to figure production. J.B. carried out non-quantitative single-cell RT–PCR and processed microarray samples. S.S. analysed microarray data, contributed to analysis of single-cell RT–qPCR data and participated in figure production. J.T. and C. Peterson contributed to analysis of single-cell RT–qPCR data and contributed to data interpretation. T.E. supervised all aspects of the study, and wrote the paper. C. Pina, C.F. and A.J.T. contributed equally to this work.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturecellbiology
Reprints and permissions information is available online at http://www.nature.com/reprints

1. Enver, T., Pera, M., Peterson, C. & Andrews, P. W. Stem cell states, fates, and the rules of attraction. *Cell Stem. Cell* **4,** 387–397 (2009).
2. Canham, M. A., Sharov, A. A., Ko, M. S. H. & Brickman, J. M. Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. *PLoS Biol.* **8,** e1000379 (2010).
3. Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450,** 1230–1234 (2007).
4. Hayashi, K., Lopes, S. M., Tang, F. & Surani, M. A. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem. Cell* **3,** 391–401 (2008).
5. Hough, S. R., Laslett, A. L., Grimmond, S. B., Kolle, G. & Pera, M. F. A continuum of cell states spans pluripotency and lineage commitment in human embryonic stem cells. *PLoS One* **4,** e7708 (2009).
6. Kalmar, T. *et al.* Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol.* **7,** e1000149 (2009).
7. Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E. & Huang, S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453,** 544–547 (2008).
8. Bussmann, L. H. *et al.* A robust and highly efficient immune cell reprogramming system. *Cell Stem. Cell* **5,** 554–566 (2009).
9. Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51,** 987–1000 (1987).
10. Feng, R. *et al.* PU.1 and C/EBPα/β convert fibroblasts into macrophage-like cells. *Proc. Natl Acad. Sci. USA* **105,** 6057–6062 (2008).
11. Graf, T. & Enver, T. Forcing cells to change lineages. *Nature* **462,** 587–594 (2009).
12. 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).
13. Iwasaki, H. *et al.* GATA-1 converts lymphoid and myelomonocytic progenitors into the megakaryocyte/erythrocyte lineages. *Immunity* **19,** 451–462 (2003).
14. Weintraub, H. *et al.* Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. *Proc. Natl Acad. Sci. USA* **86,** 5434–5438 (1989).
15. Tsai, S., Bartelmez, S., Sitnicka, E. & Collins, S. Lymphohematopoietic progenitors immortalized by a retroviral vector harboring a dominant-negative retinoic acid receptor can recapitulate lymphoid, myeloid, and erythroid development. *Genes Dev.* **8,** 2831–2841 (1994).
16. Ye, Z. J., Kluger, Y., Lian, Z. & Weissman, S. M. Two types of precursor cells in a multipotential hematopoietic cell line. *Proc. Natl Acad. Sci. USA* **102,** 18461–18466 (2005).
17. Kim, S. I. & Bresnick, E. H. Transcriptional control of erythropoiesis: emerging mechanisms and principles. *Oncogene* **26,** 6777–6794 (2007).
18. Capron, C. *et al.* LYL-1 deficiency induces a stress erythropoiesis. *Exp. Hematol.* **39,** 629–642 (2011).
19. Pronk, C. J. *et al.* Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. *Cell Stem. Cell* **1,** 428–442 (2007).
20. Pina, C., May, G., Soneji, S., Hong, D. & Enver, T. MLLT3 regulates early human erythroid and megakaryocytic cell fate. *Cell Stem. Cell* **2,** 264–273 (2008).
21. Rodrigues, N. P. *et al.* Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. *Blood* **106,** 477–484 (2005).
22. Tipping, A. J. *et al.* High GATA-2 expression inhibits human hematopoietic stem and progenitor cell function by effects on cell cycle. *Blood* **113,** 2661–2672 (2009).
23. Porcher, C. *et al.* The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. *Cell* **86,** 47–57 (1996).
24. Wray, J., Kalkan, T. & Smith, A. G. The ground state of pluripotency. *Biochem. Soc. Trans.* **38,** 1027–1032 (2010).
25. Coulombel, L. Identification of hematopoietic stem/progenitor cells: strength and drawbacks of functional assays. *Oncogene* **23,** 7210–7222 (2004).
26. 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 (2005).
27. Huang, S., Eichler, G., Bar-Yam, Y. & Ingber, D. E. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.* **94,** 128701 (2005).

METHODS

EML cell culture. EML cells were maintained in Iscove's modified Dulbecco's medium (IMDM, Invitrogen) with 5% horse serum (PAA), 10% SCF-conditioned medium (SCF-CM) from BHK-MKL cells, 2 mM L-glutamine and 1% penicillin/streptomycin (self-renewal medium). Cell densities were kept between 2×10^4 and 2×10^5 cell ml^{-1}. For clonal cultures, single cells were double-sorted into the wells of a 96-well plate containing 100 µl of self-renewal medium; the wells were scored every 2–3 days and live and dead cells counted. For erythroid differentiation, a previously published protocol[28] was modified: briefly, 2×10^4 cells ml^{-1} were seeded in self-renewal medium supplemented with 10^U ml^{-1} recombinant human erythropoietin (rhEpo, Ortho Biotech or Amgen). At day 2, cells were washed and seeded at the initial density in IMDM with 5% horse serum, 2.5% SCF-CM, 2 mM L-glutamine, 1% penicillin/streptomycin and 10 U ml^{-1} rhEpo. At day 5, 10 U ml^{-1} rhEpo was added to the culture. For differentiation to the neutrophil lineage, 2×10^4 cells ml^{-1} were seeded in self-renewal medium with 10 ng ml^{-1} mIL3 (Peprotech) and 10 µM retinoic acid (Sigma Aldrich). At day 2, cells were washed and seeded at the initial density in IMDM 5% horse serum, 2.5% SCF-CM, 2 mM L-glutamine, 1% penicillin/streptomycin, 10 ng ml^{-1} mGM-CSF (Peprotech) and 10 µM retinoic acid. At day 5, cultures were supplemented with 10 ng ml^{-1} mGM-CSF. Differentiating cell cultures were followed for up to 7 days.

Isolation of mouse bone marrow stem and progenitor cells. C57Bl/6 mice were kept in the BMS facility at the John Radcliffe Hospital site, Oxford, in accordance with Home Office regulations. Long bones were dissected from humanely killed 7–9-week-old mice and bone marrow was isolated following standard procedures. In each experiment, 2/3 of the pooled cells were used for sorting of HSCs and 1/3 for isolation of erythro-myeloid progenitors, following published protocols with minor modifications[19,26]. After red blood cell lysis, mature cells were depleted in a column-based step using a cocktail of lineage antibodies. For sorting of long-term repopulating HSCs and short-term repopulating HSCs, this included B220, CD3, CD4, CD8, Gr1, Mac1 and Ter119; residual lineage staining was detected with a goat anti-rat antibody, and cells were further stained for surface markers kit, Sca1, CD34 and Flt3. For sorting of progenitors, Sca1 and CD41 were included in the lineage cocktail; further staining included kit, FcγR, CD105 and CD150. Detailed antibody information is included in Supplementary Table S1.

Flow cytometry, cell sorting, antibody staining and divisional tracking. The antibodies used for cell surface staining are described in Supplementary Table S1. Briefly, antibody staining was carried out on ice in either medium (EML) or PBS/fetal calf serum (bone marrow); cells were washed and resuspended in staining medium/buffer, and used for downstream flow cytometry. For divisional tracking experiments, cells were stained with cell surface antibodies and then loaded with 5 µM of CellTrace CFSE (Invitrogen) following the manufacturer's protocols. Cells were sorted on the basis of surface phenotype and the main CFSE fluorescence intensity range to exclude poorly loaded cells. Cells were analysed after sorting to define the CFSE loading peak and then on a daily basis, using the same flow cytometer settings, to follow their divisional history. At day 4, all cells in culture were loaded with 5 µM of CellTrace violet (Invitrogen) following the manufacturer's protocols without further sorting; definition of the loading peak and daily follow-up analysis were carried out as above. CFSE and CellTrace violet are lipophylic membrane dyes that, on cell division, are split approximately equally between daughter cells, resulting in halving of the fluorescence intensity of the label. After 5–6 divisions the fluorescence intensity overlaps with that of unlabelled cells (Fig. 4c, grey area), which prevents long-term follow-up with a single dye. We have overcome this limitation by using a second label; the occurrence of similar changes in the profiles of both dyes at day 5 highlights the validity of the approach (Fig. 4c). Divisional history was obtained for a total of 9 days; confirmation of Sca1 and CD34 profile reconstitution was obtained through analyses at days 0, 3, 6 and 8. Cells were sorted on FACSAria II (Benton Dickinson) or MoFlo (Beckman Coulter) instruments; analysis was carried out on CyAn ADP or Gallios instruments (Beckman Coulter). Appropriate unstained and single-colour controls were used for gate definition and compensation set-up. Data were acquired using FACS Diva (BD), Summit (BC) or Gallios (BC) software, and analysed in FlowJo, Summit or Kaluza.

Immunofluorescence microscopy. Cells (5×10^4–1×10^5) in PBS were deposited onto slides within poly-L-lysine-coated spots (Sigma) and allowed to adhere; slides were then air-dried and stored at −20 °C in the presence of desiccant. Staining was carried out at room temperature; all reagents and washes between steps were in PBS. Cells were immediately fixed in 4% paraformaldehyde (Electron Microscopy Sciences) and permeabilized in 1% Triton X-100 (Sigma); slides were then blocked in 2% fetal calf serum, 2% bovine serum albumin (BSA; Sigma) followed by staining with a 1:100 dilution of anti-GATA1 N6 antibody (Santa Cruz sc-265) in blocking solution. Secondary antibody staining was carried out with a 1:200 dilution of minimally crossreactive Affinipure Cy3-conjugated Donkey anti-rat antibody (Jackson Immunoresearch 712-165-153) in blocking solution. Cells were re-fixed in 2% paraformaldehyde before staining nuclei in 1:5,000 ToPro3 (Invitrogen). Slides were mounted in Vectashield and 22 mm coverslips (VWR) and imaged on a confocal microscope (Olympus BX51 with BioRad Radiance 2000 laser scanning system).

Immunoblotting. Total cell extracts were obtained by cell pellet lysis in RIPA buffer. Cleared lysates were run in Bis-Tris 6–12% NuPAGE gels in 1× MES running buffer (Invitrogen). Lanes were loaded with approximate cell equivalents. Proteins were transferred to pre-equilibrated Immobilon-P membranes (Millipore) using a wet transfer system. Membranes were blocked in TBS containing 0.05% Tween 20 and 5% fat-free milk, and stained with monoclonal anti-mouse GATA1 antibody (N6; sc-265, Santa Cruz Biotechnologies; 1:1,000 dilution), followed by HRP-conjugated goat anti-rat antibody (Pierce; 1:2,000 dilution). Antibody binding was detected using an electrochemiluminescence system (Amersham).

Quantitative real-time PCR. Equal amounts of total RNA extracted using TRIzol reagent (Invitrogen) were reverse-transcribed using SuperScriptII (SSII; Invitrogen) as per the manufacturer's instructions and complementary DNA samples were analysed by quantitative PCR (qPCR) using inventoried TaqMan gene expression assays (Applied Biosystems; see Supplementary Table S2). Standard curves were run for each gene to determine PCR efficiencies; relative gene expression was calculated using the Pfaffl method[29]. Analyses were carried out in duplicate or triplicate.

Single-cell multiplex PCR with reverse transcription. Single EML cells were double-sorted into the wells of a 96-well plate containing a detergent-based lysis buffer (0.4% IGEPAL in the presence of deoxynucleoside triphosphates, dithiothreitol and RNase inhibitor—RNase OUT, Invitrogen) and snap-frozen before storage at −80 °C. Control wells containing 0 or 10 cells were included. On thawing, plates were processed following published protocols[30] with minimal modifications, namely the use of SSII for gene-specific reverse transcription and inclusion of a reverse-transcriptase heat-inactivation step; this was followed by multiplex PCR (HotStar Taq, Qiagen; 35 cycles, Ta = 60 °C). Two per cent of the DNA obtained was amplified in individual gene-specific reactions using nested primers (35 cycles, Ta = 60 °C). Products were detected in ethidium-bromide-stained agarose/TAE gels; only cells with detectable Hprt1 expression were considered for further analysis. Primer sequences are in Supplementary Table S3.

Gene expression profiling and microarray analysis. Total RNA was isolated from triplicate or quadruplicate samples of EML cell fractions using TRIzol reagent, and RNA concentration and integrity were determined using an RNA 6000 Nano RNA kit (Agilent) on a BioAnalyzer 2100 (Agilent). Total RNA (150 ng) and Agilent One-Color RNA Spike-In were reverse-transcribed and linearly amplified in the presence of Cy3-labelled CTP using a Low RNA Input One-Color kit (Agilent) following Agilent protocols, to generate Cy3-labelled complementary RNA for hybridization to high-density microarrays. Quality and yield/specific activity of cRNA were determined using an RNA 6000 Nano kit and spectrophotometer (NanoDrop ND-1000). All samples generated comparable quality cRNA profiles, with specific activities in the range of 10.3 to 12.2 pmol Cy3 µ^{-1}g^{-1} cRNA. Each Cy3-labelled cRNA sample (1.6 µg as per NanoDrop reading) was hybridized to an individual 44 K sub-array of a 4 × 44 K high-density microarray slide (Whole Mouse Gene Expression Microarrays, Agilent), washed, scanned and feature-extracted as per Agilent protocols. Data were normalized using cyclic loess, and differentially expressed genes were identified using LIMMA (ref. 31) for all pairwise comparisons. All complete hierarchical clustering and K-means clustering analyses were carried out in Genesis on the basis of Pearson correlation and using Z-score values (mean centred expression values divided by the s.d.).

Quantitative single-cell PCR with reverse transcription. For single-cell qPCR with reverse transcription (RT–qPCR) analysis, single cells were double-sorted into 10 µl of CellsDirect reaction mix containing 1 µl of SSIII/PlatinumTaq mix (CellsDirect One-Step RT–qPCR kit, no ROX, Invitrogen), RNase OUT and up to 48 multiplexed inventoried TaqMan assays, final dilution 0.05× (Supplementary Table S2). Zero-, 5- and 50-cell controls were included. RT–PCR pre-amplification cycling conditions were: 50 °C, 15 min; 95 °C, 2 min; 22 or 25 × (95 °C, 15 s; 60 °C, 4 min) for EML and primary bone marrow cells, respectively. The DNA obtained was quantified in standard qPCR conditions using the same inventoried TaqMan assays in single or duplicate reactions, depending on the number of genes tested. Nine per cent of the pre-amplification products mixed with TaqMan Universal PCR Master Mix (Applied Biosystems), and the TaqMan assays were loaded separately onto the wells of 48.48 Gene expression Dynamic Arrays (Fluidigm) in the presence of the appropriate loading reagents. The arrays were read in a Biomark genetic analysis system (Fluidigm) and the data were exported into Microsoft Excel for downstream analysis. The amplification curves were quality-controlled and the data filtered according to no-reverse-transcription control reactions and to exclude Ct > 30. ΔCt values were calculated in reference to

DOI: 10.1038/ncb2442

Atp5a1; readings from duplicate assays were averaged; only cells with expression of all tested housekeeping controls (*Atp5a1*, *Hprt1* and *Ubc* ± *B2m*) were considered for downstream analysis. Data used for multidimensional scaling did not include housekeeping controls or *Ly6a* (Sca1). The R^2 value (or percentage of variance explained) for each population was calculated using genes as factors in a linear model.

Accession numbers. GEO microarray data accession number: GSE34605.

28. Orford, K. *et al*. Differential H3K4 methylation identifies developmentally poised hematopoietic genes. *Dev. Cell* **14,** 798–809 (2008).
29. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic. Acids Res.* **29,** e45 (2001).
30. Hu, M. *et al*. Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev.* **11,** 774–785 (1997).
31. Smyth, G. K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet Mol. Biol.* **3,** 1–25 (2004).

DOI: 10.1038/ncb2442

nature cell biology

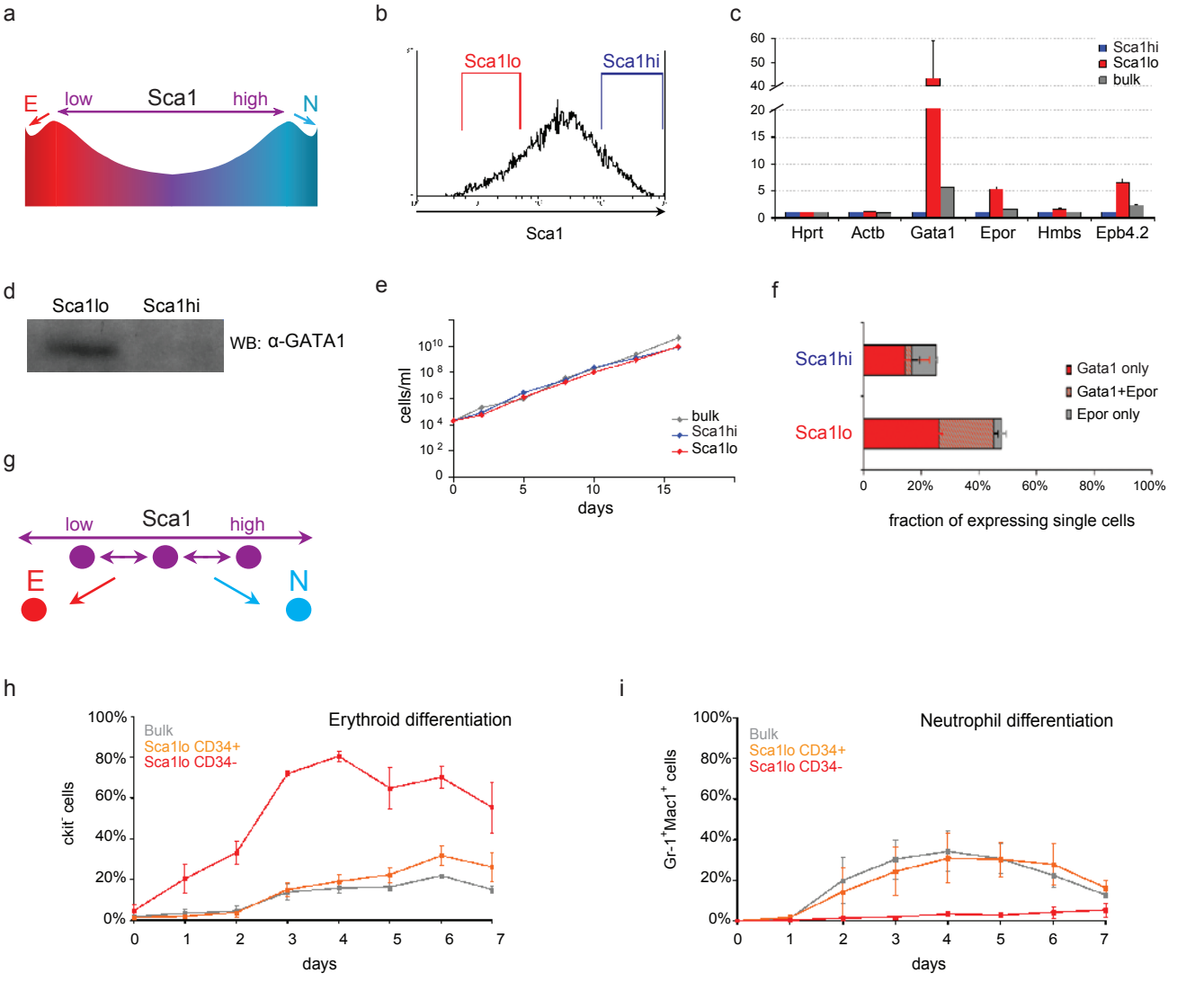

Figure S1 Investigation of self-renewal potential and molecular lineage bias in EML Sca1lo cells (relates to Figure 1 in main text). a. Lineage bias and self-renewal potential of EML cells as per the attractor metaphor. The study by Huang and collaborators has suggested that self-renewing EML cells can experience transcriptome-wide activation of nearly complete lineage programmes in a reversible manner, while retaining culture-reconstitution potential. Erythroid (E) and neutrophil (N)-like states correlate with low and high expression of the surface marker Sca1, respectively. b. Representative flow cytometry histogram of Sca1 expression in EML cells; 15% lowest and highest-staining cells sorted as Sca1lo and Sca1hi, respectively. c. Quantitative RT-PCR (RqPCR) analysis of selected erythroid-affiliated genes in Sca1lo and Sca1hi EML populations. Data plotted are fold-change in gene expression relative to housekeeping control Hprt1 and summarise 3 independent paired experiments as mean + standard error of the mean (s.e.m.); an additional control gene Actb is represented. d. Western blot analysis of Gata1 protein levels in Sca1hi and Sca1lo EML fractions; identical cell-equivalent amounts of total cell extracts were loaded. Data are representative of 2 experiments. e. Growth curves of Sca1lo, Sca1hi and total (bulk) EML cell-initiated cultures; data is representative of two independent experiments. f. Single-cell multiplex RT-PCR (non-quantitative) analysis of Gata1 and Epor expression in Sca1lo (n=153) and Sca1hi (n=160) EML cells with detectable expression of Hprt1; the results are mean + standard deviation of 2 independent paired experiments. g. Diagram depicting alternative model underlying co-existence of self-renewal capacity and molecular lineage bias at the Sca1 distribution extremes: analysis at the population level does not distinguish between functionally-identical cells with distinct molecular signatures as in (a), from heterogeneous mixtures of transcriptionally stable unbiased self-renewing cells encompassing the entire range of Sca1 expression, and distinct lineage-affiliated cells with no culture-reconstitution potential at the extremes of Sca1 distribution. h, i. Timecourse of erythroid (h) and neutrophil (i) differentiation of Sca1lo CD34+ and Sca1lo CD34- cells; bulk EML cells used as control. Proportion of differentiated cells assessed daily by flow cytometry: erythroid differentiation read by loss of c-kit (kit-), neutrophil differentiation read by Gr1/Mac1 double-positivity. Plots represent data from at least 2 independent experiments; results are mean ± s.e.m..

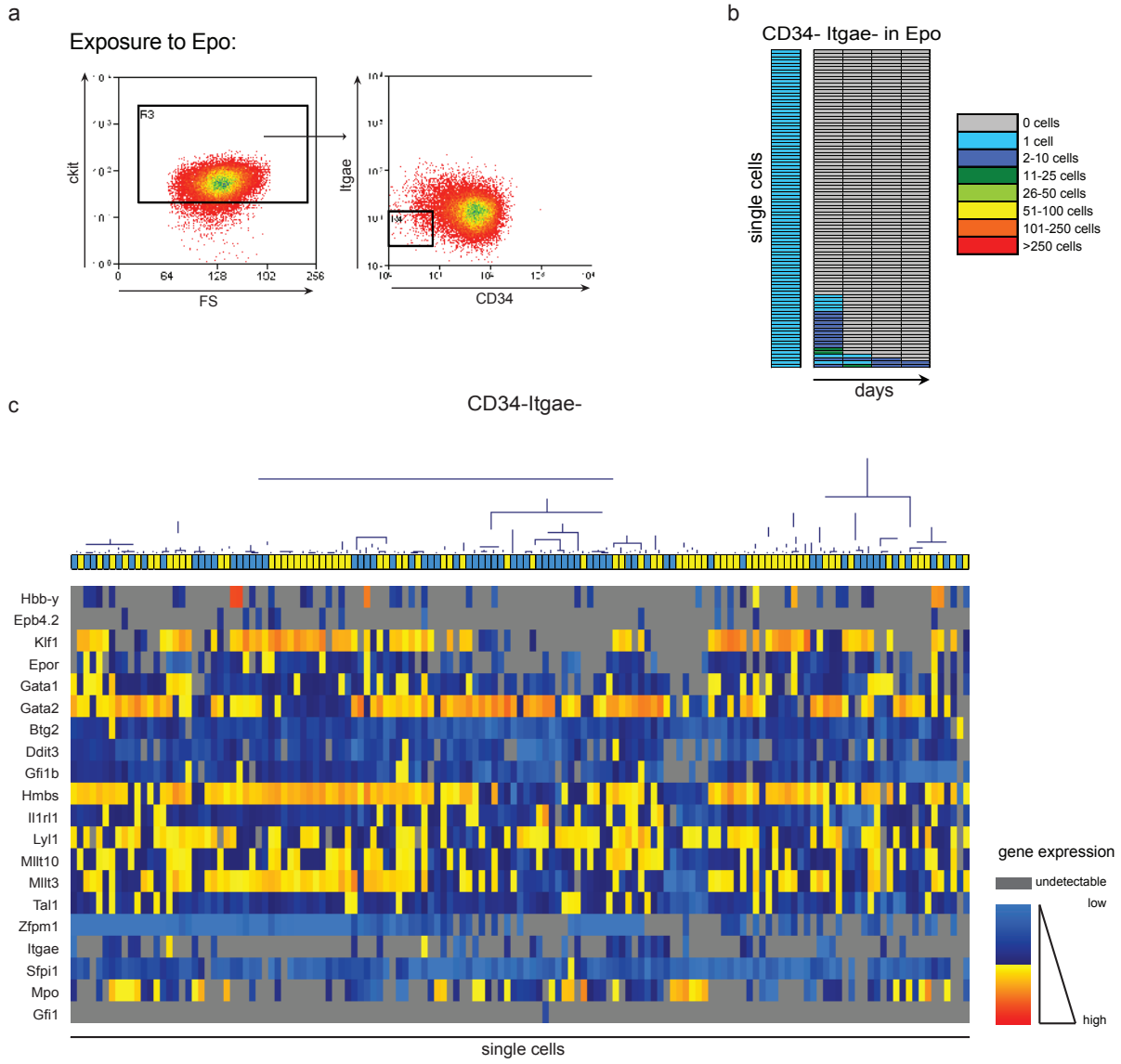

Figure S2 Characterisation of erythroid-committed CD34-Itgae- EML cells isolated in the presence and in the absence of erythropoieitin (Epo) (relates to Figure 3 in main text). a. Flow cytometry plots representing CD34-Itgae-kit+ cells first isolatable after 2 days of exposure of CD34+Itgae- EML cells to Epo. Unlike self-renewing SCF-dependent EML cultures in which virtually all the cells are kit+, differentiating cultures in Epo have a growing fraction of kit- cells. We have only sorted the kit+ cells in the CD34-Itgae- fraction to allow direct comparison with CD34-Itgae- cells obtained in self-renewing culture conditions. b. Clonal analysis of culture-reconstituting potential of CD34-Itgae-kit+ cells isolated from Epo-containing cultures; plotting scheme as in Fig. 1c. The analysis shows that these cells are devoid of culture-reconstituting capacity, thus highlighting the correspondence between surface phenotype and functional potential under different culture conditions. CD34+Itgae-kit+ cells isolated from the same Epo-containing cultures had retained self-renewal potential (data not shown). c. Expression profiles of CD34-Itgae- cells isolated from EML cultures in the presence and in the absence of Epo. The heatmap represents the data in the cluster dendrogram in Fig. 3a, which is reproduced here. Data are Z-score normalised expression values for the complete panel of 20 lineage-relevant genes tested in EML cells. In the cluster dendrogram, blue boxes are CD34-Itgae- cells obtained under self-renewing culture conditions (no Epo), yellow boxes are CD34-Itgae- cells obtained from Epo-containing cultures.

Supplementary Files

Supplementary File 1 Expression profiles of genes up-regulated in Sca1lo (vs. Sca1hi) cells as per Chang et al microarray data in Sca1hi, Sca1lo CD34+ and Sca1lo CD34- EML compartments (Excel datasheet)

Supplementary File 2 Expression profiles of differentially expressed genes up-regulated in E-diff EML cells (Excel datasheet)

Supplementary File 3 Expression profiles of differential E-diff signature genes in CD34+ Itgae-, CD34- Itgae- (no Epo) and CD34- Itgae- (+Epo) EML compartments (Excel datasheet)

Supplementary File 4 Expression profiles of genes up-regulated in the CD34+ Itgae- (vs. CD34- Itgae-) EML compartment (Excel datasheet)

Supplementary Table 1. Antibodies used for flow cytometry and cell sorting.

Antibody	Fluorochrome	Clone	Supplier	Catalogue number	Dilutions
CD3ε	purified	145-2C11	BioLegend	100301	1:100
CD4	purified	G41.5	eBioscience	16-0041	1:100
CD8a	purified	53-6.7	eBioscience	16-0081	1:100
CD11b / Mac1	purified biotinylated PE APC	M1/70	eBioscience	16-0112 13-0112 12-0112 17-0112	1:100 1:100 1:100 1:100
CD16/32 / FcγIII/IIR	FITC	93	eBioscience	11-0161	1:100
CD34	FITC Alexafluor 647	RAM34	eBioscience	11-0341 51-0341	1:20 1:100
CD41	purified	MWREG30	BD Pharmingen	553847	1:100
CD103 / Itgae	biotinylated	2E7	eBioscience	13-1031	1:50
CD105 / endoglin	PE	MJ7/18	eBioscience	12-1051	1:50
CD150 / SLAMF1	PECy7	TC15-12F12.2	BioLegend	115914	1:20
B220	purified	RA3-6B2	eBioscience	16-0452	1:100
Flt3	biotinylated	A2F10	eBioscience	13-1351	1:50
Ly-6A/E / Sca-1	PE PECy7 PB	E13-161.7 or D7	BD Pharmingen BD Pharmingen BioLegend	553336 558162 122520	1:50 1:50 1:50
Ly-6C / Gr1	purified PE PECy7	RB6-8C5	BD Pharmingen BD Pharmingen eBioscience	550291 553128 25-5931	1:100 1:100 1:100
Ter-119	purified APC	TER-119	BD Pharmingen	550565 557909	1:100 1:100
F(ab')$_2$ goat anti-rat IgG (H/L)	PECy5		Invitrogen	A10691	1:100
Streptavidin	PECy7 APCeFluor780 PB APC		eBioscience eBioscience Invitrogen eBioscience	25-4317 47-4317 S11222 17-4317	1:100 1:100 1:200 1:200

Note: Lineage cocktail used for sorting HSC from mouse BM was CD3ε, CD4, CD8a, B220, Gr1, Mac1, Ter119; CD41 and Sca1 were added when sorting progenitors. All lineage antibodies were purified and detected with PECy5-conjugated F(ab')$_2$ goat anti-rat IgG (H/L).

Supplementary Table 2. Taqman assays used in RqPCR analysis

Gene	Assay number	RqPCR	Fluidigm (EML)	Fluidigm (BM)
Actb	Mm00607939_s1	X		
Atp5a1	Mm00431960_m1		X	X
B2m	Mm00437762_m1		X	
Hprt1	Mm00446968_m1	X	X	X
Ubc	Mm01201237_m1		X	X
Arid5a	Mm00524454_m1		X	
Btg2	Mm00476162_m1		X	X
Cd34	Mm00519283_m1		X	X
Cebpa	Mm00514283_s1		X	
Cebpg	Mm01266786_m1		X	
Cxcr3	Mm00438259_m1		X	
Cxxc5	Mm00505000_m1		X	
Ddit3	Mm00492097_m1		X	X
Dlx1	Mm00438424_m1		X	
Dnmt3b	Mm01240113_m1		X	
Epb4.2	Mm00469107_m1	X	X	X
Epor	Mm00438760_m1	X	X	X
Erg	Mm01214246_m1		X	
Gata1	Mm00484678_m1	X	X	X
Gata2	Mm00492300_m1		X	X
Gfi1	Mm00515855_m1		X	X
Gfi1b	Mm00492319_m1		X	X
Hbb-y	Mm00433936_g1		X	
Hlf	Mm00723157_m1		X	
Hmbs	Mm01143545_m1	X	X	X
Hmga2	Mm04183367_g1		X	
Il1rl1	Mm00516117_m1		X	X
Irf1	Mm01288580_m1		X	
Irf6	Mm00516797_m1		X	
Itga1	Mm01306375_m1		X	
Itgae	Mm00434443_m1		X	

Itgax	Mm00498698_m1		**X**	
Jak3	Mm00439962_m1		**X**	
Klf1	Mm00516096_m1		**X**	**X**
Ly6a	Mm00726565_s1		**X**	
Lyl1	Mm00493219_m1		**X**	**X**
Lyz2	Mm01612741_m1			**X**
Meis1	Mm00487664_m1		**X**	
Mllt3	Mm00550927_m1		**X**	**X**
Mllt10	Mm00487708_m1		**X**	
Mpl	Mm00440310_m1		**X**	
Mpo	Mm00447886_m1		**X**	**X**
Sfpi1	Mm00488140_m1		**X**	**X**
Smad3	Mm01170760_m1		**X**	
Smad5	Mm03024001_g1		**X**	
Sox4	Mm00486317_s1		**X**	
Stat3	Mm01219775_m1		**X**	
Tal1	Mm00441665_m1		**X**	**X**
Wnt10a	Mm00437325_m1		**X**	
Zfp810	Mm00521242_m1		**X**	
Zfpm1	Mm00494336_m1		**X**	**X**

Note: Control gene expression assays are highlighted in grey.

Supplementary Table 3. Primers used in non-quantitative single-cell PCR.

Primer name	**Sequence**
mGATA1 outer forward	TCACCATCAGATTCCACAGG
mGATA1 nested forward	GCCAGAGGGTTTGGATGCAG
mGATA1 nested reverse	TGAGGCAGGGTAGAGTGCCG
mGATA1 outer reverse	CCAAGAACGTGTTGTTGCTC
mEpoR outer forward	CGCTACACCTTCGCTGTTCG
mEpoR nested forward	TCTGGAGTGCCTGGTCTGAG
mEpoR nested reverse	GCTCTCTGGGCTTGGGATGC
mEpoR outer reverse	CAAACTCGCTCTCTGGGCTT
mHPRT outer forward	GGGGGCTATAAGTTCTTTGC
mHPRT nested forward	GTTCTTTGCTGACCTGCTGG
mHPRT nested reverse	TGGGGCTGTACTGCTTAACC
mHPRT outer reverse	TCCAACACTTCGAGAGGTCC

Note: Annealing temperature is 60^{o}C for all primer pairs.