

# Supporting Information Appendix

Hamey et al.

## Supporting Text

**Single-cell gene expression analysis.** Single-cell gene expression analysis was performed as previously described [1, 2]. Expression data were collected using the Fluidigm Data Collection software.  $\Delta Ct$  values were calculated by normalizing the mean expression level to the housekeeping genes *Ubc* and *Polr2a* [3]. All housekeeping genes (*Ubc*, *Polr2a*, *Eif2b1*) along with *Cdkn2a*, *Egfl7*, *Gfi1*, and *Sfpi1* were removed from the data set for downstream analysis. *Cdkn2a* was not expressed in any cells and *Egfl7*, *Gfi1* and *Sfpi1* were removed due to technical issues. Quality control and normalization were performed in R ([www.r-project.org](http://www.r-project.org)). The data collected for FSR-HSC2, MPP and PreMegEs during this investigation were re-analyzed with data from Wilson *et al.* [2]. qRT-PCR gene expression data were also collected for HoxB8-FL cells and processed as described above, assaying the same 48 genes as for the primary bone marrow [2] and 416B [4] cells.

**Chromatin immunoprecipitation sequencing.** ChIP assays were performed as previously described [5] using polyclonal antibodies against GATA2 (Santa Cruz, sc9008x) and control nonspecific rabbit IgG (Sigma, I5006). Samples were amplified using the Illumina TruSeq ChIP Sample Prep Kit and sequenced using the Illumina HiSeq 2500 System. Sequencing reads were mapped to the mm10 mouse reference genome using Bowtie2 [6], converted to a density plot, and displayed as UCSC genome browser custom tracks. Peaks were called using MACS2 Software [7]. Raw and processed ChIP-Seq data have been submitted to the NCBI Gene Expression Omnibus with identifier GSE84328.

**Luciferase assays.** Luciferase and LacZ constructs were generated using standard recombinant DNA techniques. The mm10 mouse genome coordinates of chromosomal regions tested are as follows: chr8:122699004-122701098 for the *Cbfa2t3h* promoter, chr8:122699111-122699377 for the *Cbfa2t3h* min promoter, and chr15:103258245-103258850 for the *Nfe2* Enhancer. Both wild-type and GATA2 mutant constructs were generated, where GATA2 binding sites were fully mutated to prevent any binding activity. Luciferase assays were performed as previously described [8]. Luciferase assays were performed in the 416B cell line for stable transfections.

1. Moignard V *et al.* (2013) Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. *Nature Cell Biology* 15(4):363–372.
2. Wilson NK *et al.* (2015) Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. *Cell Stem Cell* 16(6):712–724.
3. Guo G *et al.* (2010) Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Developmental Cell* 18(4):675 – 685.
4. Schutte J *et al.* (2016) An experimentally validated network of nine haematopoietic transcription factors reveals mechanisms of cell state stability. *Elife* 5:e11469.
5. Wilson NK *et al.* (2010) Combinatorial transcriptional control in blood stem/progenitor cells: Genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell* 7(4):532–544.
6. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with bowtie 2. *Nature Methods* 9(4):357–359.
7. Zhang Y *et al.* (2008) Model-based analysis of chip-seq (MACS). *Genome Biology* 9(9):1–9.
8. Bockamp E *et al.* (1995) Lineage-restricted regulation of the murine SCL/TAL-1 promoter. *Blood* 86(4):1502–1514.

## Supporting Tables

**Table S1.** Boolean rules for expression of each gene in both the MEP and LMPP networks. Network rule columns describe the equally high scoring rules for each gene. Agreement columns show agreement level of the rule with pseudotime input-output pairs. For example, an agreement of 0.98 means that these Boolean rules agreed with 98% of the pseudotime pairs.

Gene	MEP network rules	LMPP network rules	MEP agreement	LMPP agreement
Bptf	Myb OR Gata2 OR Ikzf1 OR Lmo2 Erg OR Ikzf1 Smarcc1 Nfe2	Ikzf1 Nfe2 Lmo2 Erg OR Smarcc1 Myb OR Smarcc1	0.98	0.97
Cbfa2t3h	Nfe2 Myb OR Meis1 OR Ikzf1 Gata2 OR Fli1 OR Gata1 OR Meis1 Ikzf1 OR Fli1 OR Gata2 OR Myb	Fli1 Nfe2 Meis1 Ikzf1	0.95	0.82
Erg	Erg AND Meis1	Bptf Meis1 Fli1	0.72	0.88
Ets1	(Notch AND Tcf7) AND NOT(Etv6)	Ets1 AND Notch	0.77	0.57
Ets2	Smarcc1 OR Gfi1b OR Fli1	Ets2	0.78	0.58
Etv6	Smarcc1	Fli1 Meis1	0.89	0.90
Fli1	Fli1 OR Meis1	Meis1 Nfe2 Runx1 OR Erg OR Cbfa2t3h Smarcc1 OR Etv6 OR Cbfa2t3h	0.91	0.99
Gata1	Gata1	Smarcc1 AND NOT Fli1 Tcf AND NOT Fli1 Tcf7 AND NOT Erg Gfi1b AND NOT Fli1 Gfi1b AND NOT Lyl1 Tcf7 AND NOT Nkx2.3 Tcf7 AND NOT Lyl1 Tcf7 AND NOT Hoxa9 Gata2 AND NOT Fli1 Myb AND NOT Fli1 Tal1 AND NOT Fli1 Hoxa5 AND NOT Fli1 Cbfa2t3h AND NOT Fli1 Gata2 AND NOT Lyl1 Gfi1b AND Tcf7 Hoxa5 AND NOT(Hoxa9 OR Nkx2.3) Gfi1b AND NOT (Hoxa9 OR Nkx2.3) Hoxa5 AND NOT(Erg OR Nkx2.3) Myb AND NOT(Erg OR Hoxa9) Smarcc1 AND NOT(Hoxa9 OR Lyl1) Hoxa5 AND NOT(Erg OR Hoxa9) Cbfa2t3h AND NOT(Erg OR Hoxa9) Tal1 AND NOT(Erg OR Hoxa9) Hoxa5 AND Myb AND NOT Hoxa9 Gata2 AND NOT(Erg OR Hoxa9) Smarcc1 AND NOT(Erg OR Lyl1) Hoxa5 AND NOT(Lyl1 OR Nkx2.3) Myb AND NOT(Lyl1 OR Nkx2.3)	0.88	0.98

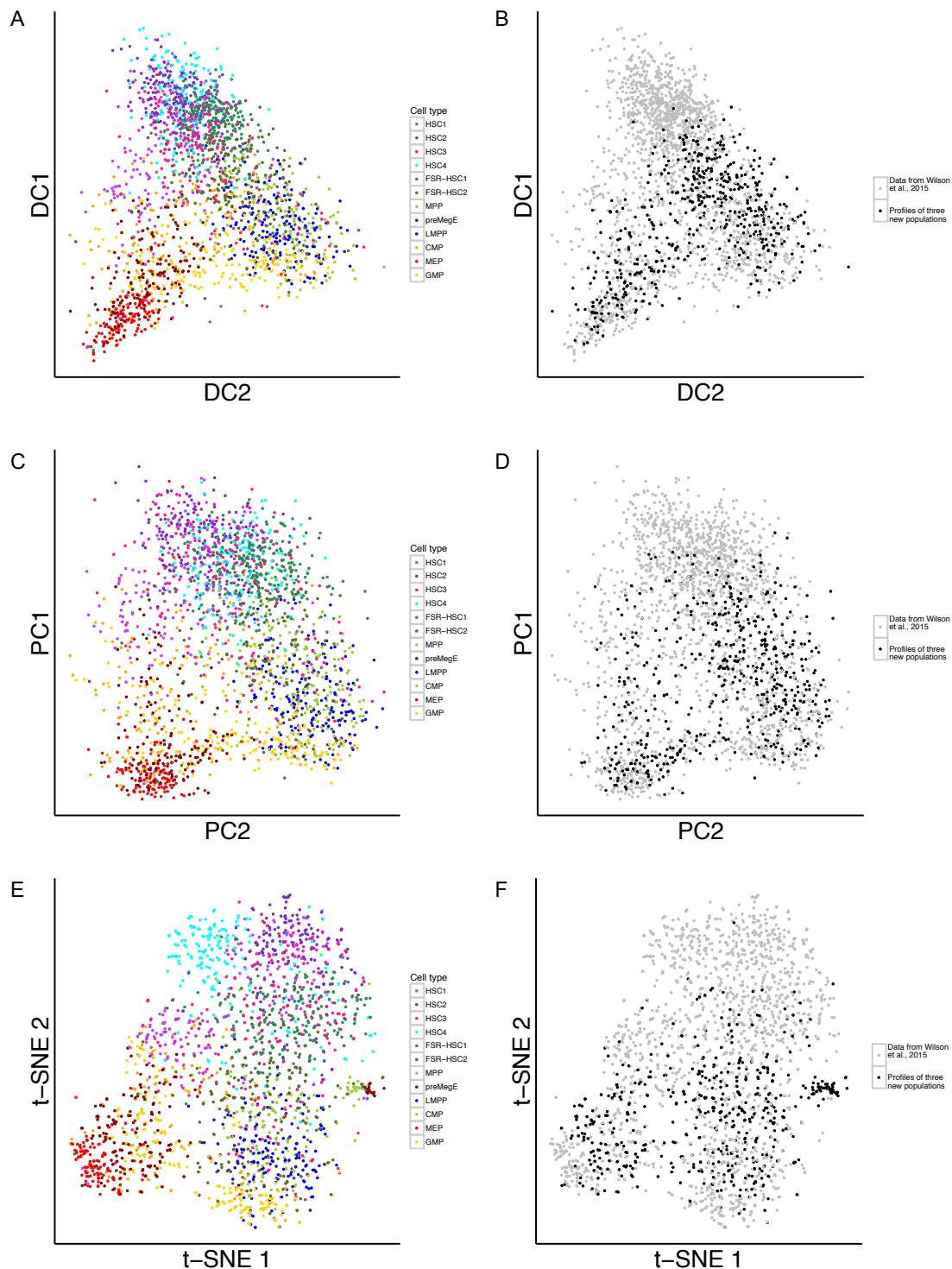
		Smarcc1 AND NOT(Lyl1 OR Nkx2.3) Hoxa5 AND NOT(Erg OR Lyl1) Myb AND NOT(Erg OR Lyl1) Tal1 AND NOT(Lyl1 OR Nkx2.3) Cbfa2t3h AND NOT(Lyl1 OR Nkx2.3) Cbfa2t3h AND NOT(Erg OR Lyl1) Hoxa5 AND NOT(Hoxa9 OR Lyl1) Smarcc1 AND NOT(Erg OR Hoxa9) Gfi1b AND Hoxa5 AND NOT Erg Tal1 AND NOT(Erg OR Lyl1) Cbfa2t3h AND NOT(Hoxa9 OR Lyl1) Tal1 AND NOT(Hoxa9 OR Lyl1)		
Gata2	Nfe2 Bptf Gfi1b OR Meis1 Gata1 OR Meis1 Cbfa2t3h OR Meis1 Cbfa2t3h OR Pbx1	Gata2	0.97	0.79
Gata3	Gata3 AND NOT Myb	Tal1 AND Gata3	0.72	0.65
Gfi1b	Gata1 OR Ets2 OR Gata2 OR Tal1	Gata2 AND Gfi1b AND NOT Notch	0.86	0.74
Hhex	Nfe2	Hhex	0.74	0.56
Hoxa5	Tcf7	Prdm16 AND Tcf7	0.87	0.85
Hoxa9	Ikzf1 AND Meis1	Meis1 Ikzf1 Nkx2.3 OR Ets1 OR Lyl1 OR Hoxa5	0.65	0.87
Hoxb4	Tcf7	Tcf7	0.84	0.80
Ikzf1	Smarcc1 Bptf Cbfa2t3h OR Hoxa9	Bptf Hoxa9 OR Smarcc1 OR Myb OR Cbfa2t3h	0.93	0.96
Ldb1	Smarcc1 Myb OR Lmo2 OR Ikzf1	Ikzf1 Lmo2 Myb OR Smarcc1	0.87	0.811068702
Lmo2	Bptf Nfe2 Lyl1 Ldb1 OR Meis1	Bptf Meis1 Nfe2 Lyl1 OR Notch Ldb1 OR Lyl1 Lyl1 OR Tal1 Lyl1 OR Nkx2.3	0.96	0.98
Lyl1	Smarcc1 Nfe2 Myb OR Hoxa9 OR Lmo2	Nfe2 Lmo2 Erg OR Smarcc1 OR Hoxa9 Myb OR Smarcc1 Erg OR Lmo2	0.95	0.98
Meis1	Meis1 OR Erg	Lmo2 Nfe2 Fli1 Nkx2.3 OR Runx1 Cbfa2t3h OR Etv6 Erg OR Hoxa9 Cbfa2t3h OR Erg OR Runx1	0.88	0.99
Mitf	Fli1 AND Mitf	Mitf	0.67	0.66
Myb	(Gata1 OR Runx1) AND NOT (Gata3 AND Nkx2.3)	Myb OR Ldb1 AND NOT(Gata3 AND Prdm16)	0.83	0.69
Nfe2	Bptf Lyl1	Bptf Fli1	0.99	0.99

	Fli1 OR Myb OR Lmo2 OR Meis1 Cbfa2t3h OR Lmo2 OR Gata2 Gata2 OR Myb Cbfa2t3h OR Fli1 Hhex OR Lmo2 Cbfa2t3h OR Meis1	Lmo2 Meis1 Cbfa2t3h OR Lyl1 OR Myb OR Fli1		
Nkx2.3	Nkx2.3 AND NOT Gata1	Lmo2 Meis1 Hoxa9 OR Lyl1	0.85	0.77
Notch	Tcf7  Ets1 AND NOT Gata2 Ets1 AND NOT Nfe2 Lmo2 AND NOT Nfe2 Lmo2 AND NOT(Gata2 OR Gfi1b)	Lmo2 AND NOT(Gata2 AND Gfi1b)	0.84	0.66
Pbx1	Meis1	Gata2 OR Runx1	0.75	0.71
Prdm16	Fli1 AND NOT Myb	Hoxa5 AND NOT Myb	0.75	0.71
Runx1	Fli1 OR Myb OR Meis1	Fli1 Meis1	0.90	0.88
Smarcc1	Bptf Lyl1 Etv6 OR Ldb1 OR Ikzf1 OR Fli1 Fli1 OR Myb OR Ikzf1 OR Ets2 Ets2 OR Etv6 Fli1 OR Gata1	Ikzf1 Fli1 Bptf Lyl1 OR Etv6 OR Ldb1 Lyl1 OR Myb OR Ets2	0.99	0.98
Tal1	Lmo2 OR Gata1 OR Gfi1b OR Gata2	Gata2 OR Tal1	0.85	0.64
Tcf7	Notch AND NOT Ikzf1 Hoxa5 AND NOT Ikzf1 Nkx2.3 AND NOT Ikzf1 Gata3 AND NOT Ikzf1 Hoxb4 AND NOT Ikzf1 Ets1 AND NOT Ikzf1 Gata1 AND NOT Ikzf1 Gata1 AND Hoxa5	Hoxb4 AND NOT Ikzf1 Hoxa5 AND NOT Ikzf1 Gata1 AND NOT Ikzf1 Notch AND NOT Ikzf1 Nkx2.3 AND NOT Ikzf1 Ets1 AND NOT Ikzf1 Runx1 AND NOT Ikzf1 Gata3 AND NOT Ikzf1	0.97	0.97

**Table S2. Antibody list. Antibodies used to sort MPP (Lin<sup>-</sup>c-Kit<sup>+</sup>Sca1<sup>+</sup>IL-7Ra<sup>-</sup>CD34<sup>-</sup>Flt3<sup>+</sup>), preMegE (Lin<sup>-</sup>c-Kit<sup>+</sup>CD16/32<sup>lo/-</sup>CD41<sup>-</sup>CD150<sup>+</sup>CD105<sup>lo/-</sup>) and FSR-HSC2 (Lin<sup>-</sup>c-Kit<sup>+</sup>Sca1<sup>+</sup>IL-7Ra<sup>-</sup>CD34<sup>-</sup>Flt3<sup>-</sup>) cells.**

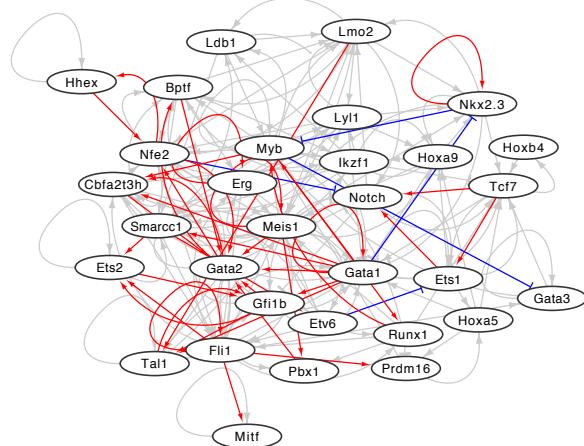
Fluorophore	Antibody	Company	Cat. No.	Clone No.	Used for sorting
BV510	Lineage cocktail	StemCell Tech	19756C.1		All
APC-Cy7	c-kit	Biolegend	105826	2B8	All
Pacific blue	Sca1	Biolegend	122520	E13-161.7	All
7AAD	7aad	BD Pharmingen	51-6898E		All
BV510	IL-7Ra	BD Pharmingen	555288	B12-1	MPP, FSR-HSC2
FITC	CD34	BD Biosciences	560238	RAM34	MPP, FSR-HSC2
PE	CD16/32 (Fc $\gamma$ R)	Biolegend	101308	93	PreMegE, FSR-HSC2
PE-Cy7	CD150	Biolegend	115914	TC15-12F12.2	PreMegE, FSR-HSC2
PE	Flt3	eBioscience	12-1351-83	A2F10	MPP
FITC	CD41	BD Pharmingen	553848	MWReg30	PreMegE
APC	CD105	Biolegend	120414	MJ7/18	PreMegE
PE-Cy5	Flt3	eBioscience	15-1351-81	A2F10	FSR-HSC2
APC	CD48	Biolegend	103412	HM48-1	FSR-HSC2

## Supporting Figures

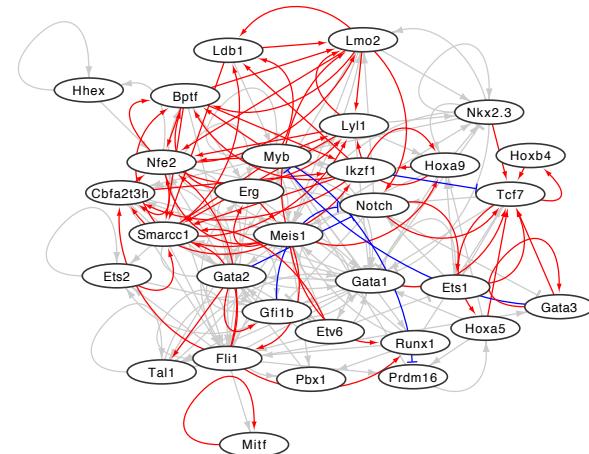


**Fig. S1.** FSR-HSC2, MPP and preMegE qRT-PCR data integrates with original data of (Wilson *et al.*, 2015). (A) Diffusion map dimensionality reduction with all cell types indicated in different colors. (B) The three new populations profiled for this study (FSR-HSC2, MPP and preMegE) are shown in black on the diffusion map, with single-cell profiles from (Wilson *et al.*, 2015) in grey. (C) Principal component analysis of the data set shows a similar structure to diffusion maps. Cell types are colored as in panel (A). (D) Principal component analysis indicating different data sources, as in panel (B). (E) t-distributed stochastic neighbourhood embedding is an alternative dimensionality reduction technique that also shows a similar structure to the diffusion map. Cell types are colored as in panel (A). (F) t-distributed stochastic neighbourhood embedding indicating data different sources, as in panel (B).

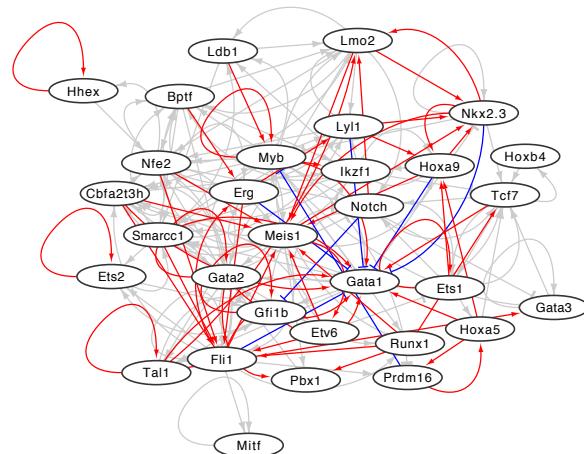
### A Unique to MEP network model



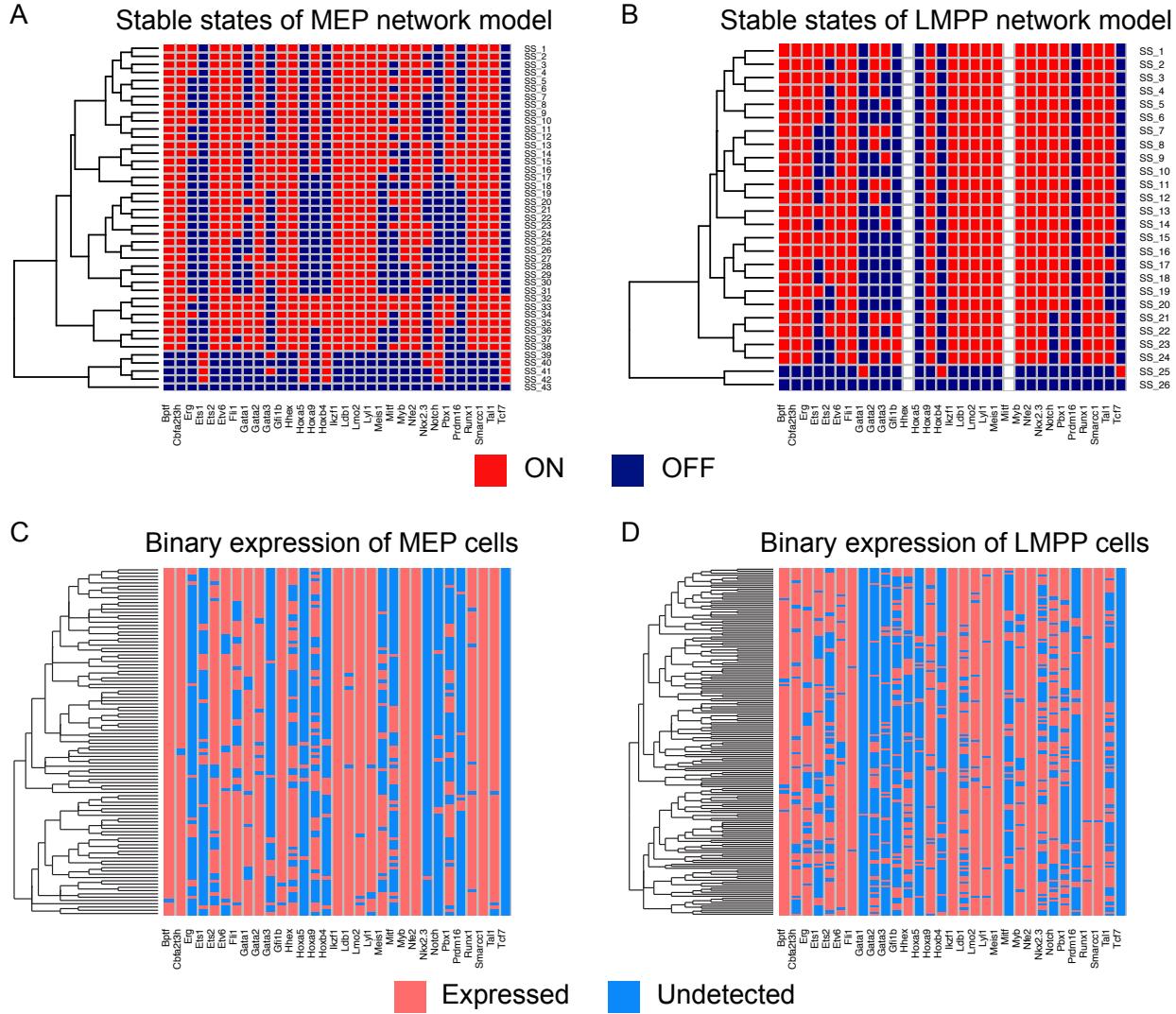
### B Links shared by both models



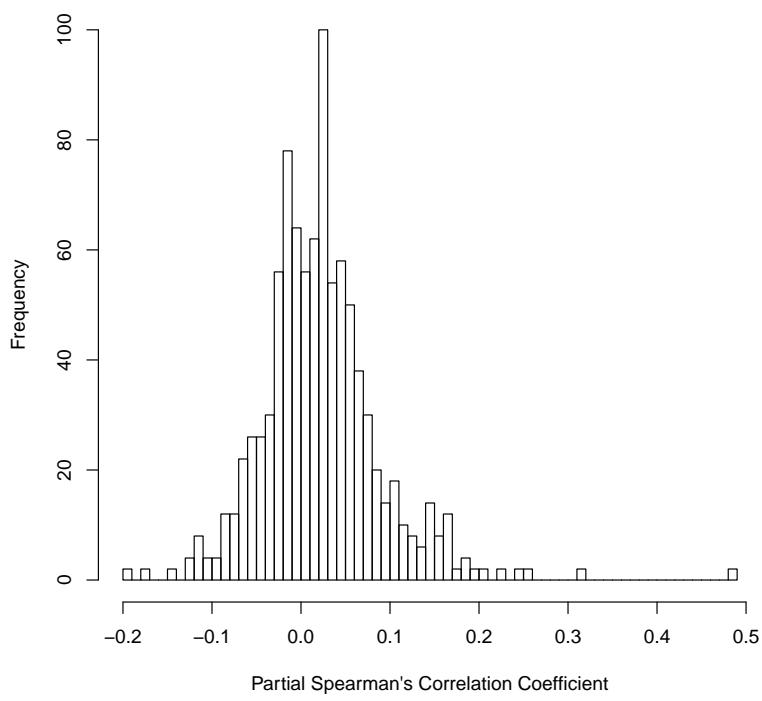
### C Unique to LMPP network model



**Fig. S2.** Visual comparison of regulation in MEP and LMPP network models. (A) Regulatory links unique to the MEP network model and not present in the LMPP network model. Blue flat headed arrows represent repression and red pointed arrows represent activation. Grey links represent regulation present in one or both network models but not unique to the MEP network model. (B) Regulatory links shared between the MEP and LMPP networks. Grey links represent links present in one of the network models but not shared by both. (C) Regulatory links unique to the LMPP network. Grey links represent links present in one or both network models but not unique to the LMPP network model.



**Fig. S3.** Expression of MEP/LMPP network model stable states can be compared to binary *in vivo* data. (A) Stable states of the MEP network model. Genes are colored in red (ON) or blue (OFF). (B) Stable states of the LMPP network model. Genes in white were not connected in the model, and so excluded from stable state analysis. (C) Clustering on qRT-PCR data for primary bone marrow MEP cells converted to binary gene expression. (D) Clustering on qRT-PCR data for primary bone marrow LMPP cells converted to binary gene expression.



**Fig. S4.** Distribution of partial Spearman's correlation coefficients calculated between transcription factor gene pairs by using single-cell qRT-PCR data.