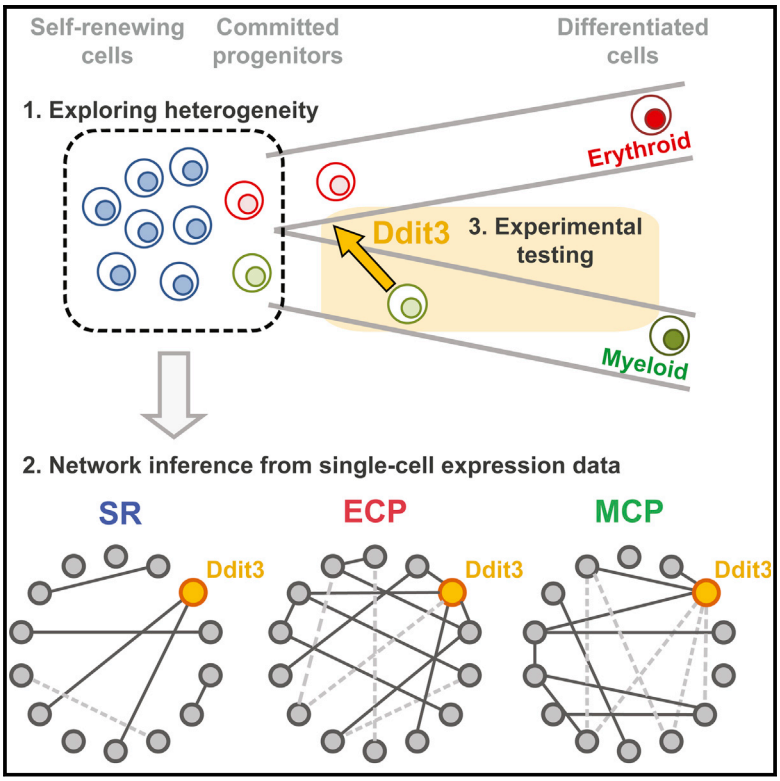
Cell Reports

Report

Single-Cell Network Analysis Identifies DDIT3 as a Nodal Lineage Regulator in Hematopoiesis

Graphical Abstract



Authors

Cristina Pina, José Teles, Cristina Fugazza, ..., Mattias Ohlsson, Carsten Peterson, Tariq Enver

Correspondence

t.enver@ucl.ac.uk

In Brief

Pina et al. develop a gene regulatory network inference method using single-cell gene expression data and identify *Ddit3* as a regulatory node in erythroid lineage programming. The authors explore this inference and show that Ddit3 can antagonize myeloid programming and enable erythroid signatures and forms a regulatory axis with Gata2.

Highlights

- We present a method for inferring gene regulatory networks (GRNs) from single cells
- Lineage cross-antagonism is a key property of GRNs of early lineage commitment
- *Ddit3* is a regulatory node in erythroid lineage programming
- A Ddit3-Gata2 regulatory axis antagonizes myeloid and enables erythroid programs

Accession Numbers

GSE68754
SRP045672

Pina et al., 2015, Cell Reports *11*, 1503–1510
June 16, 2015 ©2015 The Authors
http://dx.doi.org/10.1016/j.celrep.2015.05.016

CellPress

Cell Reports
Report

Single-Cell Network Analysis Identifies DDIT3 as a Nodal Lineage Regulator in Hematopoiesis

Cristina Pina,[1,3,4] José Teles,[1,2,3,5] Cristina Fugazza,[1,3,6] Gillian May,[1] Dapeng Wang,[1] Yanping Guo,[1] Shamit Soneji,[1,7] John Brown,[1] Patrik Edén,[2] Mattias Ohlsson,[2] Carsten Peterson,[2] and Tariq Enver[1,*]
[1]Stem Cell Laboratory, UCL Cancer Institute, University College London, London W1CE 6BT, UK
[2]Computational Biology and Biological Physics, Department of Astronomy and Theoretical Physics, Lund University, 223 62 Lund, Sweden
[3]Co-first author
[4]Present address: National Health Service Blood and Transplant, Department of Haematology, University of Cambridge, Cambridge CB2 0PT, UK
[5]Present address: Sainsbury Laboratory, University of Cambridge, Cambridge CB2 1LR, UK
[6]Present address: Dipartimento di Biotecnologie e Bioscienze, Università degli Studi di Milano-Bicocca, 20126 Milano, Italy
[7]Present address: Lund Stem Cell Center, Lund University Hospital, Biomedical Centre, 221 84 Lund, Sweden
*Correspondence: t.enver@ucl.ac.uk
http://dx.doi.org/10.1016/j.celrep.2015.05.016

SUMMARY

We explore cell heterogeneity during spontaneous and transcription-factor-driven commitment for network inference in hematopoiesis. Since individual genes display discrete OFF states or a distribution of ON levels, we compute and combine pairwise gene associations from binary and continuous components of gene expression in single cells. *Ddit3* emerges as a regulatory node with positive linkage to erythroid regulators and negative association with myeloid determinants. *Ddit3* loss impairs erythroid colony output from multipotent cells, while forcing *Ddit3* in granulo-monocytic progenitors (GMPs) enhances self-renewal and impedes differentiation. Network analysis of *Ddit3*-transduced GMPs reveals uncoupling of myeloid networks and strengthening of erythroid linkages. RNA sequencing suggests that *Ddit3* acts through development or stabilization of a precursor upstream of GMPs with inherent Meg-E potential. The enrichment of Gata2 target genes in Ddit3-dependent transcriptional responses suggests that *Ddit3* functions in an erythroid transcriptional network nucleated by *Gata2*.

INTRODUCTION

Development and differentiation are characterized by genetic circuitry or gene regulatory networks (GRNs) that have inherent forward momentum encoded by a number of regulatory motifs (Davidson, 2010). To self-renew and maintain differentiation potential, stem cells must structure their GRNs so as to arrest or buffer this forward trajectory. Networks at early multipotent stages may bear little relation to those of mature differentiated cells, making comparison between them difficult. Detailed time-series data are useful in this regard (Bruno et al., 2004; May et al., 2013), but there is likely substantial asynchrony between individual cells at any given time point. Cells also may undergo lineage commitment through different initial gene expression trajectories (Pina et al., 2012). Together, these factors can confound attempts to infer network architectures and gain molecular insights into commitment and subsequent lineage specification from averaged gene expression profiles.

Analysis of gene expression in single cells offers a different approach that, in principle, makes use of cellular heterogeneity as a source of variation for establishing gene-gene associations. Recent studies have used all expression data (Guo et al., 2013; Moignard et al., 2013, 2015) or inferred pairwise gene associations using only co-expressing cells (Ståhlberg et al., 2011). It has been suggested that levels of expression are better accounted for in cells that co-express both genes, and may be obscured by presence/absence effects when all cells are considered (Rusnakova et al., 2013). We have tried to address this constraint in our exploration of gene expression networks around the erythroid versus myelo-monocytic lineage choice of multipotent hematopoietic progenitor cells. Additionally, we have focused on capturing networks from closely related cells in the vicinity of the commitment boundary, to gain insight into the evolution of GRNs relevant to lineage specification.

The erythro-myeloid bifurcation is an intensively studied paradigm, and both transcription factors (TFs) and regulatory motifs involved in physiological transcriptional programming of these alternative fates have been described (Wolff and Humeniuk, 2013). Key players include Gata factors, the Ets family protein Pu.1, and C/ebp family members, whose potency has been demonstrated in the experimental reprogramming of blood lineages (Graf and Enver, 2009). We used two distinct cell commitment scenarios to obtain high resolution around the early phase of commitment and lineage specification. First, we identified and prospectively isolated cells spontaneously committing to both lineages under culture conditions that maintain self-renewing (SR) cells and lack pro-differentiative cytokines. Second, we

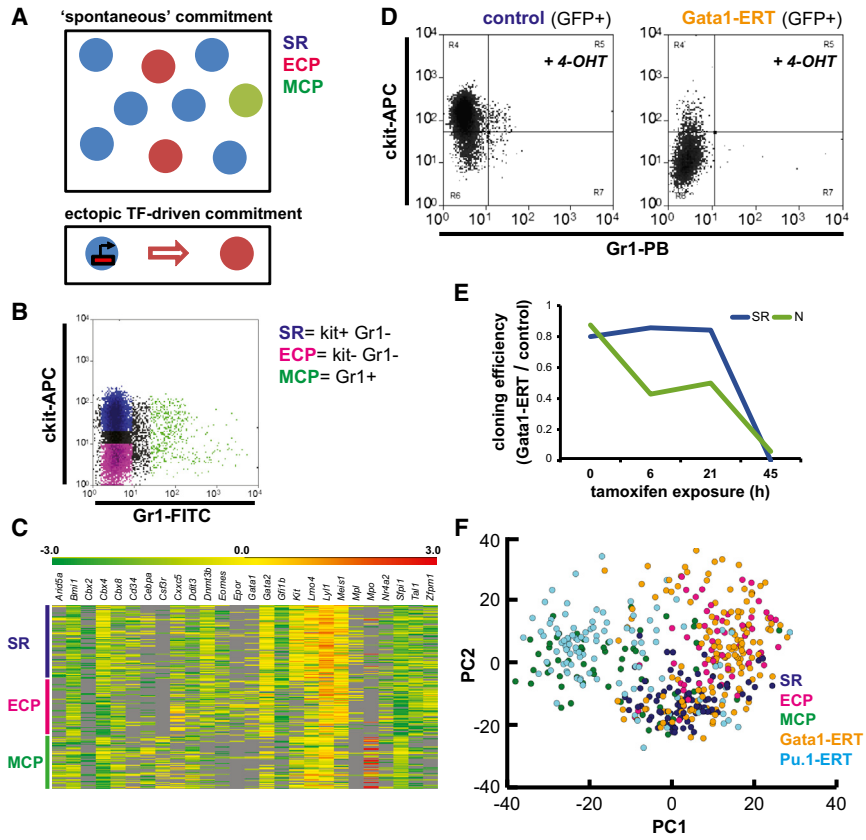


Figure 1. Single-Cell Transcriptional Profiling Captures Molecular Spaces of Lineage Commitment

(A) Depiction shows the two modes of commitment surveyed in this study: (top) hierarchically related SR cells, ECPs, and MCPs in equilibrium in a multipotent cell culture system; (bottom) unilineage commitment of SR cells driven by a single TF.

(B) Flow cytometry plot shows co-existing SR, ECP, and MCP cells in an FDCPmix culture under SR conditions.

(C) Heatmap of expression profiles of 26 genes in individual FDCPmix SR, ECP, and MCP cells. The gene panel represents the consensus analyzed in sufficient cell numbers in all compartments between replicate experiments. Data are *Z* score-normalized ΔCt values; undetectable expression, gray.

(D) Flow cytometry plots show FDCPmix SR cells transduced with control empty vector or with a GATA1-ERT fusion after 21-hr activation with tamoxifen (4-OHT, 1 μM).

(E) Cloning efficiency of control and Gata1-ERT GFP+ cells cultured under SR and neutrophil differentiation (N) conditions during a time course of 4-OHT activation. Sampling times correspond to those of single-cell qRT-PCR analysis. At each time point, 60 individual cells of each genotype were plated into SR or N conditions, and individually seeded wells were inspected at regular intervals for a 7-day period.

(F) PCA plot of the transcriptional profiles of individual FDCPmix cells undergoing distinct modes of lineage commitment. The first two PC explain 31% of the data variance; n = 82 (SR), 60 (ECP), 59 (MCP), 147 (Gata1-ERT), and 103 (Pu.1-ERT). A consensus set of 22 genes was analyzed in all five compartments.

used inducible variants of the lineage-affiliated TFs Gata1 and Pu.1 to drive cells into lineage specification, again in the absence of pro-differentiative cytokines. This approach allows timed sampling after instigation of a discrete lineage trigger, and it may provide a more homogeneous molecular entry into commitment and lineage development.

We describe state-distinct networks in multipotent and early lineage-committed cells and, in particular, highlight the existence of lineage-conflicting programs at the emergence of lineage choice. We further identify an axis involved in lineage specification that includes *Gata2* and *Ddit3*, a C/ebp family member previously implicated in stress response (Zinszner et al., 1998) and described as a potential target of erythropoietin signaling in erythro-leukemic cell lines (Coutts et al., 1999).

RESULTS

We have explored commitment in the non-transformed bone marrow (BM)-derived hematopoietic multipotent cell line FDCPmix. FDCPmix is karyotypically normal, IL-3 dependent, and capable of multilineage differentiation in response to the appropriate environmental cues. Under maintenance culture conditions, SR and lineage-committed cells (erythroid- or myeloid-committed progenitors [ECPs or MCPs]) co-exist in culture (Figure 1A) and can be isolated on the basis of their surface phenotype (Figure 1B). SR cells are Kit+Gr1− cells (Figure 1B) with proliferative capacity in bulk and clonal cultures (Figures S1A and S1B) and are uniquely able to faithfully reconstitute the cellular heterogeneity observed in maintenance cultures (Figure S1C). Lineage-committed cells devoid of SR potential are Gr1+ MCPs, with an early myeloid morphology (Figure S1D) and no erythroid differentiation capacity (Figure S1E), and kit−Gr1− ECPs, with accelerated erythroid differentiation (Figure S1E) but minimal or no contribution to neutrophil cultures (Figure S1F). The transcriptional signatures of SR, MCP, and ECP compartments are readily distinct (Figure S1G) and confirm their lineage affiliation (Table S1).

We next explored cellular heterogeneity within these cell compartments using single-cell multiplex qRT-PCR. The results showed substantial cell-to-cell heterogeneity (Figure 1C) and overlap (Figure 1F) within and among all three compartments; nevertheless, the compartments may be robustly identified using single and dual-gene classifiers (Table S1). The transcriptional heterogeneity observed within individual populations may highlight the capture of multiple contemporaneous molecular programs underlying lineage commitment under self-renewal conditions. Since TFs are potent instigators of lineage determination or reprogramming, we next examined cellular and transcriptional heterogeneity in FDCPmix cells driven to erythroid or myeloid lineages through expression of inducible Gata1 or

Pu.1 estrogen receptor fusions, respectively (Figure S1H). We analyzed the transcriptional programs of single cells captured at various time points (Figure S1I) after induction. In parallel, we studied the temporal dynamics of lineage commitment in this setting and functionally tested their commitment status by evaluating the following: (1) their retention of self-renewal potential, i.e., their capacity to re-initiate maintenance cultures; and (2) their lineage potential in response to various cytokine cues. This experimental design affords a dynamic appreciation of cellular and molecular mechanisms employed in lineage specification.

Activation of Gata1 in SR cells led to phenotypic changes at as early as 4–6 hr (Figure 1D), accompanied by the loss of neutrophil potential (at 6 hr) and followed by the loss of SR capacity (at 45 hr) (Figure 1E). Enforced Pu.1 activity resulted in the loss of clonogenic SR potential and elicited a myeloid differentiation bias in a more extended time frame (Figures S1J and S1K). Single-cell transcriptional profiling during TF-driven commitment confirmed the lineage identity of the cells obtained, which broadly separated away from the SR state along erythroid (Gata1) and myeloid (Pu.1) axes (Figure 1F). The analysis also revealed significant heterogeneity of molecular programs throughout the process of commitment. Gata1-driven cells co-occupied the same transcriptional space as ECP cells and showed a similar extent of cell-to-cell variation. In contrast, Pu.1-driven cells, while similarly heterogeneous to MCPs, appeared to occupy a distinct territory (data not shown). This presumably reflects the neutrophilic status of MCPs and a monocytic bias of Pu.1-ERT-differentiated cells, consistent with prior reports of Pu.1-driven cell fate (Laslo et al., 2006).

We exploited the heterogeneity of cells at early stages of spontaneous and TF-driven lineage commitment to explore the transcriptional networks controlling lineage specification. Inspection of patterns of expression for individual genes revealed a fraction of cells in which the gene is off, and a fraction of cells expressing the gene to varying levels (on) (Figure 2A). The on/off status can be described as binary while the distribution of on values represents a continuous component of the data. Thus, for any given gene pair, both binary and continuous relationships are possible; this is exemplified for *Gata1* and *Epor* in Figure 2B. We sought to capture both kinds of information to infer putative transcriptional networks. Methodologically, we used odds ratio (OR) to quantify on/off gene-to-gene associations (Figure 2C) and Spearman rank correlation to measure correlations between gene expression levels (Figure 2D). We combined gene associations obtained by both methodologies to infer putative regulatory networks characterizing SR states and the different modes of lineage commitment (Figure 2E).

At coarse grain, the networks revealed increased connectivity in the lineage-committed compared to the SR state. Also, commitment appeared associated with a higher frequency of negative associations between genes (Figures 2F and S2A), including known lineage-determining factors (Table S1). While this may be, to some extent, a function of the genes analyzed, it also may reflect mechanistically distinct processes governing acquisition of lineage identity versus exit from self-renewal. Negative associations are less prominent in the full-activation time courses of TF-driven commitment, as the networks capture not only the early processes of lineage specification, but also the later consolidation of the differentiation program. This likely increases the proportion of positive associations between lineage-affiliated genes. In contrast, detailed temporal analysis of Gata1-ERT-driven lineage specification revealed that cross-antagonistic associations between lineage determinants peak at 6 hr (Figure 2G), coincident with early loss of neutrophil differentiation potential (Figure 1E) en route to lineage commitment, suggesting that resolution of lineage conflicts is an early step in acquisition of lineage identity. In this respect, *Ddit3* emerges as an interesting candidate in lineage cross-antagonism: it is positively associated with *Gata2*, both in SR and committed cells (Figure S2B), and negatively associated with the neutrophil determinant *Cebpa*, either directly or through *Gata2*, in MCPs and at early stages of erythroid lineage commitment (Figure S2B; Table S1). Since Ddit3 has not previously been tabled as a central regulator of erythro-myeloid lineage specification, we functionally tested its impact in loss- and gain-of-function experiments.

Knockdown of *Ddit3* (Figures S3A and S3B) in FDCPmix cells resulted in the loss of erythroid and mixed-lineage colonies, with no change to myelo-monocytic potential (Figure 3A). The same loss of erythroid potential in colony-forming assays was observed in stem and progenitor cells (KLS) from mouse BM upon knockdown (Figure 3B) and constitutive knockout (Figure 3C) of *Ddit3* expression. The data are compatible with a requirement for Ddit3 in the erythroid lineage, while it is dispensable for the development of the myeloid lineage. The negative association observed between *Ddit3* and *Cebpa* in inferred transcriptional networks from early stages of lineage specification (Figures 2E and S2B; Table S1) suggests that Ddit3 contributes to the erasure of myeloid potential. We tested this in myeloid-committed granulo-monocytic progenitors (GMPs) by enforcing *Ddit3* expression (Figure S3C), resulting in a transient re-acquisition of self-renewal potential (Figure 3D) and a dramatic change in the nature of the colonies obtained (Figure 3E), with the predominance of large GM colonies of immature appearance (Figure 3F). Cells in these colonies expressed immature surface markers and were predominantly lineage-negative kit+CD34+CD16/32+, thus presenting an essentially GMP phenotype albeit with variable levels of Sca1 expression; in contrast, cells in control colonies exhibited a differentiated Gr1+Mac1+ phenotype (data not shown). Taken together, the data suggest that ectopic expression of *Ddit3* in GMPs blocks lineage progression and transiently re-activates self-renewal capacity.

We used single-cell gene expression profiling of GMPs, either wild-type or transduced with a control vector or a *Ddit3*-expressing lentivirus, to interrogate the transcriptional program changes imposed by enforced expression of *Ddit3* and to inspect its role in remodeling of the transcriptional networks underlying lineage progression and/or identity. Enforcement of *Ddit3* changed the expression of two-thirds of genes (Figures S3D and S3E) predicted as its neighbors in our inferred transcriptional networks (Figure S2B), attesting to the robustness of our inference approach. Principal component analysis (PCA) of the populations of individual wild-type and transduced GMP cells separated *Ddit3*-expressing cells from controls (Figure 3G). This separation is mostly attributable to the increased expression of

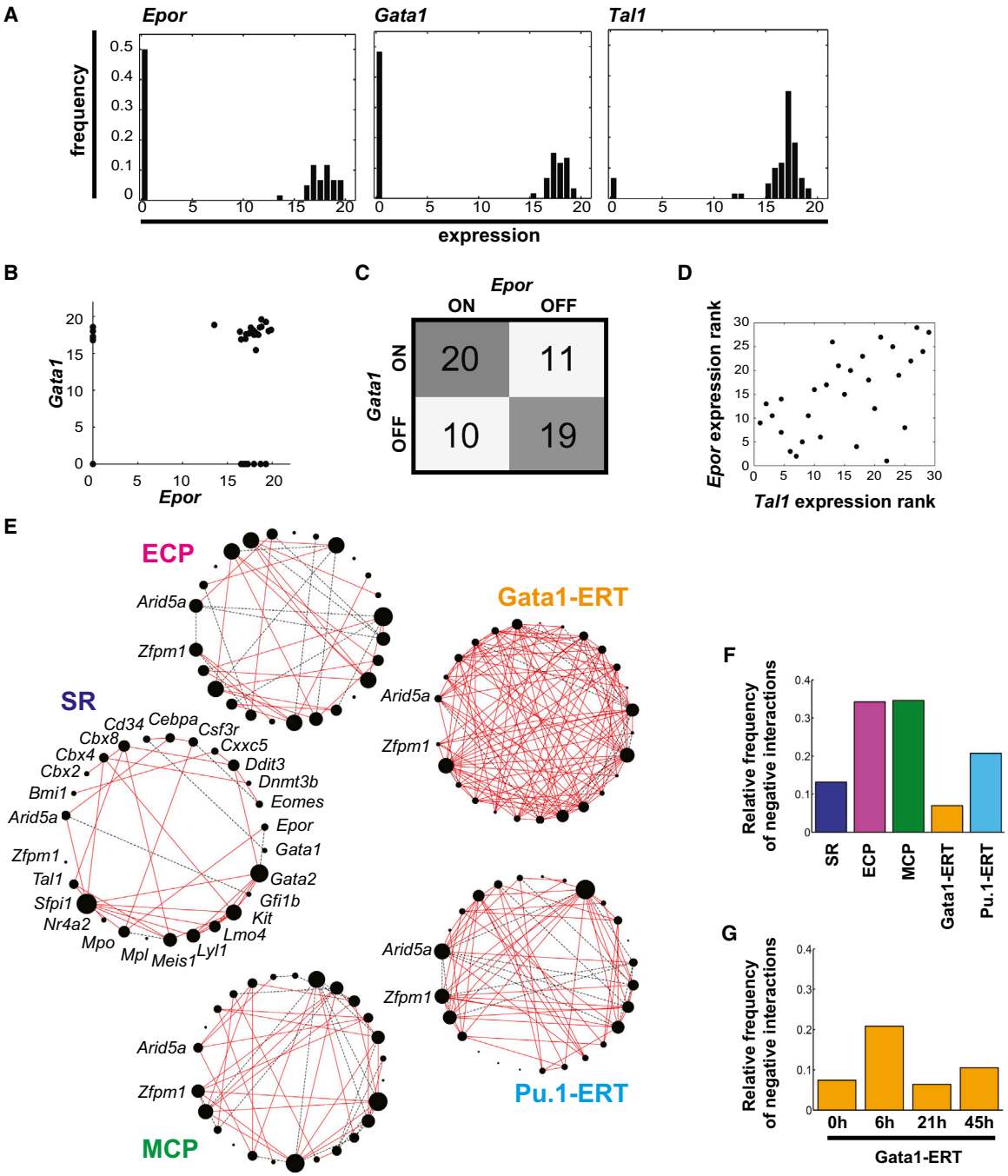


Figure 2. Combined Single-Cell Transcriptional Network Inference Methods Implicate Ddit3 in Lineage Specification

(A) Representative gene expression distributions for *Epor*, *Gata1*, and *Tal1* in ECPs are shown.

(B) Scatterplot of *Gata*1 and *Epor* single-cell expression highlights the dual aspect of the data with both binary (on/off) and continuous (expression-level) components.

(C) Contingency table summarizing on/off combination patterns of individual cells for *Epor* and *Gata1*. OR quantifies the diagonal versus off-diagonal of this matrix to infer significant positive and negative associations in the binary component of the data. *Gata1* and *Epor* show significant positive association (OR = 3.18; lower95CI > 1).

(D) Scatterplot of *Epor* and *Tal1* expression ranks in co-expressing cells. *Epor* and *Tal1* show significant positive correlation in the continuous component of the data inferred by Spearman rank correlation (r = 0.56; p = 0.002).

(E) Single-cell transcriptional networks in SR, ECP, MCP, Gata1-ERT, and Pu.1-ERT compartments were inferred by combined use of OR and Spearman rank correlation. Solid red lines, positive associations; dashed black lines, negative associations. Node size is proportional to the relative connectivity in each network.

(F) Proportion of negative interactions in the networks in (E) is shown.

(G) Proportion of negative interactions in Gata1-ERT networks at each time point is shown.

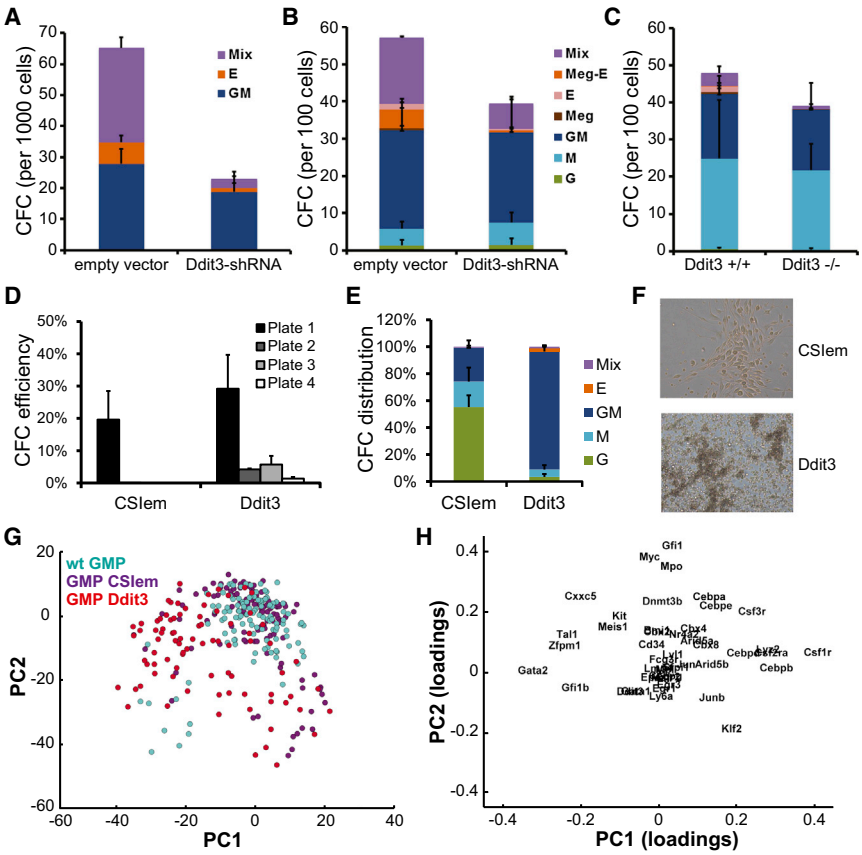


Figure 3. *Ddit3* Is Required in Early Erythroid Specification and Blocks Myeloid Lineage Progression

(A–C) Lineage potential of multipotent mouse BM cells upon the loss of *Ddit3* expression. CFC assays of FDCPmix cells (n = 3) (A) and primary KLS cells (n = 3) (B) upon *Ddit3* knockdown and of *Ddit3* knockout KLS cells (n = 4) (C) are shown. Error bars, SD.

(D) Re-plating capacity of primary BM GMPs upon enforced expression of *Ddit3* read in CFC assays (CSIem, empty vector; n = 4). Colonies were scored 7–10 days after plating of transduced cells (plate 1). The cellular content of the colonies obtained was re-seeded into successive CFC assays (plates 2–4) until the exhaustion of colony production.

(E) Distribution of colony types in CFC plate 1. Most GM colonies obtained upon *Ddit3*-enforced expression have a blast-like appearance. Error bars, SEM.

(F) Representative images of GM colonies in (E) are shown.

(G) PCA plot of the transcriptional profiles of individual GMPs, either untransduced (WT) or transduced with *CSIem*- or *Ddit3*-expressing lentiviral vectors, analyzed for the expression of 44 genes. The first two PC explain 24% of the data variance; n = 114 (CSIem), 84 (Ddit3), and 118 (WT).

(H) Gene loadings of PC1 and PC2 in (G). Genes with the most extreme positions along each axis contribute the most to cell separation along the respective PC.

early erythroid regulators *Gata2*, *Tal1*, *Zfpm1*/*Fog1*, and *Gfi1b* and the relative loss of M, GM, and G-CSF receptors (*Csf1r*, *Csf2ra*, and *Csf3r*, respectively) as well as of *C/ebp* family members (Figure 3H). The relative gain in the expression of erythroid-affiliated genes and loss of myeloid *Csf* receptors and *C/ebp* family TFs further developed with prolonged expression of *Ddit3* in GMPs under differentiation conditions that support multilineage output (Figures S3F and S3G). These data confirm *Ddit3* as a positive regulator of erythroid lineage specification at the expense of myeloid fate, providing an experimental validation of the predictive power of the networks we derived by analyzing the heterogeneity of single cells undergoing lineage specification.

We next asked if the relative gain in importance of erythroid-affiliated regulators in *Ddit3*-transduced GMPs was associated with a global remodeling of the transcriptional networks underlying GMP lineage identity. Indeed, we observed an overall loss in network connectivity specific to the activity of Ddit3 (Figure 4A) Moreover, there was a relative gain in connectivity of *Gata2* at the expense of myeloid hubs, as quantified in Figure S4A. For a broader appreciation of the transcriptional changes induced by Ddit3, we performed RNA sequencing (RNA-seq). GMPs transduced with *Ddit3* or control vector (CSIem) were cultured for up to 5 days under conditions supportive of multilineage output. Similarly to cells obtained from colony-forming assays, *Ddit3*-transduced cells retained a GMP-like phenotype, while control cells acquired differentiated myeloid surface markers (Figure S4B). The global transcriptional profiles of cells with enforced expression of *Ddit3* (Figure S4C) were clearly distinct from control-transduced and wild-type GMPs. Gene set enrichment analysis (GSEA) showed that *Ddit3* expression is associated with global loss of GMP programs and concomitant upregulation of Meg-E-affiliated signatures (Figure 4B). Interestingly, signatures representative of pre-GM cells, the developmental precursors of GMPs (Figure 4C), also were upregulated (Figure 4B). These data suggest that Ddit3 acts through the development or stabilization of a more primitive precursor with inherent Meg-E potential (Figure 4C). Analysis of the networks derived from wild-type GMPs (Figure S4D) as well as *Ddit3*-transduced cells exposed to conditions supportive of multilineage output for 2 days (Figure 4A) revealed increased importance of specific erythroid versus myeloid regulatory nodes.

Given the association of *Gata2* and *Ddit3* seen in our network analysis of lineage commitment (Figure 2E) and the increase in activity of *Gata2* as a hub in *Ddit3*-transduced GMPs (Figure 4D), we explored the behavior of Gata2 target genes, previously identified by chromatin immunoprecipitation sequencing (ChIP-seq) in FDCPmix (May et al., 2013), in response to *Ddit3* expression. GSEA on the RNA-seq data from *Ddit3* and control vector-transduced cells provided evidence for a coincidence of Gata2 and Ddit3-driven gene expression programs (Figure 4E). These data position *Ddit3* in an erythroid transcriptional network nucleated by *Gata2*. To further explore the *Ddit3*-*Gata2* axis in lineage specification, we focused on gene expression programs resident

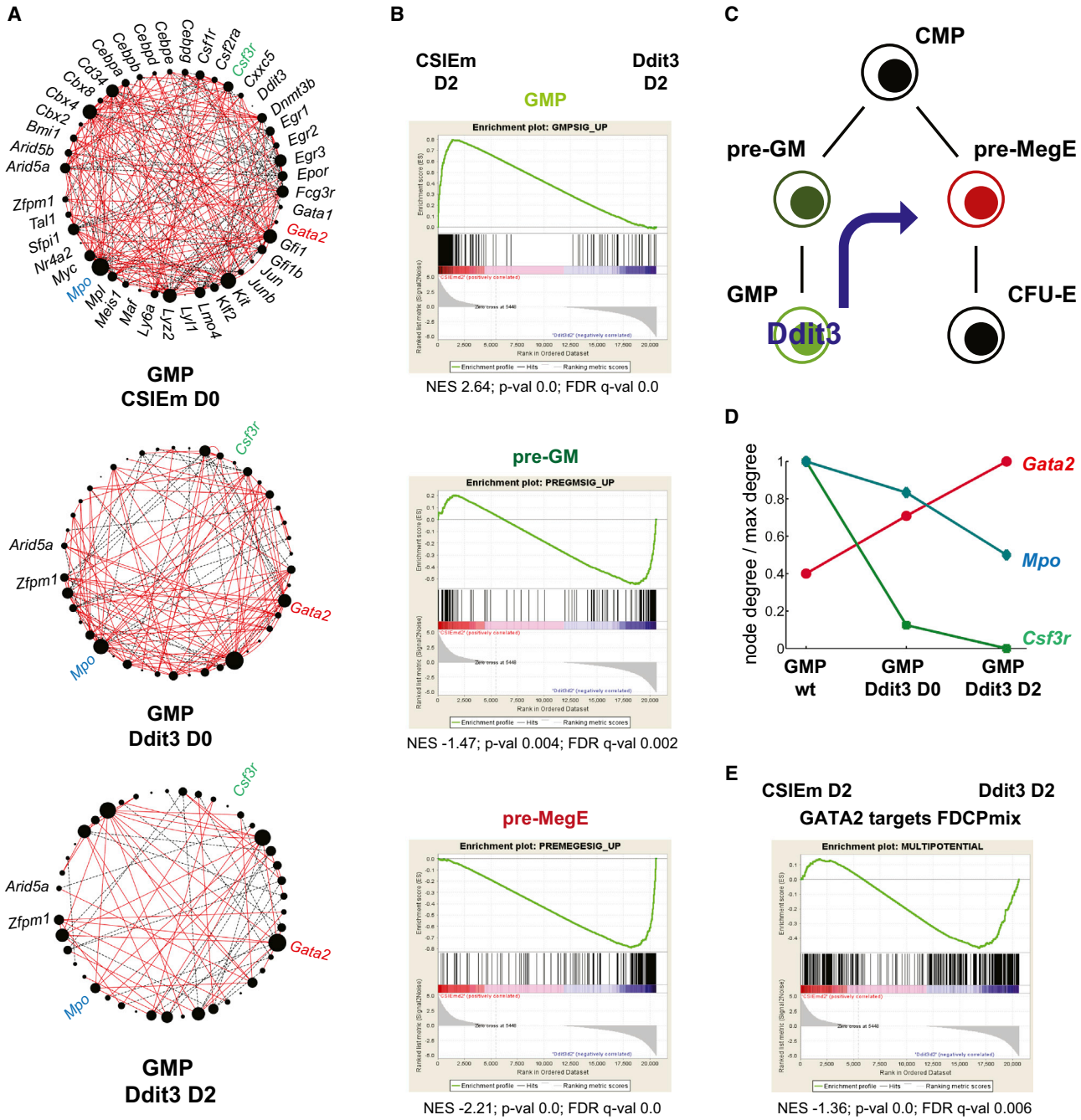


Figure 4. Ddit3 Remodels the GMP Transcriptional Landscape around *Gata2*

(A) Transcriptional networks of GMPs transduced with *CSIEm* and *Ddit3* and after 2-day culture under multilineage differentiation conditions. Network inference and representation are as in Figure 2. Highlighted in color are differential regulatory hubs (*Gata2*, *Csf3r*, and *Mpo*); their connectivity is quantified in Figure S4A.

(B) GSEA of early progenitor-affiliated signatures in the transcriptional programs of *CSIem*- and *Ddit3*-transduced GMPs is shown.

(C) Diagram shows *Ddit3*-driven lineage remodeling of GMPs.

(D) Connectivity of the differential regulatory hubs in (A) upon culture of *Ddit3*-transduced GMPs. Networks are represented in (A) and Figure S4D.

(E) GSEA of Gata2 targets in *CSIem*- and *Ddit3*-transduced GMPs. Transduced cells in (B, D, and E) were cultured for 2 days under multilineage differentiation conditions.

in those single cells that co-expressed the two genes, and derived networks from multipotent and lineage-committed compartments (Table S1). Within the ECP compartment, this analysis revealed the following: (1) an association between the two genes that was not seen when cells were used irrespective of *Ddit3*/*Gata2* status, and (2) the involvement of both these genes in

anti-correlations with myeloid-affiliated regulators. A similar picture now also emerges in multipotent cells, highlighting the gains in network information obtainable from interrogating single cells selected on the basis of specific co-expression patterns.

DISCUSSION

Our analysis of cells in the vicinity of commitment highlights the existence of lineage-conflicting programs at the emergence of lineage choice, and it identifies the importance of a *Ddit3-Gata2* axis in this process. In revealing this association, it was instrumental that our network inference approach acknowledged and took account of both binary and continuous components of single-cell gene expression. As such, we captured both on/off relationships and the associations seen between distributions of on expression levels in co-expressing cells. Furthermore, having established potentially interesting associations, we zoomed in on single cells where the associations were present to refine their specific network identity. In the case of *Ddit3-Gata2*, this supported the importance of cross-antagonistic interactions in lineage decisions. While in principle this approach can be applied to any gene-gene interaction in an unsupervised manner, it would require profiling of considerably higher numbers of individual cells to ensure statistical robustness and systematically explore all possible interactions.

To validate our inferred networks, we elected to experimentally test the predicted role of Ddit3 in erythroid lineage specification. Functional experiments revealed an early erythroid effect from *Ddit3* loss of function produced by knockdown or genetic knockout. However, we note that the constitutive *Ddit3* knockout mice used in this study did not display any discernible erythroid defect (C.F., unpublished data). Thus, lineage-determining networks may be robust to *Ddit3* deletion or constitutive deletion may be compensated during development. Forced *Ddit3* expression does promote erythroid programming, highlighting a potential role for the *Ddit3* node in early development of the erythroid lineage. Previous reports of Ddit3 function in late differentiation of erythro-leukemic cells (Coutts et al., 1999) are compatible, albeit distinct, from our proposed role in erythroid specification. Interestingly, our inspection of CFU-e potential in mouse BM stem and erythroid progenitor compartments upon Ddit3 ablation did not reveal a late erythroid defect (C.F., unpublished data).

The capacity of TFs to re-program cells has been used as a test of their lineage-determining capacity. In the case of Gata1, its introduction into GMPs results in the expression of erythroid potential with the appearance of large blast-like multipotential colonies in vitro (Heyworth et al., 2002). Ddit3 likewise is able to confer erythroid potential on GMPs. Molecular analysis suggests that *Ddit3*-enforced expression in GMPs leads to expression of erythroid-affiliated genes and an overall transcriptional state similar to that of pre-GM, the precursors of GMPs. One may presume that since pre-GM cells lie upstream of GMPs their transcriptional programs should have diverged less from multipotent cells that retain both GM and E potential. This would explain the E signatures seen in *Ddit3*-expressing GMPs. By which cellular mechanism does Ddit3 effect these changes? One possibility is that rare pre-GMs exist within prospectively isolated GMP populations and that these are preferentially selected for by Ddit3. This seems unlikely, given the kinetics of the changes in cells and gene expression observed. Alternatively, Ddit3 may regulate a subset of the erythroid program that is simply overlaid on the existing GMP program, resulting in a mixed-lineage program that is reminiscent of pre-GM cells. In such a model, Ddit3 effects a pre-GM state from GMPs, recapitulating its physiological role in lineage programming. Recently, Nerlov and colleagues have suggested that high levels of erythropoietin (EPO) may have instructive effects on lineage specification in vivo (Grover et al., 2014), and interestingly *Ddit3* has been suggested as a target of EPO signaling in erythro-leukemic cells (Coutts et al., 1999). The strong network association observed between Ddit3 and Gata2 may indicate that Ddit3 acts on an early erythroid signature primed by Gata2 in cells of mixed-lineage potential (May et al., 2013) and present in the pre-GM state. Elements of this signature may be required for erythroid lineage progression and, thus, explain the erythroid defect observed upon Ddit3 loss of function.

It is interesting to speculate as to the existence of cross-antagonistic interactions between Gata2-centered networks and C/ebp-driven myeloid programs putatively effected through Ddit3. Ddit3 heterodimerizes with C/ebp family members to form complexes that cannot bind DNA, thus blocking activation of C/ebp-driven programs, a mechanism that has been described to block differentiation in mesenchymal lineages (Han et al., 2013; Shirakawa et al., 2006). A combination of Ddit3 structure-function mutant studies and direct investigation of Gata2 and C/ebp DNA binding in Ddit3-expressing GMPs will contribute to clarifying dislodgement of C/ebp complexes from their target genes as a putative mechanism of GMP lineage remodeling.

EXPERIMENTAL PROCEDURES

Mice

B6.129S-*Ddit3*tm1Dron/J (Ddit3 KO) mice (Jackson ImmunoResearch Laboratories) and C57BL/6 mice were maintained in the John Radcliffe Hospital and CR-UK London Research Institute animal facilities in accordance with Home Office regulations.

Cell Culture and Lentiviral Transductions

FDCPmix culture conditions, lentiviral transductions with *Gata1-ERT* and *Pu.1-ERT* constructs, and tamoxifen activation were performed as described previously (May et al., 2013). Lentiviral transductions of FDCPmix cells with *Ddit3*-small hairpin RNA (shRNA) were performed under maintenance culture conditions; transductions of primary BM cells were performed in serum-free expansion medium (SFEM) (STEMCELL Technologies) supplemented with mouse stem cell factor (SCF) and Flt3L (50 ng/ml). GFP^+ cells were sorted after 2 days for downstream assays. In some experiments, GFP^+ GMPs were cultured for up to 5 days in Iscove's modified Dulbecco's medium (IMDM) + 10% fetal calf serum (FCS) supplemented with mouse SCF (50 ng/ml), mouse IL-3 and IL-6 (10 ng/ml), and human recombinant erythropoietin (EPREX, 10 U/ml). Colony-forming cell (CFC) assays used M3234 supplemented with rat SCF (100 ng/ml); mouse IL-3 (0.01 ng/ml) and EPREX (10 U/ml) (FDCPmix); and M3434 or M3234 supplemented with mouse SCF (100 ng/ml), mouse IL-3, IL-11, GM-CSF, and Tpo (10 ng/ml), and EPREX (10 U/ml). All mouse cytokines were from PeproTech and CFC media were from STEMCELL Technologies.

Single-Cell qRT-PCR

Transcriptional profiling of up to 48 genes in individual cells was performed on a Fluidigm platform and the data retrieved and quality-controlled as described previously (Teles et al., 2014). The ΔCt values were calculated to the mean of the three control genes utilized. Heatmap representation of

Z score-normalized ΔCt values was performed in Genesis; PCA plots used the Statistical Toolbox in MATLAB (MathWorks). The Taqman probes used are listed in the Supplemental Experimental Procedures.

Classification and Network Inference

Single-cell gene expression data were linearly transformed as described previously (Teles et al., 2013). Logistic regression linear classifiers were used to infer the best predictor genes in the separation between two cell populations (Teles et al., 2013). Single-cell transcriptional networks were inferred by calculating significant pairwise associations using both continuous (Spearman rank correlations) and binary (OR) components of linearly transformed expression data. Spearman rank correlations were calculated between all pairs of genes co-expressed by a minimum of ten cells in a given population. Correlation coefficients >0.4 with $p < 0.01$ were considered to be significant. OR and respective 95% confidence intervals (CIs) were calculated based on presence/absence patterns of expression for all pairs of genes in a given population. Significant positive and negative associations were called when Lower95CI > 1 and Upper95CI < 1, respectively. Network representations of significant pairwise associations in both methods were produced using Cytoscape (Smoot et al., 2011)

Additional methods are described in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

The accession numbers for the microarray and RNA-seq data reported in this paper are GEO: GSE68754 and SRA: SRP045672, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2015.05.016.

AUTHOR CONTRIBUTIONS

C. Pina and C.F. designed and performed experiments and analyzed data. J.T. developed and implemented computational methods and analyzed data. G.M., Y.G., and J.B. performed experiments. D.W. and S.S. analyzed transcriptome data. P.E., M.O., and C. Peterson developed computational methods. T.E. supervised the project. C. Pina and T.E. wrote the paper with input from J.T., C.F., G.M., and C. Peterson.

ACKNOWLEDGMENTS

We thank Kevin Clark, Arnold Pizzey, Alex Tipping, and Craig Waugh for expert cell sorting. This work was funded by the Swedish Research Council (C. Peterson and J.T.); the Swedish Foundation for Strategic Research through Create Health (P.E. and M.O.); and the UCL/UCLH Biomedical Research Centre, Leukaemia and Lymphoma Research and Cancer Research UK program grants and Children with Cancer and Great Ormond Street Hospital charity grants (T.E.).

Received: August 21, 2014
Revised: April 2, 2015
Accepted: May 10, 2015
Published: June 4, 2015

REFERENCES

Bruno, L., Hoffmann, R., McBlane, F., Brown, J., Gupta, R., Joshi, C., Pearson, S., Seidl, T., Heyworth, C., and Enver, T. (2004). Molecular signatures of self-renewal, differentiation, and lineage choice in multipotential hemopoietic progenitor cells in vitro. Mol. Cell. Biol. *24*, 741–756.

Coutts, M., Cui, K., Davis, K.L., Keutzer, J.C., and Sytkowski, A.J. (1999). Regulated expression and functional role of the transcription factor CHOP (GADD153) in erythroid growth and differentiation. Blood *93*, 3369–3378.

Davidson, E.H. (2010). Emerging properties of animal gene regulatory networks. Nature *468*, 911–920.

Graf, T., and Enver, T. (2009). Forcing cells to change lineages. Nature *462*, 587–594.

Grover, A., Mancini, E., Moore, S., Mead, A.J., Atkinson, D., Rasmussen, K.D., O'Carroll, D., Jacobsen, S.E., and Nerlov, C. (2014). Erythropoietin guides multipotent hematopoietic progenitor cells toward an erythroid fate. J. Exp. Med. *211*, 181–188.

Guo, G., Luc, S., Marco, E., Lin, T.W., Peng, C., Kerenyi, M.A., Beyaz, S., Kim, W., Xu, J., Das, P.P., et al. (2013). Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. Cell Stem Cell *13*, 492–505.

Han, J., Murthy, R., Wood, B., Song, B., Wang, S., Sun, B., Malhi, H., and Kaufman, R.J. (2013). ER stress signalling through eIF2α and CHOP, but not IRE1α, attenuates adipogenesis in mice. Diabetologia *56*, 911–924.

Heyworth, C., Pearson, S., May, G., and Enver, T. (2002). Transcription factor-mediated lineage switching reveals plasticity in primary committed progenitor cells. EMBO J. *21*, 3770–3781.

Laslo, P., Spooner, C.J., Warmflash, A., Lancki, D.W., Lee, H.J., Sciammas, R., Gantner, B.N., Dinner, A.R., and Singh, H. (2006). Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. Cell *126*, 755–766.

May, G., Soneji, S., Tipping, A.J., Teles, J., McGowan, S.J., Wu, M., Guo, Y., Fugazza, C., Brown, J., Karlsson, G., et al. (2013). Dynamic analysis of gene expression and genome-wide transcription factor binding during lineage specification of multipotent progenitors. Cell Stem Cell *13*, 754–768.

Moignard, V., Macaulay, I.C., Swiers, G., Buettner, F., Schütte, J., Calero-Nieto, F.J., Kinston, S., Joshi, A., Hannah, R., Theis, F.J., et al. (2013). Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. Nat. Cell Biol. *15*, 363–372.

Moignard, V., Woodhouse, S., Haghverdi, L., Lilly, A.J., Tanaka, Y., Wilkinson, A.C., Buettner, F., Macaulay, I.C., Jawaid, W., Diamanti, E., et al. (2015). Decoding the regulatory network of early blood development from single-cell gene expression measurements. Nat. Biotechnol. *33*, 269–276.

Pina, C., Fugazza, C., Tipping, A.J., Brown, J., Soneji, S., Teles, J., Peterson, C., and Enver, T. (2012). Inferring rules of lineage commitment in haematopoiesis. Nat. Cell Biol. *14*, 287–294.

Rusnakova, V., Honsa, P., Dzamba, D., Ståhlberg, A., Kubista, M., and Anderova, M. (2013). Heterogeneity of astrocytes: from development to injury - single cell gene expression. PLoS ONE *8*, e69734.

Shirakawa, K., Maeda, S., Gotoh, T., Hayashi, M., Shinomiya, K., Ehata, S., Nishimura, R., Mori, M., Onozaki, K., Hayashi, H., et al. (2006). CCAAT/enhancer-binding protein homologous protein (CHOP) regulates osteoblast differentiation. Mol. Cell. Biol. *26*, 6105–6116.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L., and Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics *27*, 431–432.

Ståhlberg, A., Andersson, D., Aurelius, J., Faiz, M., Pekna, M., Kubista, M., and Pekny, M. (2011). Defining cell populations with single-cell gene expression profiling: correlations and identification of astrocyte subpopulations. Nucleic Acids Res. *39*, e24.

Teles, J., Pina, C., Edén, P., Ohlsson, M., Enver, T., and Peterson, C. (2013). Transcriptional regulation of lineage commitment—a stochastic model of cell fate decisions. PLoS Comput. Biol. *9*, e1003197.

Teles, J., Enver, T., and Pina, C. (2014). Single-cell PCR profiling of gene expression in hematopoiesis. Methods Mol. Biol. *1185*, 21–42.

Wolff, L., and Humeniuk, R. (2013). Concise review: erythroid versus myeloid lineage commitment: regulating the master regulators. Stem Cells *31*, 1237–1244.

Zinszner, H., Kuroda, M., Wang, X., Batchvarova, N., Lightfoot, R.T., Remotti, H., Stevens, J.L., and Ron, D. (1998). CHOP is implicated in programmed cell death in response to impaired function of the endoplasmic reticulum. Genes Dev. *12*, 982–995.

Cell Reports

Supplemental Information

Single-Cell Network Analysis Identifies DDIT3 as a Nodal Lineage Regulator in Hematopoiesis

Cristina Pina, José Teles, Cristina Fugazza, Gillian May, Dapeng Wang, Yanping Guo, Shamit Soneji, John Brown, Patrik Edén, Mattias Ohlsson, Carsten Peterson, and Tariq Enver

SUPPLEMENTARY METHODS

Mice

B6.129S-*Ddit3*tm1Dron/J (Ddit3 KO) mice (Marciniak et al., 2004) (Jackson Laboratories) and C57BL/6 mice were maintained in the John Radcliffe Hospital (Oxford) and CR-UK London Research Institute animal facilities in accordance with Home Office regulations.

Flow cytometry

KLS and GMP cells were isolated from mouse BM essentially as described (Akashi et al., 2000), using pre-enrichment on magnetic columns (Miltenyi Biotech) with either lineage$^+$ cell depletion (LD columns) or Kit$^+$ cell enrichment (LS columns) as per manufacturer's protocols. Cell sorting was performed on MoFlo (Dako Cytomation) and FACSAriaII (BD Biosciences) instruments; analysis used Cyan ADP and Gallios analyzers (Dako Cytomation). For single-cell experiments, cells were sorted into tubes and re-sorted on the same gates for single-cell deposition into 96-well plates. Antibodies used are listed below.

Supplementary Methods Table 1: Flow cytometry antibodies used in this study (refers to Experimental Procedures)

Antibody	Fluorochrome	Clone	Supplier
CD3ε	purified biotinylated	145-2C1	BioLegend
CD4	purified biotinylated	G41.5	eBioscience
CD8a	purified biotinylated	53-6.7	eBioscience
CD11b / Mac1	purified biotinylated PE APC	M1/70	eBioscience
CD16/32 / FcgIII/IIR	FITC PECy7	93	eBioscience
CD34	FITC	RAM34	eBioscience

	Alexafluor 647 eFluor 660		
CD117 / c-kit	PE APC Alexafluor 780 Pacific Blue/eFluor 450	2B8	eBioscience/BD Pharmingen
B220	purified biotinylated	RA3-6B2	eBioscience
Ly-6A/E / Sca-1	PE PECy7 Pacific Blue	E13-161.7 or D7	BioLegend/BD Pharmingen
Ly-6C / Gr1	purified FITC PE PECy7 Pacific Blue	RB6-8C5	BioLegend/BD Pharmingen
Ter-119	purified biotinylated	TER-119	BD Pharmingen
$F(ab')_2$ goat anti-rat IgG (H/L)	PECy5		Invitrogen
Streptavidin	PECy7 APCCy7 Pacific Blue		eBioscience/Invitrogen

Note: The lineage antibody cocktail was CD3ε, CD4, CD8a, B220, Gr1, Mac1 and Ter119.

Lentiviral constructs

Full-length mouse *Ddit3* and *Ddit3-ΔC* mutant with N-terminal FLAG tags and engineered BamHI flanking restriction sites were PCR-cloned into T-easy vector (Promega) and subcloned into the BamHI site of CSIEm (*pHR-SIN-CSGW-ires-EmGFP*), as described (May et al., 2013). All cloning primer sequences are available upon request. Ddit3-shRNA (target sequence: AAGAGCAAGGAAGAACTAGGAAA; sense oligo: TGAGCAAGGAAGAACTAGGAAAGGGATCCTTTCCTAGTTCTTCCTTCGTCT TTTTTC; antisense oligo: TCGAGAAAAAAGACGAAGGAAGAACTAGGAAAGGATCCCTTTCCTAGTT CTTCCTTGCTCA) was cloned into the HpaI and XhoI sites of *pLentilox 3.7*, also as described (May et al., 2013).

Single-cell quantitative RT-PCR

The Taqman probes used in this study are listed below.

Supplementary Methods Table 2: Taqman probes used in this study
(refers to Experimental Procedures)

Gene	Assay number
Atp5a1	Mm00431960_m1
Hprt1	Mm00446968_m1
Ubc	Mm01201237_m1
Arid5a	Mm00524454_m1
Arid5b	Mm00517818_m1
Cbx2	Mm00483084_m1
Cbx4	Mm00483089_m1
Cbx8	Mm00489229_m1
Cd34	Mm00519283_m1
Cebpa	Mm00514283_s1
Cebpb	Mm00843434_s1
Cebpd	Mm00786711_s1
Cebpe	Mm02030363_s1
Cebpg	Mm01266786_m1
Csf1r	Mm01266652_m1
Csf2ra	Mm00438331_g1
Csf3r	Mm00432735_m1
Cxxc5	Mm00505000_m1
Ddit3	Mm00492097_m1
Dnmt3b	Mm01240113_m1
Egr1	Mm00656724_m1
Egr2	Mm00456650_m1
Egr3	Mm00516979_m1
Eomes	Mm01351985_m1
Epor	Mm00438760_m1
Fcgr3	Mm00438882_m1
Gata1	Mm00484678_m1

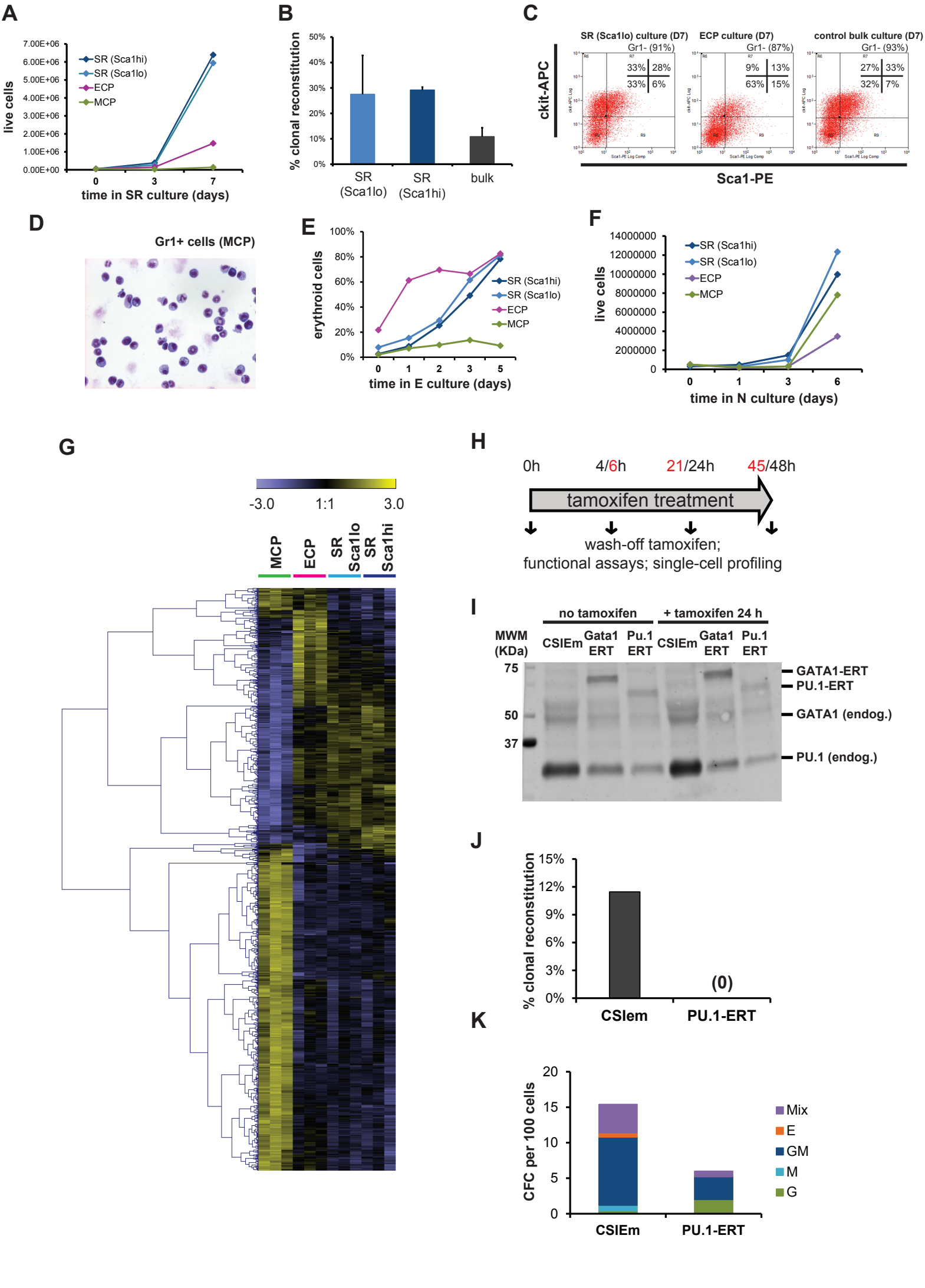
Gata2	Mm00492300_m1
Gfi1	Mm00515855_m1
Gfi1b	Mm00492319_m1
Jun	Mm00495062_s1
Kit	Mm00445212_m1
Klf2	Mm01244979_g1
Lmo4	Mm00495373_m1
Ly6a	Mm00726565_s1
Lyl1	Mm00493219_m1
Meis1	Mm00487664_m1
Mpl	Mm00440310_m1
Mpo	Mm00447886_m1
Myc	Mm00487803_m1
Nr4a2	Mm00443060_m1
Sfpi1	Mm00488140_m1
Tal1	Mm00441665_m1
Zfpm1	Mm00494336_m1

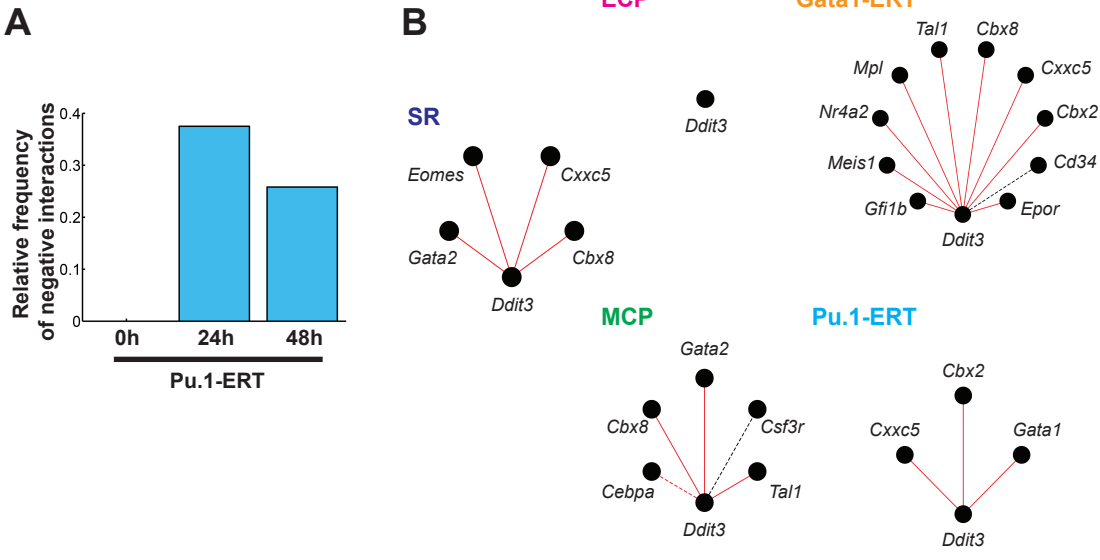
Note: Control gene expression assays are highlighted in gray.

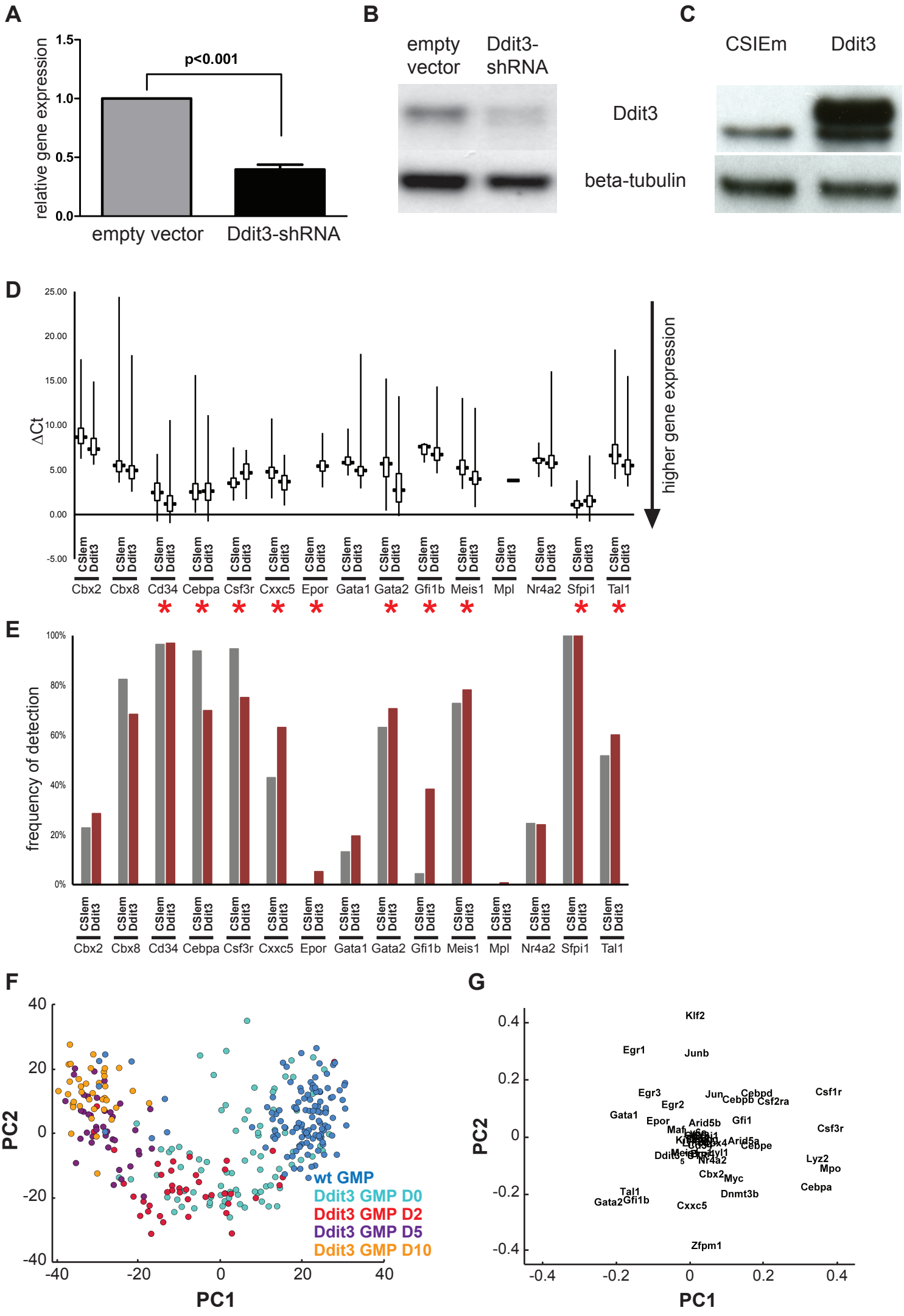
Microarray analysis and RNA-sequencing

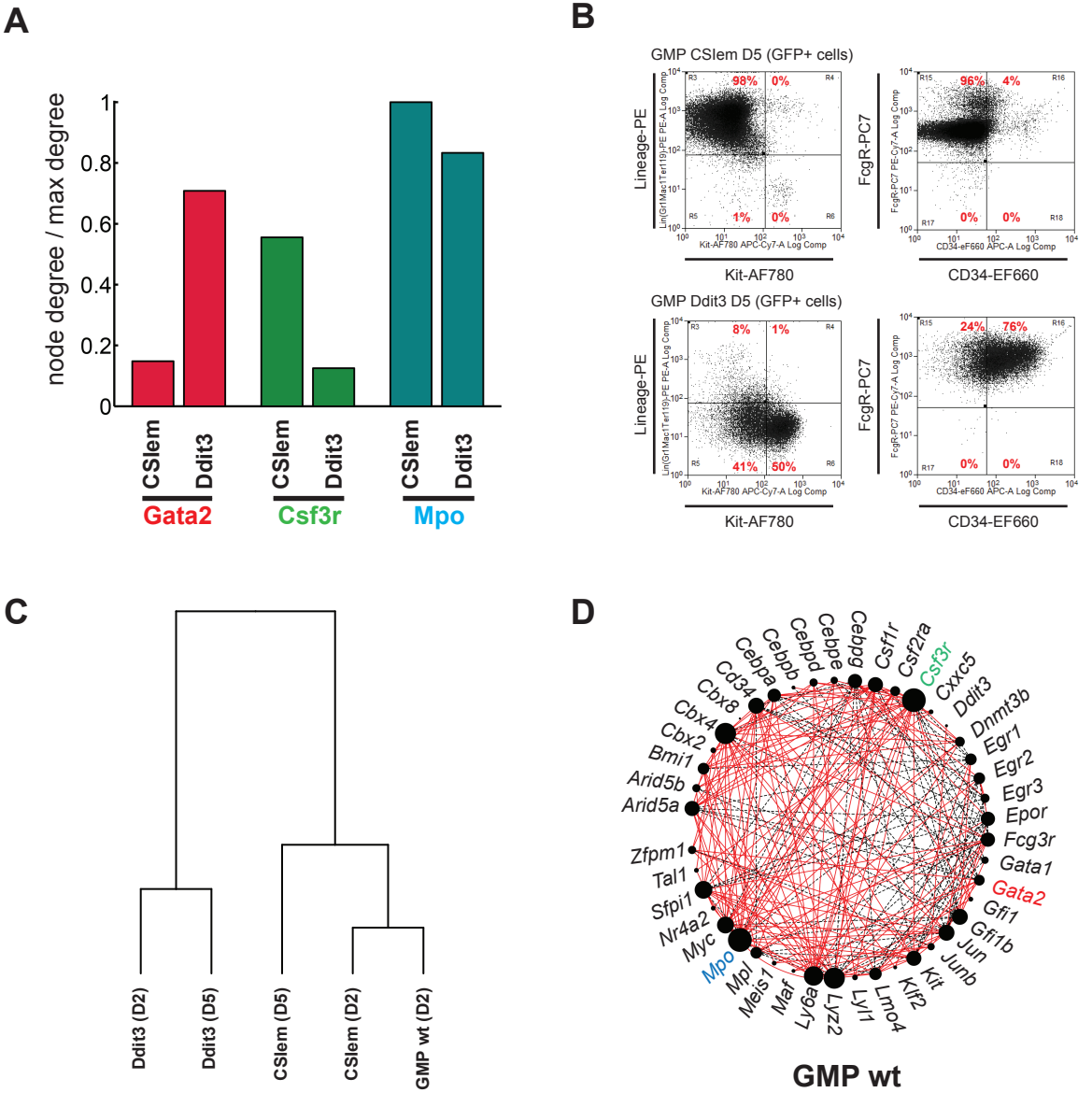
RNA was prepared from Trizol (Life Technologies) lysates. Microarray analysis was performed on an Agilent platform, as described (May et al., 2013). Differentials between any 2 compartments were calculated using LIMMA (Smyth, 2004); hierarchical clustering was performed in Genesis. For RNA sequencing, cDNA was prepared using Clontech SMARTer ultra-low input RNA kit for Illumina sequencing kit with sonication in a Covaris sonicator. 15ng cDNA were used for library preparation, using individual NEB reagents and NEB Next multiplex oligos for Illumina sequencing (E7335). Briefly, cDNA ends were repaired by incubation with T4 DNA polymerase, Klenow and T4 polynucleotide kinase (30', 20^{o}C) before addition of dA tails by Klenow 3'-5- exo minus enzyme (30', 37 ^{o}C). Quick T4 DNA

ligase was used to add NEB adapters (15', rt), followed by digestion with USER enzyme (15', 37 ^{o}C). DNA was amplified using Phusion master mix with high fidelity buffer, universal primer and an indexed oligo, with the following PCR program: 98^{o}C, 30s; 15-18x (98 ^{o}C, 10s; 65 ^{o}C, 30s; 72 ^{o}C, 30s); 72 ^{o}C, 5'. Library size selection (150-350 bp) was done by agarose gel electrophoresis. Libraries were sequenced by standard procedures on HiSeq, using 50bp single-end sequencing. FASTQ files for GMP wt, CSIEm D2, Ddit3 D2, Ddit3-ΔC D2, CSIEm D5, Ddit3 D5 and Ddit3-ΔC D5 were mapped onto the mouse genome (version mm10) using Tophat, and FPKM values measured with Cufflinks to represent gene expression levels (Trapnell et al., 2012). Only protein-coding genes were considered and expression values from individual isoforms were averaged. Gene set enrichment analysis (Subramanian et al., 2005) used the data for CSIEm D2 and Ddit3 D2 as phenotype-relevant datasets, and pre-MegE, CFUe, pre-GM and GMP gene signatures (calculated as 4-fold enrichments over the global average) as well as Gata2 ChIP targets in FDCPmix multipotential cells as background datasets (May et al., 2013). Following software developers' recommendations for conditions of no phenotypic replicates, analysis used up to 1500 genes per gene set, gene-set permutation methodology and an FDR cutoff of 0.05.


A
live cells
7.00E+06
6.00E+06
5.00E+06
4.00E+06
3.00E+06
2.00E+06
1.00E+06
0.00E+00
SR (Sca1hi)
SR (Sca1lo)
ECP
MCP
0
3
7
time in SR culture (days)
B
% clonal reconstitution
50%
40%
30%
20%
10%
0%
SR (Sca1lo)
SR (Sca1hi)
bulk
C
SR (Sca1lo) culture (D7)
Gr1- (91%)
ECP culture (D7)
Gr1- (87%)
control bulk culture (D7)
Gr1- (93%)
ckit-APC
33%
28%
33%
6%
9%
13%
63%
15%
27%
33%
32%
7%
Sca1-PE
D
Gr1+ cells (MCP)
E
erythroid cells
100%
80%
60%
40%
20%
0%
SR (Sca1hi)
SR (Sca1lo)
ECP
MCP
0
1
2
3
5
time in E culture (days)
F
live cells
14000000
12000000
10000000
8000000
6000000
4000000
2000000
0
SR (Sca1hi)
SR (Sca1lo)
ECP
MCP
0
1
3
6
time in N culture (days)
G
-3.0
1:1
3.0
MCP
ECP
SR Sca1lo
SR Sca1hi
H
0h
4/6h
21/24h
45/48h
tamoxifen treatment
wash-off tamoxifen;
functional assays; single-cell profiling
I
no tamoxifen
+ tamoxifen 24 h
MWM (KDa)
CSIEm
Gata1 ERT
Pu.1 ERT
CSIEm
Gata1 ERT
Pu.1 ERT
75
50
37
GATA1-ERT
PU.1-ERT
GATA1 (endog.)
PU.1 (endog.)
J
% clonal reconstitution
15%
12%
9%
6%
3%
0%
(0)
CSIem
PU.1-ERT
K
CFC per 100 cells
20
15
10
5
0
Mix
E
GM
M
G
CSIEm
PU.1-ERT


A
Relative frequency of negative interactions
0.4
0.3
0.2
0.1
0
0h
24h
48h
Pu.1-ERT
B
ECP
Ddit3
Gata1-ERT
Tal1
Cbx8
Mpl
Cxxc5
Nr4a2
Cbx2
Meis1
Cd34
Gfi1b
Epor
Ddit3
SR
Eomes
Cxxc5
Gata2
Cbx8
Ddit3
MCP
Gata2
Cbx8
Csf3r
Cebpa
Tal1
Ddit3
Pu.1-ERT
Cbx2
Cxxc5
Gata1
Ddit3


A
relative gene expression
1.5
1.0
0.5
0.0
p<0.001
empty vector
Ddit3-shRNA
B
empty vector
Ddit3-shRNA
Ddit3
beta-tubulin
C
CSIEm
Ddit3
D
ΔCt
25.00
20.00
15.00
10.00
5.00
0.00
-5.00
higher gene expression
CSlem
Ddit3
Cbx2
Cbx8
Cd34
Cebpa
Csf3r
Cxxc5
Epor
Gata1
Gata2
Gfi1b
Meis1
Mpl
Nr4a2
Sfpi1
Tal1
E
frequency of detection
100%
80%
60%
40%
20%
0%
F
PC2
PC1
wt GMP
Ddit3 GMP D0
Ddit3 GMP D2
Ddit3 GMP D5
Ddit3 GMP D10
G
PC2
PC1
Klf2
Egr1
Junb
Egr3
Egr2
Jun
Cebpd
Cebpb
Csf2ra
Csf1r
Gata1
Epor
Arid5b
Gfi1
Csf3r
Cebpe
Arid5a
Nr4a2
Lyz2
Mpo
Cbx2
Myc
Cebpa
Tal1
Gata2
Gfi1b
Cxxc5
Dnmt3b
Zfpm1


A
node degree / max degree
1
0.8
0.6
0.4
0.2
0
CSlem
Ddit3
Gata2
Csf3r
Mpo
B
GMP CSlem D5 (GFP+ cells)
Lineage-PE
Kit-AF780
FcgR-PC7
CD34-EF660
98%
0%
1%
0%
96%
4%
0%
0%
GMP Ddit3 D5 (GFP+ cells)
8%
1%
41%
50%
24%
76%
0%
0%
C
Ddit3 (D2)
Ddit3 (D5)
CSlem (D5)
CSlem (D2)
GMP wt (D2)
D
Cebpg
Csf1r
Csf2ra
Csf3r
Cxxc5
Ddit3
Dnmt3b
Egr1
Egr2
Egr3
Epor
Fcg3r
Gata1
Gata2
Gfi1
Gfi1b
Jun
Junb
Kit
Klf2
Lmo4
Lyl1
Lyz2
Ly6a
Maf
Meis1
Mpl
Mpo
Myc
Nr4a2
Sfpi1
Tal1
Zfpm1
Arid5a
Arid5b
Bmi1
Cbx2
Cbx4
Cbx8
Cd34
Cebpa
Cebpb
Cebpd
Cebpe
GMP wt

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. Functional and molecular characterization of FDCPmix compartments (refers to Figure 1 in main text). **A.** Growth curves of FDCPmix fractions cultured for 7 days under maintenance culture conditions; data are representative of >3 experiments. SR (Sca1hi) and SR (Sca1lo) represent Kit+Gr1- cells sorted as the highest and lowest 20% of a broad Sca1 distribution, as is apparent in C. As seen in panels A, B and E-G, and similarly to our findings in the multipotent hematopoietic EML cell line (Pina et al., 2012), these 2 fractions were indistinguishable both functionally and molecularly, and also behaved like total Kit+Gr1- cells (data not shown). They are thus treated as SR cells. **B.** Culture reconstitution capacity of total FDCPmix and SR fractions plated as single cells and cultured for 7 days under maintenance conditions. Clones containing more than 100 cells are considered to reflect the progeny of a cell with culture-reconstituting potential, used as an operational definition of self-renewal capacity. Data are mean + SD of 2 independent experiments, n=60 cells/experiment. No clones were obtained in ECP-initiated cultures (data not shown). **C.** Flow cytometry analysis of FDCPmix cultures initiated with SR and ECP fractions, as well as total cells (control bulk culture). **D.** Representative cytospin of MCP cells stained by the May Grunewald Giemsa (MGG) method. **E.** Kinetics of production of erythroid cells from FDCPmix fractions grown under erythroid differentiation conditions. Cytospins were performed at the days indicated and stained with MGG and O-dianosidine to detect hemoglobinized cells; cells were scored as blasts, erythroid or myeloid cells on the basis of their morphology (May et al., 2013). Data are representative of 3 experiments. **F.** Growth curve of FDCPmix fractions cultured under neutrophil

differentiation conditions; representative experiment. **G.** Unsupervised hierarchical clustering of differential genes between any 2 FDCPmix compartments and/or bulk FDCPmix cells analyzed on Agilent microarrays and represented in Genesis. The Z-score transformed data used for the analysis are represented in Supplementary File 1. **H.** Experimental protocol of Gata1-ERT and Pu.1-ERT activation in FDCPmix SR cells. Times of cell collection after activation are shown in red and black for Gata1-ERT and Pu.1-ERT, respectively. **I**. Western blot of Gata1-ERT and Pu.1-ERT expression in transduced FDCPmix cells with and without tamoxifen activation (24h). Endogenous Gata1 and Pu.1 proteins are also indicated. Antibodies used were anti-PU.1 (T-21), Santa Cruz Biotechnology, and anti-Gata1 ab11963, Abcam; secondary antibody detection with Odissey anti-rabbit 700. **J.** Clonal reconstitution capacity of FDCPmix SR cells transduced with *Pu.1-ERT* or control virus (*CSIEm*) and activated for 48h with tamoxifen. Assay design as in B. **K.** Lineage potential of control and Pu.1-ERT FDCPmix SR cells after 48h of tamoxifen activation, as tested by CFC assays in multipotential methylcellulose medium.

Figure S2. Network analysis of single-cell transcriptional programs in lineage commitment (refers to Figure 2 in main text). **A.** Proportion of negative interactions in Pu.1-ERT networks at each individual time point. **B.** *Ddit3* direct interactions in the single-cell transcriptional networks of SR, ECP, MCP, Gata1-ERT and Pu.1-ERT populations represented in Figure 2E.

Figure S3. Modulation of *Ddit3* levels of expression and analysis of the transcriptional consequences of enforcing *Ddit3* in GMP cells (refers to Figure 3 in main text). **A.** Quantification of *Ddit3* knockdown in FDCPmix cells by quantitative

RT-PCR; mean + SEM of 4 experiments. Mean knockdown = 0.397. **B.** Western blot analysis of *Ddit3* knockdown in mouse erythro-leukaemic MEL cells. **C.** Western blot analysis of enforced *Ddit3* expression in FDCPmix cells; CSIEm; empty control vector. Antibodies used were anti-DDIT3 (R-20) and anti-beta tubulin (H-235), Santa Cruz Biotechnology; secondary antibody detection with Hrp-conjugated anti-rabbit (GE Healthcare). **D**-**E**. Validation of the network neighbors of *Ddit3* identified in the analysis of FDCPmix cells (Figure S2B) using single-cell quantitative RT-PCR analysis of GMP cells where *Ddit3* expression was enforced. A total of 133 *Ddit3*-transduced and 114 control (CSIEm) cells from 4 independent experiments were analyzed by Fluidigm single-cell RT-qPCR and the respective ΔCt values (D, box plots) and frequency of detection (E) of neighbor genes are represented. Differentially expressed genes were determined using a Mann-Whitney test and are marked by asterisks. **F.** PCA plot of the gene expression profiles of individual GMP cells transduced with *Ddit3*-expressing lentiviral vector (D0) and cultured for 2 (D2) and 5 days (D5) in the presence of SCF, IL-3, IL-6 and EPO or tested in CFC assays under similar cytokine conditions (D10); untransduced GMP (wt) are included as control. The first 2 PC explain 31% of the data variance. N= 118 (wt), 84 (D0), 41 (D2), 39 (D5) and 37 (CFC). **G.** Gene loadings of PC1 and PC2 in F. Genes with the most extreme positions along each axis contribute the most to cell separation along the respective PC.

Figure S4. Ddit3 enforces global remodeling of lineage-affiliated programs in GMP (refers to Figure 4 in main text). **A.** Relative connectivity of *Gata2*, *Mpo* and *Csf3r* hubs in the transcriptional networks of *CSIEm* and *Ddit3*-transduced GMP represented in Figure 4A (top and middle). **B.** Flow cytometry plots of *CSIEm* and

Ddit3-transduced GMP cultured for 5 days in multi-lineage differentiation conditions (SCF, IL-3, IL-6 and EPO). Cells were re-gated as GFP^+, and $Lineage^+$ indicates detection of Gr1 and/or Mac-1 antigens. **C.** Cluster dendrogram of the RNA-seq profiles of *CSIEm* and *Ddit3* -transduced GMP, as well as wt GMP cultured for 2 and 5 days in multi-lineage differentiation conditions. **D.** Transcriptional networks of untransduced (wt) GMP. Network representation as in Figure 2E; the differential regulatory hubs highlighted in color are quantified in Figure 4D.

SUPPLEMENTARY FILES

Supplementary File 1: Pina_Supplementary_File_1.xlsx. Lists all significant pair-wise gene associations for network inference in analysis of commitment in FDCPmix (tabs 1-4); single and dual-gene classifiers from FDCPmix single-cell gene expression data (tabs 5-6); differentially expressed genes by microarray analysis of FDCPmix fractions (tab 7). Refers to Figures 1 and 2 in the main text.

Supplementary File 2: Pina_Supplementary_File_2.xlsx. Lists all significant pair-wise gene associations for network inference in wild-type and lentivirally-transduced GMP cells (1 tab). Refers to Figure 4 in the main text.

SUPPLEMENTARY REFERENCES

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

Marciniak, S.J., Yun, C.Y., Oyadomari, S., Novoa, I., Zhang, Y., Jungreis, R., Nagata, K., Harding, H.P., and Ron, D. (2004). CHOP induces death by promoting protein synthesis and oxidation in the stressed endoplasmic reticulum. Genes Dev *18*, 3066-3077.

May, G., Soneji, S., Tipping, A.J., Teles, J., McGowan, S.J., Wu, M., Guo, Y., Fugazza, C., Brown, J., Karlsson, G., *et al.* (2013). Dynamic analysis of gene expression and genome-wide transcription factor binding during lineage specification of multipotent progenitors. Cell Stem Cell *13*, 754-768.

Pina, C., Fugazza, C., Tipping, A.J., Brown, J., Soneji, S., Teles, J., Peterson, C., and Enver, T. (2012). Inferring rules of lineage commitment in haematopoiesis. Nat Cell Biol *14*, 287-294.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol *3*, Article3.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., *et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript

expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc *7*, 562-578.