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

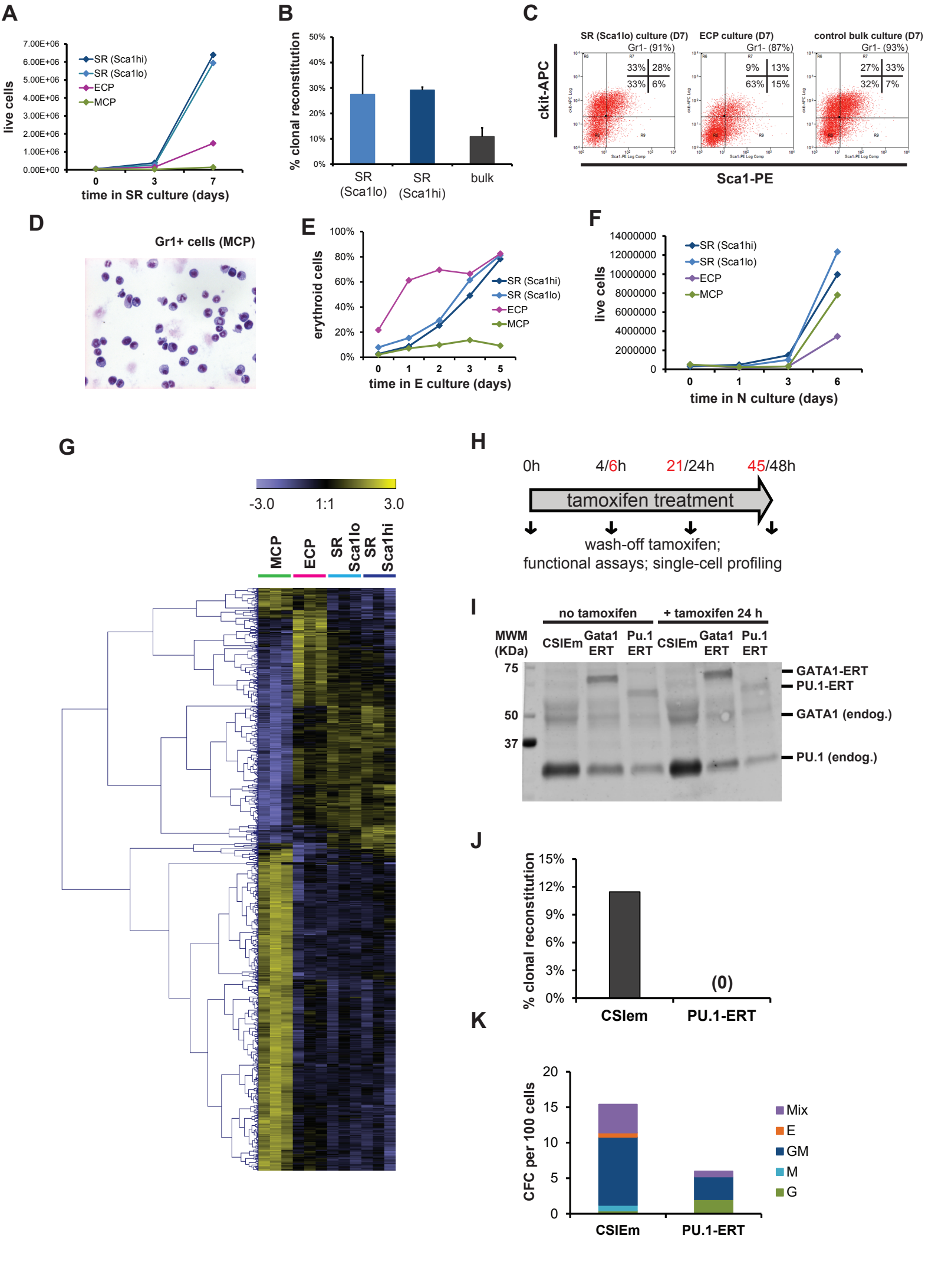
Gata2	Mm00492300_m1
Gfi1	Mm00515855_m1
Gfi1b	Mm00492319_m1
Jun	Mm00495062_s1
Kit	Mm00445212_m1
Klf2	Mm01244979_g1
Lmo4	Mm00495373_m1
Ly6a	Mm00726565_s1
Lyl1	Mm00493219_m1
Meis1	Mm00487664_m1
Mpl	Mm00440310_m1
Mpo	Mm00447886_m1
Myc	Mm00487803_m1
Nr4a2	Mm00443060_m1
Sfpi1	Mm00488140_m1
Tal1	Mm00441665_m1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

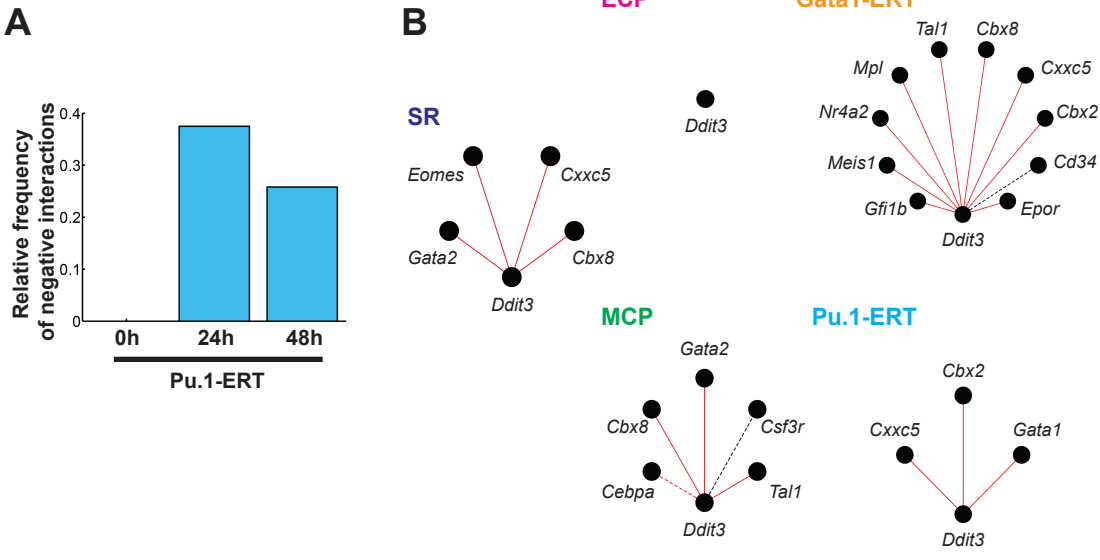
Note: Control gene expression assays are highlighted in gray.

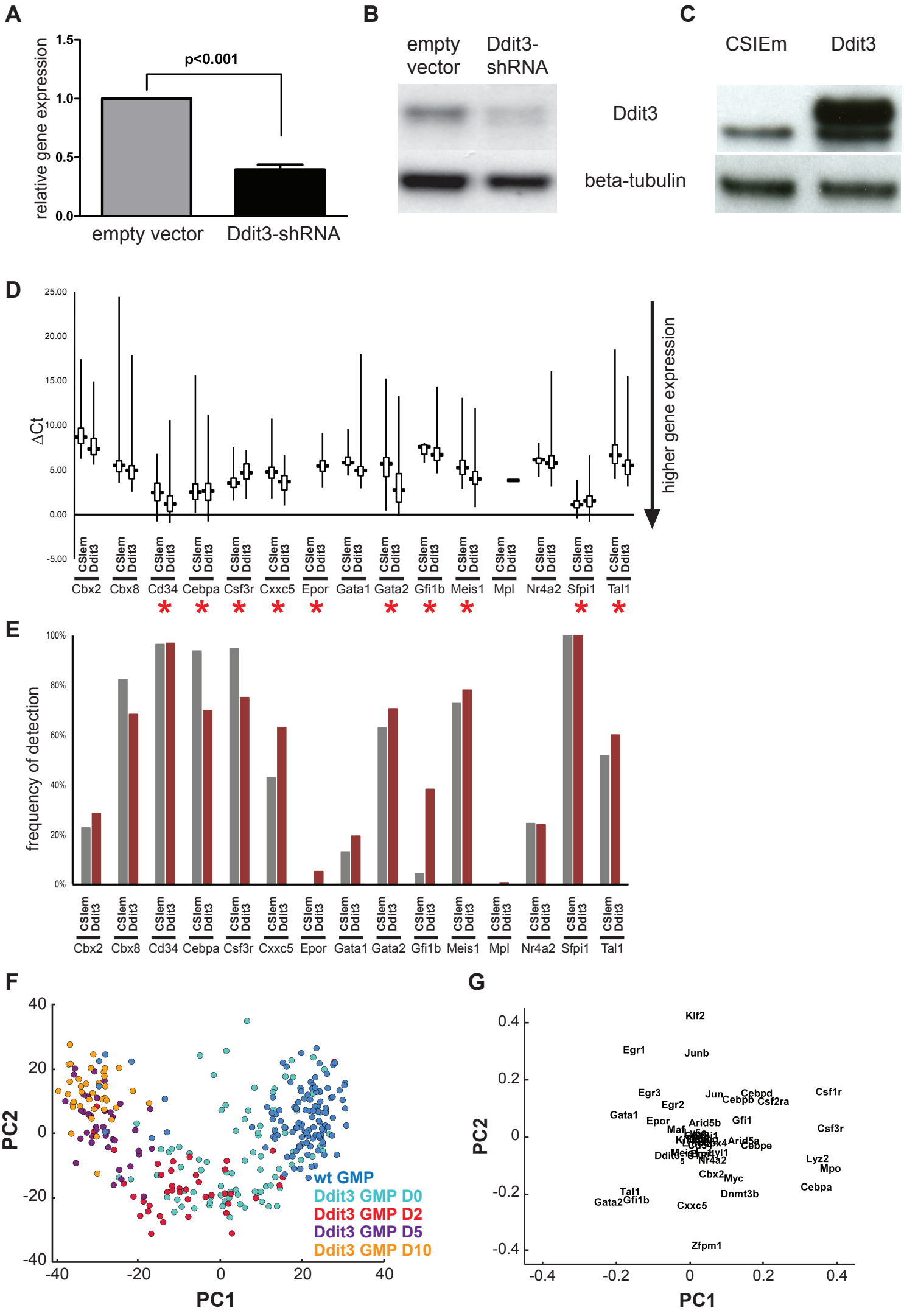
Microarray analysis and RNA-sequencing

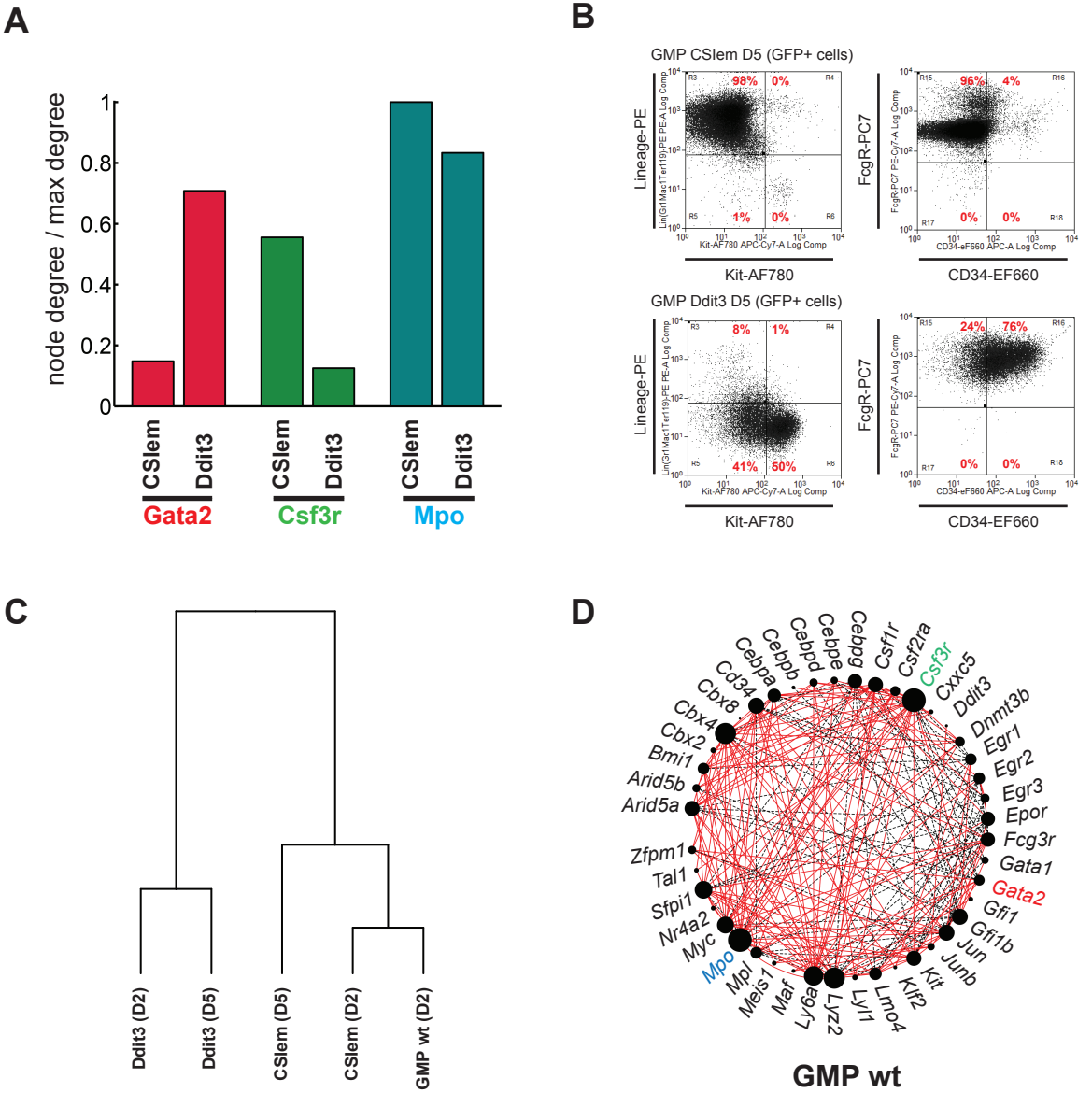
RNA was prepared from Trizol (Life Technologies) lysates. Microarray analysis was performed on an Agilent platform, as described (May et al., 2013). Differentials between any 2 compartments were calculated using LIMMA (Smyth, 2004); hierarchical clustering was performed in Genesis. For RNA sequencing, 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.