Supplemental data

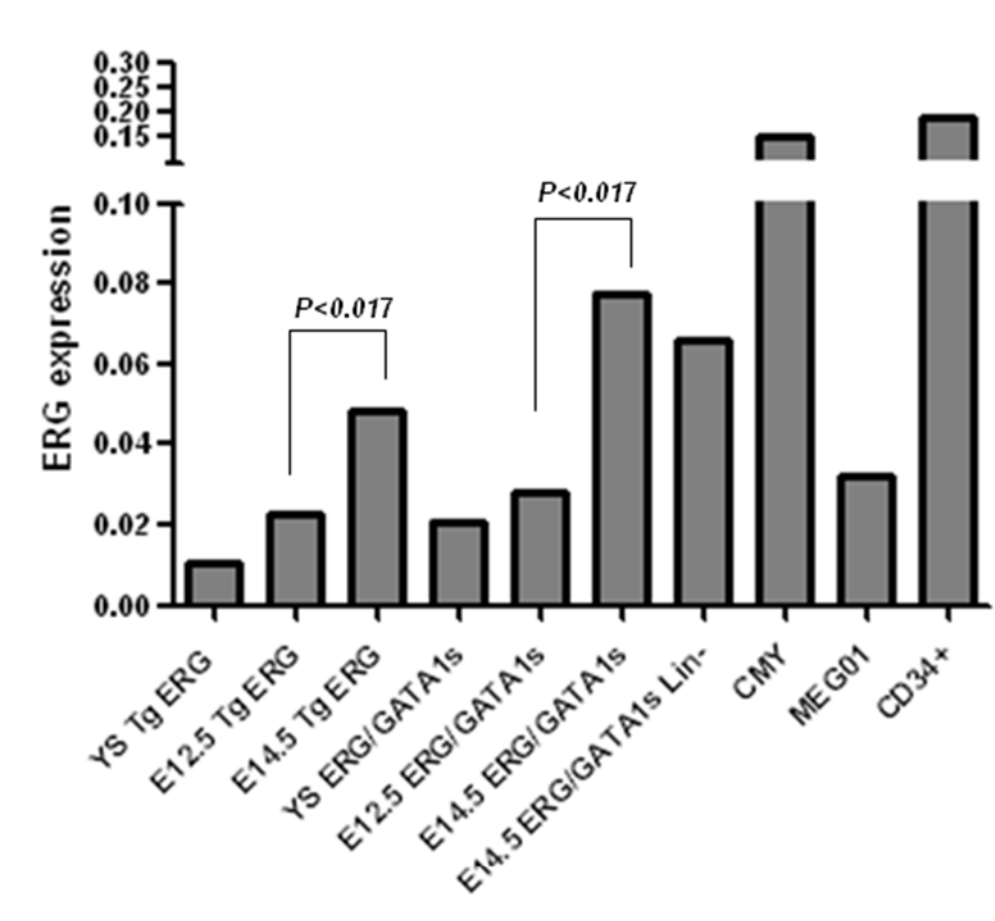


Figure S1: Human ERG3 expression as measured by real time PCR. *ERG* expression was measured from RNA extracted from FL at different embryonic stages, from *ERG/Gata1s* FL Lin- cells isolated using the lineage cell depletion kit (Miltenyi), from human CD34+ cord blood cells and from AMKL cell lines. t-test *p=0.017* between E12.5 and E14.5 FL cells.

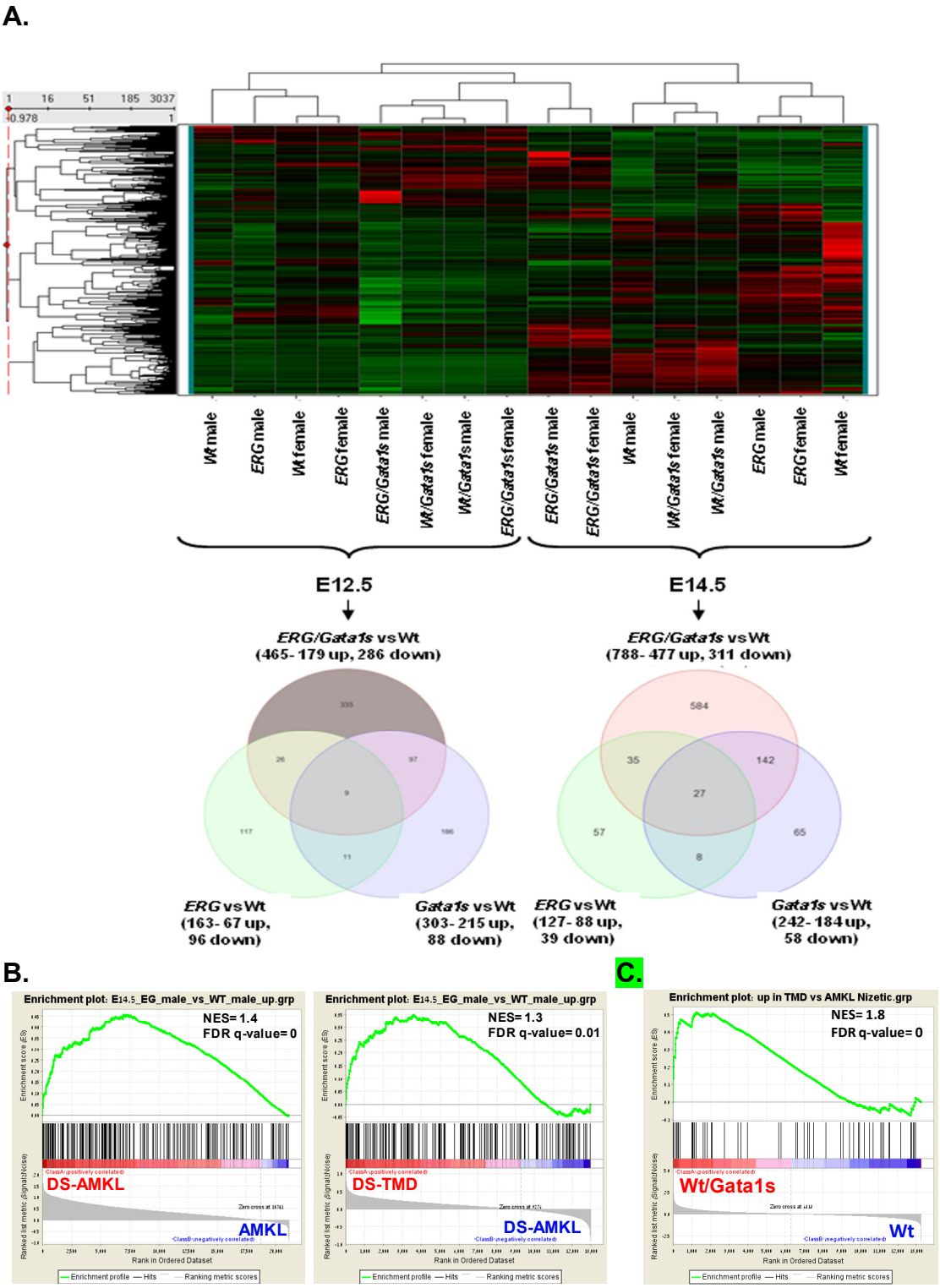


Figure S2: (A) Analysis of gene expression data. Upper panel- Heat map showing unsupervised clustering of the 16 FL samples, using the expression data corresponding to 3037 filtered genes. Lower panel- venn diagrams comparing differentially expressed genes in Tg *ERG*, *Gata1s* and *ERG/Gata1s* male FL cells (compared to the Wt control) in E12.5 and E14.5. **(B)** Gene Set Enrichment Analysis (GSEA) shows significant enrichment of *ERG/Gata1s* gene signature in DS-AMKL when compared to non DS-AMKL (GSE16677) and in DS-TMD when compared to DS-AMKL (GSE4119). NES (normalized enrichment score) and FDR (false detection rate) q values are shown. **(C)** GSEA shows significant enrichment of genes that were unregulated in E14.5 *Wt/Gata1s* FL cells compared to the respective Wt control with genes that were up regulated in human TMD (GSE4119). NES (normalized enrichment score) and FDR (false detection rate) q values are shown.

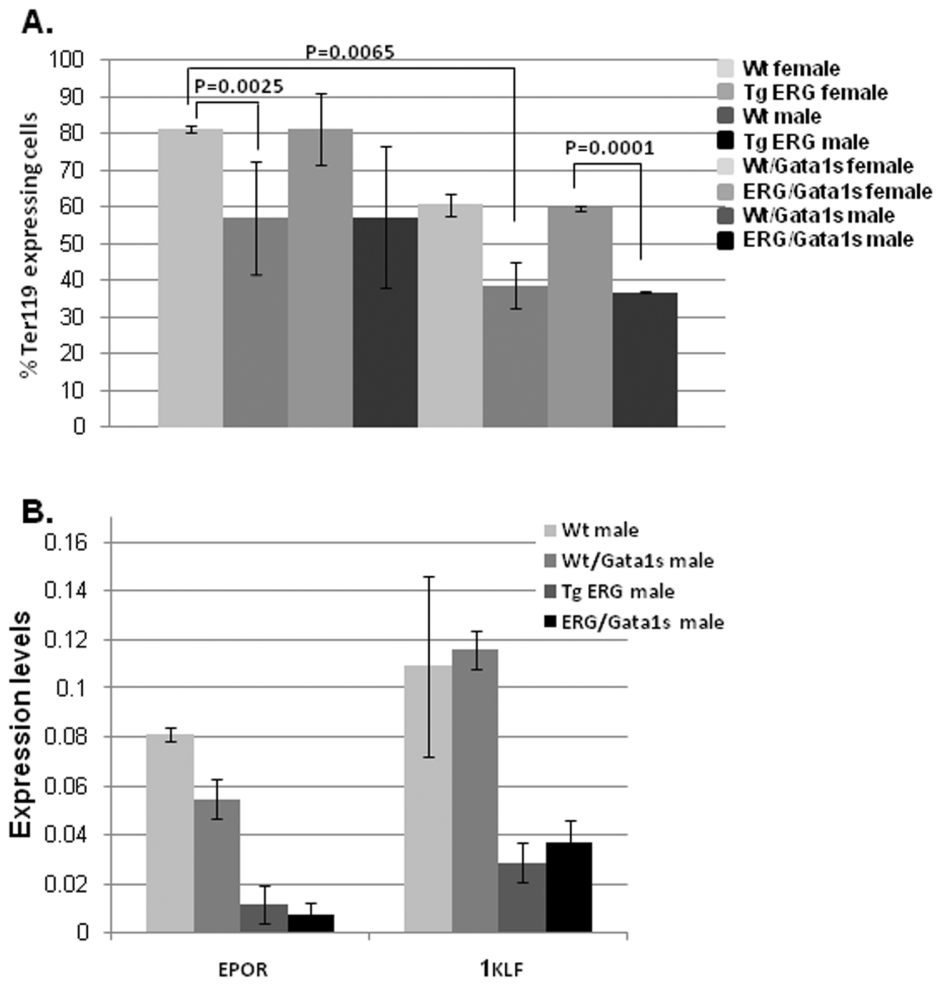
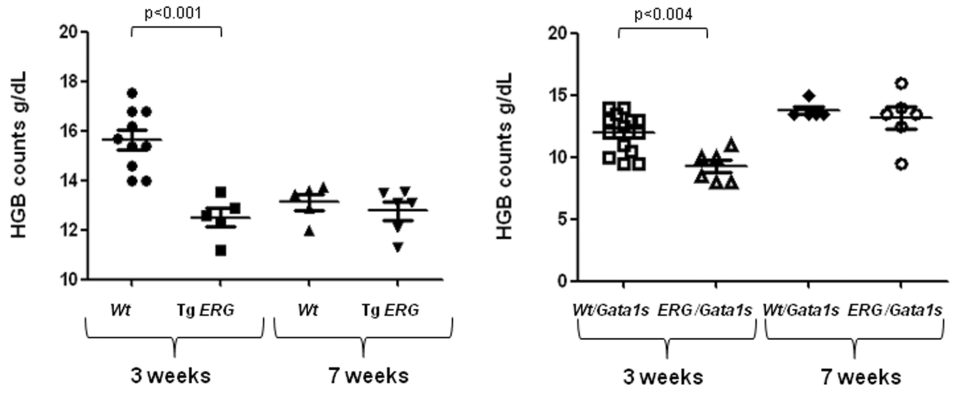
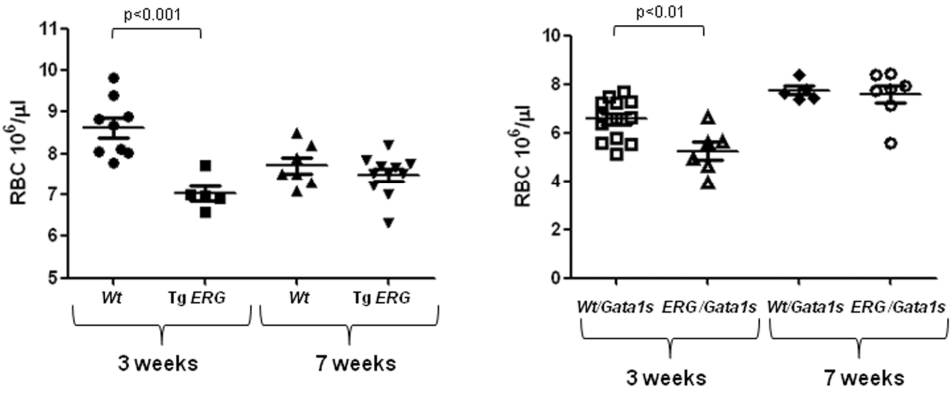


Figure S3: (A) Ter119 expression in E14.5 FL cells. Expression of Ter119 erythroid marker as measured by flow cytometry analysis in FL cells generated from E14.5 females and males *Wt*, *Tg ERG*, *Wt/Gata1s* and *ERG/Gata1s* animals. The bar graph represents an average of at least 4 experiments. Statistical significance was tested using *t-test*. **(B)** Expression of erythroid genes as measured by expression array and confirmed by real time PCR in E14.5 FL cells from *Wt*, *Tg ERG*, *Wt/Gata1s* and *ERG/Gata1s* males.

A.



B.



C.

Genotype	Age (weeks)	WBCs (X10^3 /µl) (n)	RBCs (x10^6 /µl) (n)	HGB (g/dL) (n)	PLTs (X10^3 /µl) (n)
Wt	3	3.31±0.42 (5)	8.62±0.23 (9)	15.6±0. 38 (10)	1020±42.9 (15)
	7	2.34±0.17 (8)	7.71±0.19 (7)	13.1±0. 1 (5)	1180±65.1 (17)
Tg ERG	3	2.81±0.17 (4)	7.03±0.18 (5)	12.5±0. 38 (5)	1510±177 (7)
	7	1.7±0.13 (10)	7.46±0.15 (11)	12.7±0. 36 (6)	1153±55.76 (22)
Wt/Gata1s	3	1.81±0.07 (9)	6.57±0.23 (13)	12±0. 42 (14)	718±29.9 (22)
	7	2.08±0.2 (10)	7.74±0.18 (6)	13.8±0. 3 (5)	822±37 (10)
ERG/Gata1s	3	3.05±0.5 (6)	5.247±0.386 (6)	9.25±0. 51 (6)	929±64.8 (9)
	7	1.92±0.14 (11)	7.59±0.37 (7)	13.2±0. 8 (6)	894±38.8 (10)

Figure S4: Transient decrease in HGB and RBC in Tg*ERG* and *ERG*/*Gata1s* mice. **(A,B)** Blood hemoglobin levels and red blood cell counts (RBC) were retrieved from males and females Tg*ERG* and *ERG/Gata1s* and their *Wt* and *Wt/Gata1s* littermates at the age of 3 and 7 weeks. n= at least 5 for each group. A significant difference between the groups was tested using Mann Whitney test. **(C)** Transient hematological abnormality in ERG/*Gata1s* mice.

A.

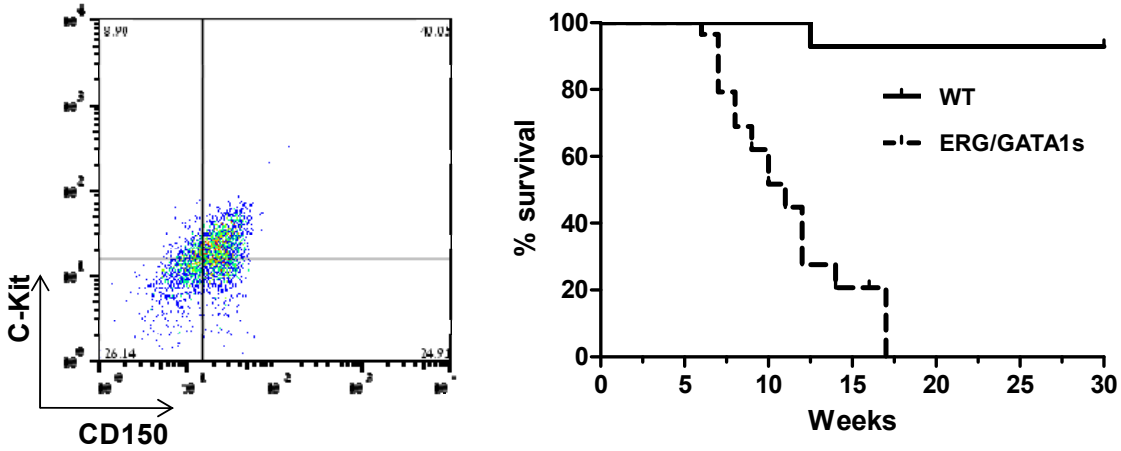


Figure S5: (A) Immunophenotype of cells extracted from the bone marrow of leukemic *ERG/Gata1s* mice shows MEP leukemia **(B)** Kaplan-Meier survival curve of ERG/Gata1s double transgenic mice with mega-erythroid progenitor leukemia.

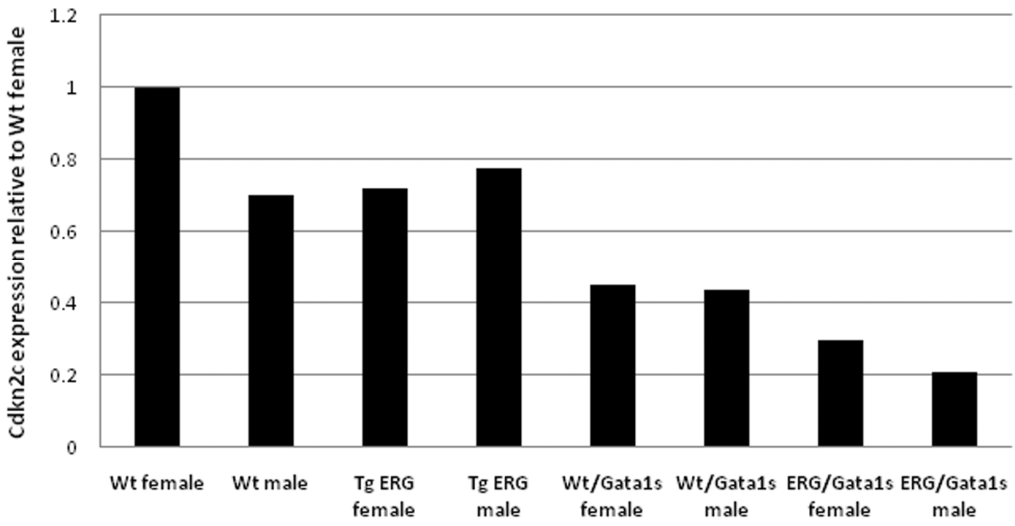


Figure S6: Expression of *Cdkn2c according to expression array of different embryos at E14.5 presented as fold change compared to Wt female.*

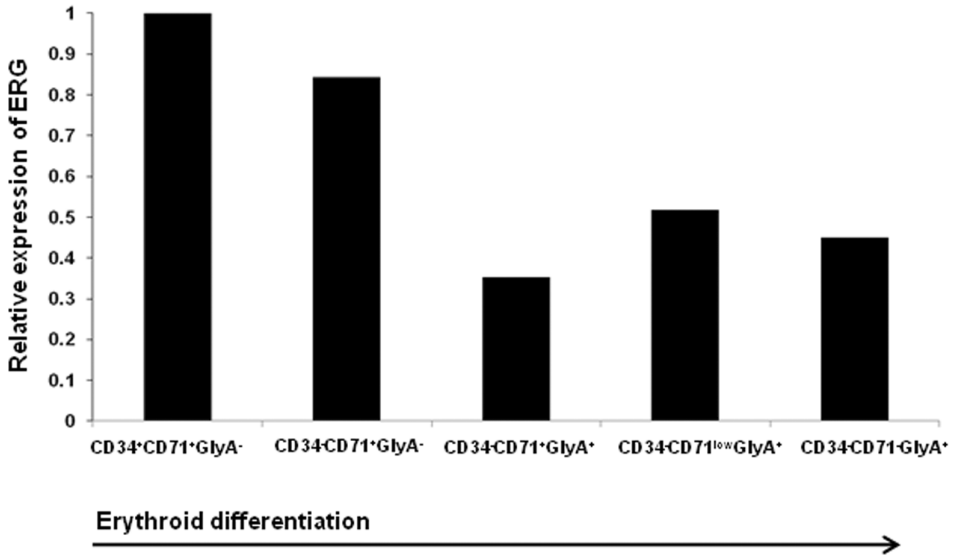


Figure S7: Expression of *ERG* in erythroid cells at different stages of erythroid differentiation. *ERG* expression was obtained from expression data published by Novershtern et al.

Table 1: Genes involved in megakaryocyte development and platelet production

Top up regulated genes common with Wt/Gata1s and ERG/Gata1s	**Relative expression in Wt/Gata1s**	**Relative expression in ERG/Gata1s**
Dock8	3.57607	2.79747
Maff	1.51034	2.05926
Klc2	1.45748	1.49197
Gata2	1.43844	4.3198
Prkar2a	1.20537	1.40004
Vps45	1.20301	2.01897
Dock10	1.16974	1.54411
Kif15	1.13532	1.51522
Irf1	-1.14	2.88013
% of up regulated genes from all the common genes in the two lists	**57%**	**72%**
Average up regulation compare to Wt FL	**1.6**	**2.2**

The gene set was taken from the Broad institute's molecular signature database (MSigDB)
http://www.reactome.org/cgi-bin/eventbrowser_st_id?ST_ID=REACT_24970).
The mice expression data is from male FL gene expression array compared to the expression measured in Wt male FL.

Table 2: Early erythroid Genes

Top genes common with Wt/Gata1s	**Wt/Gata1s**	**Top genes common with ERG/Gata1s**	**ERG/Gata1s**
Ccnb1ip1	-10.8746	Ccnb1ip1	-10.2413
Igll1	-1.85324	Pnpo	-2.97343
Fancl	-1.75817	Aldh1b1	-2.70771
Rps2	-1.60429	Hspd1	-2.47848
Acyp1	-1.56343	Fancl	-2.3628
Ptdss2	-1.52625	Ptdss2	-2.27928
Dut	-1.48067	Igll1	-2.27883
Meis2	-1.45861	Rad51c	-2.25521
Rfc4	-1.43378	Rfc4	-2.21108
Epor	-1.42643	Ndufaf1	-2.19908
Hspd1	-1.41828	Slc36a2	-2.19898
Rragd	-1.39766	Slc39a8	-2.10132
Rad51c	-1.39433	Rragd	-2.09898
Gsta4	-1.38842	Rps2	-2.08482
Pls3	-1.38747	Fam46a	-2.03105
% of down regulated genes from all the common genes in the two lists	**47%**		**51%**
Average down regulation compare to WT FL	**-1.34**		**-1.63**

The gene sets are from Novershtern et al. The mice expression is from male FL expression array compared to the expression measured in Wt male FL.

Table 3: Late Erythroid genes

Top genes common with Wt/Gata1s	Wt/Gata1s	Top genes common with ERG/Gata1s	ERG/Gata1s
Hspa1a	-3.52824	Runx1t1	-4.76153
Runx1t1	-1.68614	Txnrd2	-2.15999
Spsb3	-1.35741	Jhdm1d	-2.11134
Rab11b	-1.29556	Spsb3	-2.06049
Txnrd2	-1.27664	Bbs7	-2.04031
Dsp	-1.23163	Hspa1a	-1.89836
Zc3hav1	-1.20186	Ppp1r15a	-1.62573
Ngrn	-1.19205	Itih2	-1.59876
Jhdm1d	-1.12114	E2f2	-1.55614
Whsc1l1	-1.11692	Gldc	-1.51384
Pcmtd2	-1.06184	Ngrn	-1.44961
Cpm	-1.04604	Rab11b	-1.344
E2f2	-1.03634	Hif1an	-1.26584
Fam134c	-1.01899	Cpm	-1.24262
Polr2k	-1.01383	Whsc1l1	-1.22018
% of down regulated genes from all the common genes in the two lists	**61%**		**79%**
Average down regulation compare WT FL	**-1.3**		**-1.6**

The gene sets are from Novershtern et al. The mice expression is from male FL expression array compared to the expression measured in Wt male FL.

Supplemental methods

Primer sets used for the real time quantitative PCR:

Mpl F: TCAAGAGACCTGCTACCAG;

Mpl R: CACCGAGAGATGGCTCCA;

Pcam1 F: CTCCACTTCTGAACTCCAACAGC;

Pcam1 R: GTATTCTACATCCATGTTCTGG;

Mycn F: GAGAGGATACCTTGAGCGACTC;

Mycn R: GAACGTCTCTTCTCTACGGTG;

Mcl1 F: CTTGTAAGGACGAAACGGGAC;

Mcl1 R: CAGCAGCACATTTCTGATGC;

Epor F: CTCATTCTGGTCCTCATCTCG;

Epor R: GTGGTGAAGAGACCCTCAAACTCGCTCTCTG;

Gata2 F: GCTCTACCACAAGATGAATGGACAG;

Gata2 R: CATAAGGTGGTGGTTGTCGTC;

Ter119 F: CTGTAGGTAACCCAAATCAGC;

Ter119 R: GCAGCAGACACTTCAGTAG;

Klf1 F: GGAAGAGCTACACCAAGAG;

Klf1 R: AGCGAACCTCCAGTCACAG;

Ldb1 F: CAACAGCAAGAAGAAGAGCCCAGC;

Ldb1 R: GTGATGAGCCTCTCGTCCTC;

Myb F: CATTTACCAGGCACACAAGCGTC;

Myb R: CCAGTGGTTCTTGATAGC;

GATA1 F ex.1: CACTAAGGTGGCTGAATCCTC

GATA1 F ex.2: CTTGCCCCAGTTTGTGGATTC

GATA1 R ex.4: CTCCACAGTTCACACACTCTC

ERG F: GAGTACAGACCATGTGCGG

ERG R: AGCCTCTGGAAGTCGTCC

Gene expression profiling

Experiments were performed using Affymetrix Mouse gene 1.0 ST oligonucleotide arrays (Affymetrix, Santa Clara, CA).Total RNA prepared from pulled embryos (n>3), from each sample was used to prepare biotinylated target DNA, according to manufacturer's recommendations. Briefly, 100 ng of Total RNA was used to generate first-strand cDNA by using a T7-random hexamer primer. After second-strand synthesis, in vitro transcription was performed. The resulting cRNA is then used for a second cycle of first-strand cDNA by using a T7-random hexamer primer with UTP resulting in SS DNA used for fragmentation and terminal labeling. The target cDNA generated from each sample was processed as per manufacturer's recommendation using an Affymetrix GeneChip Instrument System. Briefly, spike controls were added to fragmented cDNA before overnight hybridization. Arrays were then washed and stained with streptavidin-phycoerythrin before being scanned on an Affymetrix GeneChip scanner. A complete description of these procedures is available at: https://www.affymetrix.com/support/technical/manuals.affx.

Quality control: The quality and amount of starting RNA was confirmed using an agarose gel or by Bioanalyser (Agilent). After scanning, array images were assessed by eye to confirm scanner alignment and the absence of significant bubbles or scratches on the chip surface. The signals derived from the array were assessed using various quality assessment metrics. Details of quality control measures can be found at: http://media.affymetrix.com/support/technical/whitepapers/exon_gene_arrays_qa_whitepaper.pdf.

Data analysis: Gene level RMA sketch algorithm (Affymetrix Expression Console and Partek Genomics Suite 6.6) was used for crude data generation.

Differentially expressed genes were filtered as changed by at least 2 fold. Genes were filtered and analyzed using unsupervised hierarchal cluster analysis (Spotfire DecisionSite for Functional Genomics; Somerville,MA) to obtain a first assessment of the data. Further processing including functional analysis was performed using Ingenuity pathway analysis software (Ingenuity systems, see Ingenuity networks and function supplemental file).