SUPPLEMENTARY INFORMATION

Supplementary Materials and Methods

Differentiation of ES cells

Blast culture: Flk1+ cells were isolated from embryoid bodies by magnetic cell sorting using MACS columns (Myltenyi Biotech) and an anti-Flk1 antibody and were plated in blast media (IMDM supplemented with 10% FCS, Pen/Strep, 1 mM glutamine, 0.45 mM MTG, 0.18 mg/ml human transferrin, 25 µg/ml Ascorbic acid, 20 % D4T conditioned media, 5 µg/L mVEGF (Peprotech), 10 µg/L mIL-6 (Miltenyi Biotec)) at a concentration of 8-10x10^3/cm^2 on gelatinized tissue culture treated dishes. D4T conditioned media was made by growing confluent D4T cells in fresh IMDM media supplemented with 10 % FCS, Pen/Strep, 1 mM glutamine, 0.15 mM MTG and 50 ng/ml acidic human FGF (R&D Sytems) for 3-4 days.

Differentiation of hemogenic endothelium (only iRunx1 cells). After 2 days of blast culture as described above, cells were trypsinised, MACS sorted for Flk-1+ cells a second time as described before and 2x10^4/cm^2 Flk-1+ cells were plated on gelatin coated tissue culture plates and cultured for 2 days in hemogenic endothelium media (IMDM supplemented with 10 % FCS, Pen/Strep, 1 mM glutamine, 0.15 mM MTG, 0.18 mg/ml human transferrin, 25 µg/ml Ascorbic acid, Oncostatin 10 µg/ml (R&D Systems 495-MO), basic mFGF 1 µg/ml (Peprotech 450-33), mSCF 10 µg/ml (Peprotech 500-P71)). The HE culture was induced on day 1 with 0.3 µg/ml doxycycline (Sigma) for an overnight incubation. For withdrawal experiments induced cells were thoroughly washed with PBS and harvested 24 h thereafter. All cells were checked for purity by flow cytometry using antibodies against c-kit (CD117, BD Pharmingen 553356), CD41 (eBioscience 25-0411), Tie2 (eBioscience 12-5987).

Differentiation of hematopoietic precursors: Hematopoietic precursor cells were isolated after 3-4 days blast culture by gently flushing non-adherent cells off the adherent endothelial cells. Progenitors were incubated with a biotinylated c-kit antibody as described for the Flk1 sort. Separated cells (c-kit+ cells) were checked for purity by flow cytometry. The vast majority of cells were c-kit+, CD41+ and Tie2-.

Macrophages: Macrophages were produced by plating c-kit+, CD41+, Tie2- cells onto low adherence plates at 1-2X10^7 cells per 15 cm dish in IMDM supplemented with 10% FCS, Pen/Strep, 1 mM glutamine, 0.15 mM MTG, 10 ng/ml murine M-CSF (Miltenyi Biotec) and 5% IL3 conditioned media (Faust et al, 1994). Macrophages were harvested after 3-5 days by washing the cells with PBS to remove non-adherent cells and then trypsinising the adherent layer. Cells were checked for purity by flow cytometry and showed high expression of CD11b (eBioscience 12-0112-82) and F4/80 (eBioscience 17-4801-80).

Cell sorting and flow cytometry

To gain pure hematopoietic cell populations of all different stages during the Brachyury wild type blast culture we harvested the whole culture on day 3 and stained the cells for Tie2, c-kit and CD41. Cells were separated according to their surface markers on a Cytomation MoFlo into two pools of hemogenic endothelium (Tie2+, c-kit+, CD41- and Tie2+, c-kit+, CD41+) and progenitors (Tie2-, c-kit+, CD41+).
In order to check the purity of all MACS sorted cells. flow cytometry was performed on BD LSRII or Cyan (Beckman Coulter, Brea, CA) flow cytometers and samples were analysed using the FlowJo (TreeStar Inc) or Summit (Dako, Beckman Coulter) software. Relevant isotype controls were performed for all antibodies

Chromatin immunoprecipitation

Crosslinked chromatin was sonicated using a Bioruptor water bath in immunoprecipitation buffer I (25 mM Tris 1 M, pH 8.0, 150 mM NaCl, 2 mM EDTA, pH 8.0, 1 % TritonX100 and 0.25 % SDS). After centrifugation the sheared 0.5-2 kb chromatin fragments (1-2 x 10^6 cells) were diluted with 2 volume immunoprecipitation buffer II (25 mM Tris, pH 8.0, 150mM NaCl, 2 mM EDTA, pH 8.0, 1 % TritonX100, 7.5 % glycerol) and precipitation was carried out for 2 hours at 4 °C using 2 µg specific antibody coupled to 15 µl protein G Dynabeads (Dynal). Beads were washed with low salt buffer (20 mM Tris 1 M, pH 8.0, 150 mM NaCl, 2 mM EDTA, pH 8.0, 1 % TritonX100, 0.1 % SDS), high salt buffer (20 mM Tris, pH 8.0, 500 mM NaCl, 2 mM EDTA, pH 8.0, 1 % TritonX100, 0.1 % SDS), LiCl buffer (10 mM Tris, pH 8.0, 250 mM lithium chloride, 1 mM EDTA, pH 8.0, 0.5 % NP40, 0.5 % sodium-deoxycholate) and TE pH 8.0 containing 50 mM sodium chloride. The immune complexes were eluted in 100 µl elution buffer (100 mM $NaHCO_3$, 1 % SDS) and, after

adding 5 µl 5M sodium chloride and proteinase K the crosslinks were reversed at 65 °C overnight. DNA was extracted by using the Ampure PCR purification kit and analysed by qRT-PCR.

Sequencing library preparation

Libraries of DNA fragments from chromatin immunoprecipitation or DNase I treatment were prepared from approximately 10 ng of DNA. Firstly, overhangs were repaired by treatment of sample material with T4 DNA polymerase, T4 PNK and Klenow DNA polymerase (all enzymes obtained from New England Biolabs UK) in a reaction also containing 50 mM Tris-HCl,10 mM $MgCl_2$, 10 mM Dithiothreitol, 0.4 mM dNTPs and 1 mM ATP. Samples were purified after each step using Qiagen MinElute columns (according to the manufacturer's guidelines). 'A' bases were added to 3' ends of fragments using Klenow Fragment (3´- 5´ exo-), allowing for subsequent ligation of adapter oligonucleotides (Illumina part #1000521) using Quick T4 DNA ligase. After a further column clean up to remove excess adaptors, fragments were amplified in an 18-cycle PCR reaction using adapter-specific primers (5'-CAAGCAGAAGACGGCATACGAGCTCTTCCGATC*T and 5'-AATGATACGGCGACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T). The libraries were purified and adapter dimers removed by running the PCR products on 2% agarose gels and excising gel slices corresponding to fragments approximately 200-300 bp in size, which were then extracted using the Qiagen gel extraction kit. Libraries were validated using quantitative PCR for known targets, and quality assessed by running 1 µl of each sample on an Agilent Technologies 2100 Bioanalyser. Once prepared, DNA libraries were subject to massively parallel DNA sequencing on an Illumina Genome Analyzer.

Data analysis:

Raw sequence data in fastq format from ChIP-seq and input material were mapped to the mm9 mouse genome build using the Bowtie program (Langmead et al., 2009). The reads in the resulting alignment files (sam format) were converted to 200bp and used to generate density maps in bigwig format.

The reads were converted to eland format for peak finding using a combination of MACS v1.4.0beta (Zhang et al., 2008) and Peakseq (Rozowsky et al., 2009). Various stringencies were used to generate 400bp peak coordinates in bed format that fit well with the density map when visually inspected using the University of California Santa Cruz (UCSC) Genome Browser, at known regulatory elements.

Peaks were allocated to genes if located in either their promoters or introns, or if intergenic, to the nearest gene located within 100 kb. The peak lists and gene lists for different factors were then intersected using the online tools Venny (http://bioinfogp.cnb.csic.es/tools/venny/index.html) (Oliveros, 2007) and Galaxy (http://main.g2.bx.psu.edu/) (Goecks et al., 2010) to find groups of overlapping peaks or genes between factors.

De novo motif discovery was performed using MEME v4.6.1 (Bailey and Elkan, 1994) and HOMER v3.2 (Heinz et al., 2010). Motifs were compared with the JASPAR_CORE database (Bryne et al., 2008) and matches to consensus sequences were determined using TFBSSearch (Chapman et al., 2004). The following sequences were used: TKNNGNAAK (CEBP), TGYGGT (RUNX), GGAAR (Fli1), CANNTG (SCL/TAL1).

The GSEA software (Subramanian et al., 2005) was used to perform gene set enrichment analysis on groups of genes against haematopoietic expression data (Chambers et al., 2007), looking for enrichment in HSCs versus all other cell types in the dataset.

Analysis of H3K9Ac profiles was performed using Seqminer (Ye et al., 2011). Groups of peak regions were used as reference coordinates against all aligned reads for the acetylation and factors of interest. Mean read density profiles were produced for 2 clusters, generated using the default K-means method.

The peaks were intersected with other published TF peaks in various cell types using the HemoChIP resource (Hannah et al., 2011), to generate a binary matrix showing regions of combinatorial binding which was then hierarchically clustered using Pearson's correlation coefficient between the binding patterns of the samples to produce a heat-map.

The heat-maps that show tag counts distribution in SCL/TAL1, FLI-1 and C/EBPβ peaks in hemogenic endothelium (-DOX) cells (Figure 2 D) were generated by first dividing the

binding-sites of SCL/TAL1, FLI-1 and C/EBPβ into promoter/ intragenic and intergenic genomic regions, then RNA POL II and H3K9Ac levels for each genomic division were counted using +/- 1kb window from SCL (left), FLI-1 (middle) and C/EBPβ (right) peak summit. The tag counts were normalised using quantile normalisation and displayed on the heatmap from high to low according to POL II level. POLII level is well correlated with H3K9Ac level where the average correlation coefficients for FLI-1, SCL and C/EBPβ are 0.60, 0.64 and 0.78 respectively.

The microarray gene expression scanned images were analyzed with Feature Extraction Software 10.7.1.1 (Agilent) (protocol GE1_107_Sep09, Grid: 028005_D_F_20100614 and platform Agilent SurePrint G3 Mouse GE 8x60K). The raw data output by Feature Extraction Software was analysed using the LIMMA R package (Smyth, 2005) with quantile normalisation and background subtraction. Contrast matrix and eBays function were used and p value <= 0.01 was applied.

To examine how factor binding, Pol II occupancy and histone acetylation correlated with gene expression we associated the binding sites of SCL/TAL1, FLI-1 and C/EBPβ peaks in hemogenic endothelium (-DOX) cells with their genomic location by first dividing the binding-sites of SCL/TAL1, FLI-1 and C-EBP/β into promoter/ intragenic and intergenic genomic regions, then RNA Pol II and H3K9Ac levels for each genomic division were counted using +/- 1kb window from FLI-1, SCL/TAL1 and C/EBPβ peak summit. The tag counts were normalised using quantile normalisation and displayed on the heat-map from high to low according to their Pol II occupancy (Figure 2 D and Supplementary Figure 3 E). We then subdivided this gene set into expressed and non-expressed genes according to their absolute expression levels and plotted the expression levels next to the ChIP data. Genes were considered as differentially expressed genes such that the log expression value is greater than or equal to 6. This showed that Pol II level is highly correlated with H3K9Ac level while there was a week correlation between Pol II occupancy, histone acetylation at the SCL/TAL1, FLI-1 and C/EBPβ binding sites and gene expression levels. KEGG pathway analysis was done using DAVID online tools.

To correlate factor binding of FLI-1, SCL/TAL1 and C/EBPβ with expression levels, we first divide factor binding genes lists into seven subsets of genes binding each combination of one, two or all three transcription factors (subsets G1- G7), each subset contains n_i genes

where i = 1, …,7. We then calculated how many high expression genes we would expect on the gene subset (n_i) by chance using {BY_CHANCE = (Nh*n_i)/(Nh+Nl)} where Nh is the high expression set (expressed genes log2 ≥ 6), Nl is the low expression set (non-expressed genes log2 ≤ 6) and n_i is the gene subset that binds each combination of one, two or all three transcription factors. Then we calculated the log enrichment using log(n_i∈Nh/ BY_CHANCE) and the hypergeometric distribution were used for calculating the significance enrichment of high expression genes in each subset of genes. The results indicated that in all seven cases the actual number of high expression genes (n_i in Nh) on the gene subsets is greater than the number by chance meaning that the subsets are overrepresented in high expression genes compared to low expression genes. For example in the case of the subset G1 (genes that bind all three factors) we found 2300 genes in total of which 2002 are high expression, and by chance we would expect only 1446 high expression genes. This over-representation of 556 high expression genes is highly statistically significant (p=1.22e-163). We then examined the relation between the total number of binding sites per gene in gene subset G1 (binding all three factors) and the expression level for each gene and found very little correlation between the number of binding sites and gene expression levels with a Pearson correlation coefficient equal to 0.1.

The overlapping peaks were calculated using BedTools (Quinlan and Hall, Bioinformatics 2010) and were decided to be overlapping if they have at least one genomic position (base pair) in common between them. The frequency plots of these peak populations, as a function of their peak-center distances (in Figure 8B) were generated using R (http://www.r-project.org). The Z scores for calculating the significance of peak overlaps between RUNX1 and SCL/TAL1+DOX or FLI-1+DOX unique peaks were obtained by bootstrapping (100,000 iterations). A random peak set (74516 peaks) was obtained from the union of the H3K9 acetylation peaks before and after RUNX1 induction. For bootstrapping, peak sets of 400bps width and a population equal to the RUNX1 peak population (15669) were randomly obtained from this random set. The mean (μ) and the standard deviation (σ) for the total overlap between the unique peaks from SCL/TAL1+DOX or FLI1+DOX and the random peak set were calculated and compared with the actual overlap (X) between the unique and RUNX1 peaks, to obtain the Z scores ($z = \frac{X-\mu}{\sigma}$). Both the SCL/TAL+DOX and the FLI1+DOX unique peaks' overlap with RUNX1 peaks were found to be significant, with Z scores of 107.6 and 34.9 respectively.

Genes with at least two fold-changes in expression (either up or down) were selected and correlated with FLI-1 SCL/TAL1 and C/EBPβ bound genes in hemogenic endothelium (-DOX) cells and RUNX1 bound genes in hemogenic endothelium (+DOX) cells. The resulting correlated genes were subdivided into RUNX1 bound and non-bound genes and the hypergeometric distribution was used for calculating the significance enrichment of direct RUNX1 target genes. Both up and down regulated RUNX1 target gene enrichment were found to be significant, with p-values of 2e-46 and 2e-4 respectively (Supplementary Figure 6 G, H).

Supplementary References:

Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009 ;10(3):R25

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137

Rozowsky J, Euskirchen G, Auerbach RK, Zhang ZD, Gibson T, Bjornson R, Carriero N, Snyder M, Gerstein MB. PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. Nat Biotechnol. 2009 Jan;27(1):66-75

Oliveros JC: VENNY: An interactive tool for comparing lists with Venn diagrams. [http://bioinfogp.cnb.csic.es/tools/venny/index.html] 2007.

Goecks J, Nekrutenko A, Taylor J; Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. 2010;11(8):R86.

Timothy L. Bailey and Charles Elkan, "Fitting a mixture model by expectation maximization to discover motifs in biopolymers", Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, pp. 28-36, AAAI Press, Menlo Park, California, 1994

Heinz S, Benner C, Spann N, Bertolino E et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol Cell 2010 May 28;38(4):576-589

Bryne JC, Valen E, Tang MH, Marstrand T, Winther O, da Piedade I, Krogh A, Lenhard B, Sandelin A. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. Nucleic Acids Res. 2008 Jan;36 (Database issue):D102-6.

Chapman MA, Donaldson IJ, Gilbert J, Grafham D, Rogers J, Green AR, Göttgens B. Analysis of multiple genomic sequence alignments: a web resource, online tools, and lessons learned from analysis of mammalian SCL loci. Genome Res. 2004 Feb;14(2):313-8.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545-50.

Chambers SM, Boles NC, Lin KY, Tierney MP, Bowman TV, Bradfute SB, Chen AJ, Merchant AA, Sirin O, Weksberg DC, Merchant MG, Fisk CJ, Shaw CA, Goodell MA. Hematopoietic fingerprints: an expression database of stem cells and their progeny. Cell Stem Cell. 2007 Nov;1(5):578-91.

Ye T, Krebs AR, Choukrallah MA, Keime C, Plewniak F, Davidson I, Tora L. seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res. 2011 Mar;39(6):e35.

Hannah R, Joshi A, Wilson NK, Kinston S, Göttgens B. A compendium of genome-wide hematopoietic transcription factor maps supports the identification of gene regulatory control mechanisms. Exp Hematol. 2011 May;39(5):531-41.

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010 March;26(6):841-42.

Smyth GK. Limma: linear models for microarray data. In: Bioinformatics and Computational Biology Solutions using R and Bioconductor, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (eds.), Springer, New York, (2005):397-420

LEGENDS TO SUPPLEMENTARY FIGURES:

Supplementary Figure 1

Comparative gene expression analysis of isolated cell populations during blast culture differentiation of ES cells. A. FACS profiles of cell populations developing during a 4-day blast culture of Brachyury wild-type ES cells. The staining of the surface markers Tie2, c-kit and CD41 nicely illustrates the kinetics and emergence of hematopoietic progenitors through the stage of the hemogenic endothelium. The blast culture on day 3 harbours enough cells be able to isolate cells representing the two hemogenic endothelium stages as well as c-kit+ precursor cells. By MoFlo cell sorting for all three markers pure populations of early HE (Tie2+ckit+CD41-), late HE (Tie2+ckit+CD41+) and progenitors (Tie2-ckit+CD41+) were obtained. All cell populations were re-analyzed after the sort and respective FACS profiles demonstrate their high purity. **B.** mRNA expression analyses of transcription factor genes in different purified cell populations representing different stages of hematopoietic development as described in Figure 1. Experiments were carried out in at least triplicates where STDEV is applied, otherwise the average of two biological duplicates plus indication of relative values is shown and expression levels were normalised to *Gapdh* expression.

Supplementary Figure 2

Analysis of genome-wide binding of SCL/TAL1, C/EBPβ and FLI1. A. FACS profiles demonstrating the isolation of pure hemogenic endothelium cells from iRUNX1 ES cells using a two-step protocol. Left panel: The protocol differs from the wild-type differentiation procedure shown in Figure 1 and Supplementary Figure 1 in that a second MACS sort is performed at day 2 of blast culture. Since iRUNX1 cells are blocked in differentiation at the Tie2+, c-kit+ and CD41- stage, Flk1-expressing hemogenic endothelial cells (Flk1+, usually 15-20%) can be easily enriched at this time-point. FACS profiles (bottom left panels) depicting the expression of Tie2, c-kit and CD41 surface markers show the HE identity of the Flk1+ population in contrast to the discarded major Flk1- population. Bottom right panel: After two days hemogenic culture Flk1 harvested for each experiment still show expression of respective surface markers and morphology indicative for early hemogenic endothelium (Tie2+ckit+CD41-) and the same level of CD41 expression, confirming the block in differentiation in the absence of RUNX1 (-DOX). FACS profiles after the overnight induction with doxycyline (+DOX) show, that these cells are released from their arrest into the next differentiation state of HE (Tie2+ckit+CD41+). **B**. UCSC

genome browser screenshots show examples of two genes (*Etv6* and *Runx1*) bound by SCL/TAL1, C/EBPβ, FLI1 and RNA-Pol II, as well as acetylated histones. **C.** Unsupervised hierarchical clustering of binding sequences for all three transcription factors confirming the structure of the Venn diagram shown in Figure 2 B. Binding sites 10 kb up- and downstream of the peak centre are shown. **D.** Manual ChIP analysis of transcription factor binding as well as RNA Polymerase II and H3 K9Ac marks in *Runx1-/-* cells. Relative enrichment is shown on the *Pu.1* gene (-14kb 5' and 3' URE, -12kb enhancer, -5kb control and promoter), *Runx1* promoter and *Csf1r* (promoter and enhancer FIRE), confirming that the binding of transcription factors and RNA Pol II at this loci is not due to leakiness of the inducible iRUNX1 system. **E**. Manual validation of ChIP-sequencing data at selected single genes. Bars show the chromatin enrichment of each transcription factor in the prepared libraries relative to the standard curve. Chr2 serves as a control amplicon.

Supplementary Figure 3

Classification of genes bound by SCL/TAL1, C/EBPβ and FLI1. A. Gene set enrichment analysis (GSEA) of each transcription factor alone and in combination with others s described in the legend to Figure 3. **B**. A gene ontology analysis of the 2400 genes which are occupied by all three factors reveal genes that are involved in the development of the hematopoietic system. **C.** Additional gene annotation was performed with the Database for Annotation, Visualization and Integrated Discovery (DAVID) and visualized as KEGG pathway. Many genes bound by the three transcription factors are significantly involved in focal adhesion processes, receptor interaction, cell motility and cell cycle (p-value: 5.4e-15). **D.** Manual expression analysis of depicted genes from these pathways in purified cell populations as described in Figure 1, showing up- and down-regulation during the differentiation of the hematopoietic system. **E.** Correlation of RNA-Pol II occupancy and histone acetylation at promoters, intergenic and intragenic regions of genes bound by FLI-1, SCL/TAL1 or C/EBPβ with the expression of associated genes. Upper panels: expressed genes (log2 ≥ 6) and lower panels: low/non-expressed genes (log2 < 6). **F.** Correlation between factor binding and gene expression. Left panel: Log enrichment of high expression genes for gene subsets (G1-G7) binding each combination of one, two or all three of FLI-1, SCL/TAL1 and C-EBPβ. Right panel: hypergeometric distribution was used to calculate the significance enrichment of genes expressed higher than log2 in each subset where p-values are annotated on top of each bar.

Supplementary Figure 4

A. Sequence of the mutation of the FLI1/PU.1 binding site knocked into the *Pu.*1 locus. **B.** The hemogenic endothelium (HE) cultures of $Pu.1^{+/ki}$ and $Pu.1^{ki/ki}$ cells show no visible effect of the FLI1 binding site mutation at the stage of progenitor emergence from the hemogenic endothelium. Images show the heterozygous $Pu.1^{+/ki}$ culture (left) and the homozygous $Pu.1^{ki/ki}$ culture (right) at day 3 of HE culture.

Supplementary Figure 5

SCL/TAL1 and FLI1 binding patterns change during development. A. UCSC genome browser screen shots show additional genes with changing transcription factor binding patterns during the development from the un-induced HE to c-kit+ progenitors (the *Mecom* locus containing the hematopoietic regulator gene *Evi1* and the *Ets1* locus). Note that at *Mecom*, SCL/TAL1 and FLI-1 co-localize in the HE, which demonstrates the reproducibility of our assays and that these peaks disappear after *Evi1* down-regulation in c-kit+ cells. **B.** Manual validation of ChIP-sequencing results. Bars show the chromatin enrichment of each transcription factor in the prepared libraries relative to the standard curve. Chr2 served as a control amplicon.

C. Unbiased analysis for enriched motifs associated with unique and shared SCL/TAL1 and FLI-1 binding sites in c-kit+ cells. **D.** Frequency of presence of the respective consensus motif in SCL/TAL1 and FLI-1 binding sites in c-kit+ cells. **E.** Venn diagrams visualising the overlap of genes bound by either FLI1 or SCL/TAL1 in the un-induced HE (HE) and c-kit+ progenitors (c-kit+). **F.** Gene set enrichment analysis of FLI-1 and SCL/TAL1 on its own and on the overlapping genes shows that genes bound by both factors have the highest score for HSC specific gene enrichment.

Supplementary Figure 6

Genome-wide RUNX1 binding analysis. A. (Upper panel): experimental strategy for RUNX1 induction experiments, (lower panel) FACS plots demonstrating the up-regulation of CD41 expression after RUNX1 induction. **B.** 4-way Venn diagram analysing the intersection between RUNX1, SCL/TAL1, FLI-1 and C/EBPβ peaks in the HE. RUNX1 binding profiles were generated by intersecting two independent biological experiments with a peak overlap of 77%. Only 158 peaks were found binding the four factors. **C.** Manual validation of ChIP-sequencing results for the indicated factors. The PCR values were calculated relative to the standard curve. **D.** Manual ChIP validation of RUNX1

independent factor binding at selected single target genes with and without induction (*Pu.1*: -14kb enhancer 5' and 3', -12kb enhancer, -5kb control and promoter; *Runx1*: distal promoter; *Csf1r:* promoter and FIRE enhancer). The values represent the mean values of two independent biological replicates analyzed in duplicate and are normalized to input and the internal Chr2 control. **E.** Unbiased motif search within RUNX1 bound sequences (left panel), 61% of all analysed sequences contain the RUNX1 motif (right panel). **F.** Schematic outline of differentiation of the inducible Runx1 ES cell line (iRunx1), which is blocked at the first stage of the hemogenic endothelium. An overnight induction of RUNX1 rescues hematopoietic gene expression. **G.** Analysis of expression microarray experiments. Proportion of RUNX1 responsive genes subdivided into RUNX1 bound and non-bound genes. **H.** Venn-diagram depicting the association of RUNX1-responsive genes with their factor-binding pattern. **I.** Response of hematopoietic regulator genes carrying (black) or not carrying (red) RUNX1 binding sites to RUNX1 induction. mRNA expression levels were analysed in undifferentiated ES cells, Flk1+ hemangioblast, uninduced (HE-) and the induced (HE+) hemogenic endothelium. Experiments were done in at least three biological replicates.

Supplementary Figure 7:

RUNX1 dependent transcription factor binding. A. Screen shots taken from the UCSC genome browser give more examples for both RUNX1 dependent and independent binding of transcription factors to the individual genes. **B.** Manual validation of ChIP-seq libraries showing alterations in binding patterns before and after induction (HE-DOX/+DOX) and in c-kit+ progenitors. PCR signals are relative to the standard curve and the Chr2 amplicon served as control. **C.** Overlap of genes bound by SCL/TAL1 (left) and FLI-1 (right) before and after induction of RUNX1 and the overlap between the hemogenic endothelium and c-kit+ progenitors (clockwise).

Supplementary Figure 8:

A. Venn diagram visualizing the peak populations analysed in Figure 8. Note that peaks shared between induced and un-induced cells were not taken into consideration. **B.** Surface marker profile of iRUNX1 cells before RUNX1 induction and after RUNX1 induction and withdrawal after 24h as indicated. **C.** Co-immunoprecipitation pulldown experiments in induced hemogenic endothelium (iRUNX1 cells) indicate a direct interaction of RUNX1 with FLI-1 (left), C/EBPβ (middle) or SCL/TAL1 (right). The pulldowns were performed with a specific antibody against each transcription factor.

RUNX1 interaction was shown by Western blotting and detection by using an antibody against the HA tag of iRUNX1. The control IPs using an IgG antibody did not show any nonspecific signal.

Tables:

Target	Antibody	Cell Type	Number of total reads	Number of mapped reads	No of peaks
RUNX1	HA (Sigma H6908 lot 038K4750)	HE -DOX	8,481,817	1,963,077	1002
RUNX1	HA (Sigma H6908 lot 038K4750)	HE +DOX	29,637,698	21267861 (intersect)	15669
SCL/TAL1	sc-12984 lot H3110	HE -DOX	13,471,876	12,049,319	10617
SCL/TAL1	sc-12984 lot H3110	HE +DOX	36,491,406	33,269,844	10395
FLI1	sc-356 lot E2110	HE - DOX	17,390,877	12,716,291	8507
FLI1	sc-356 lot E2110	HE + DOX	37,502,843	35,807,486	5577
C/EBPβ	sc-150 lot F0711	HE - DOX	12,279,078	10,536,410	10193
Pol II	ab817 (lot GR21668-1)	HE - DOX	34,264,433	33,075,290	19712
H3 K9Ac	ab4441 (lot GR25683-1)	HE - DOX	33,305,541	32,339,562	n.a.
H3 K9Ac	ab4441 (lot GR25683-1)	HE + DOX	34,340,007	32,621,745	n.a.
SCL/TAL1	sc-12984 lot H3110	c-kit+	39,282,952	34,256,409	8594
FLI1	sc-356 lot E2110	c-kit+	40,441,647	35,075,626	7734
Input		HE + DOX	4,753,442	3,707,811	n.a.

Supplementary Table 1: Antibodies and numbers of reads and peaks for each ChIP-seq experiment. Note that H3K9Ac peaks have not been determined (n.a.) as they extend over extended regions covering several nucleosomes.

Supplementary Table 2:

List of used primers:

Gene	Primer sequence expression	Reference
Runx1	GCAGGCAACGATGAAAACTACTC CAAACCTGAGGTCGTTGAATCTC	
Fli1	TCGTGAGGACTGGTCTGTATGG GCTGTTGTCGCACCTCAGTTAC	
SCL/Tal	CAACAACAACCGGGTGAAGA ATTCTGCTGCCGCCATCGTT	
Elf1	CAGCACCATGCAGGAGGAA TTGCCAGGAGACACCACTACTG	
Etv6	TGCTCTATGAACTCCTTCAGCATA TCTGGACACAGTTATCTTCGTCAT	Li *et al* Nature Immunology 12 129-136
Bmi1	GCCTAAGGAAGAGGTGAATGATAA ATCTGGAAAGTATTGGGTATGTCC	Li *et al* Nature Immunology 12 129-136
Etv2	CACCACGCCCTCCAAATC AGCAGCTCCAGGAGGAATTG	
Evi1	TGCCCTGGAGATGAGCTGTAA GATCTAGAGCAGAAAGGCCAGATT	
Notch1	GATGGCCTCAATGGGTACAAG TCGTTGTTGTTGATGTCACAGT	Primer bank 13177625a1
Pbx	GACAACTCAGTGGAGCATTCC CATCACGTGGGTGGTGAAT	Li *et al* Nature Immunology 12 129-136
Sox17	CTAAGCAAGATGCTAGGCAAG TACTTGTAGTTGGGGTGGTCC	
Sox7	GAACCCGGACCTGCACAAC GCTCTGCCTCATCCACATAGG	
Erg	AGGAGCTGTGCAAGATGACA TCAGATGTGGAAGGGGAGTC	Li *et al* Nature Immunology 12 129-136
JunB	TCACGACGACTCTTACGCAG CCTTGAGACCCCGATAGGGA	Primer bank 6680512a1
Lmo2	GAGAGACTATCTCAGGCTTTTTGG TTGAAACACTCCAGGTGATACACT	Li *et al* Nature Immunology 12 129-136
c-Myb	CAGAAGAGGAGGACAGAATCATTT TTCCAGTGGTTCTTGATAGCATTA	Li *et al* Nature Immunology 12 129-136
Stat4	GGAAGCACTCAGTAAGATGACG CTGCTGCCGCTTTTTCCAG	Primer bank 6755670a3
Pu1	CCATAGCGATCACTACTGGGATTT TGTGAAGTGGTTCTCAGGGAAGT	
NFE2	TCCTCAGCAGAACAGGAACAG GGCTCAAAAGATGTCTCACTTGG	Primer bank 33469085a1

Gfi1	GTGAGCCTGGAGCAACACAA CTCTTGAAGCTCTTGCCACAGA	
C/EBPβ	GTTTCGGGACTTGATGCAATC CGCAGGAACATCTTTAAGGTGAT	
C/EBPα	GCAGGAGGAAGATACAGGAAGCT ACACCTAAGTCCCTCCCCTCTAAA	
Ets1	TCCTATCAGCTCGGAAGAACTC TCTTGCTTGATGGCAAAGTAGTC	Primer bank 21307663a1
Csf1r	CTTTGGTCTGGGCAAAGAAGAT CAGGGCCTCCTTCTCATCAG	
Gata2	AGAACCGGCCGCTCATC TCGTCTGACAATTTGCACAACAG	
Itga2b	AGAGGGCCATTCCTGTCTG GTCGATTCCGCTTGAAGAAG	
AI467606	GCGAAATCCAAGATTGAAGC TCTGGGAGACTCAGTGTGGA	
Gene	**Primer sequence ChIP**	**Reference**
Notch1	TGGGCAGACAGGAACTTTGAC CCGCGCTGACCAGCTT	
Etv6	TCGCGCCCTGCACTTC GCTTTAAGAGAATCCACCAGGAAA	
Nfe2	TGTTTGGCAACAATGCTTGTG CAACCCACCTCCACTACGTATGT	
Erg	CCCTGAGCTGATGCCAAAAG TTGAGTGACCGACATCTGCAA	
Vav2	CAGCTGGCTTCCTTCACAATC CACCCCCAGGTCCCCTAA	
Asb6	CGCCGGTGAGCAATCCT TGACTCACATGCTCCCTCAAAG	
Evi1	ACCTGTCACCAGCCAGTCTTG CAGGGATGAGAAAGAAAGCAGAGA	
Meis1	GGCAGCAGGAAGGAATTTGA TTGTGGGAAACTTGCTTTATGGT	
Itga2b	CTGTGAAAGTCCAGCCACCAT AGTGAGCCAGGCAGCGAAT	
Ets1	AGAGGAAACCATCCAGCTAATCAG CGGAGCAAAGATTTCTGTTTTATTT	
Sox17	GCCAGGCTTCTAGTCCAGATAC GAGCTCACGACGTTATCAAAGC	
Gfi -35kb	CCACAAACAGAACAGCTGGA CCACATGACCTCATGAATGC	

AI467606	CACAGTCGCCCAGCAACA TGGATTTCGCCGTTTTCAG	
-14kb 5'URE	GCCCAGGCTAGGGAAGTTTG GAGAGCAGAGCACTTCATGGCTA	
-14kb 3'URE	GCTGTTGGCGTTTTGCAAT GGCCGGTGCCTGAGAAA	
-12 kb enhancer	CAGAGCGGGCAGCTATTTAC GGGGAAGGAAAGAGGAAGC	
-5kb control	GGCACATGGTAGAAGAGAACAAACT TTGTGTTTTCACTGTGTGTCTGATG	
Pu.1 promoter	GTAGCGCAAGAGATTTATGCAAAC GCACAAGTTCCTGATTTTATCGAA	
Runx1 promoter	CAGCAGGCAGGACGAATCA CGCCTATGCTGTGGGTTGA	
Csf1r promoter	CTGCTGCTGGCCACAGTTT CAGCGATGCCCTCTTTGC	
FIRE	GGTGCCAGCAATGTGTTTCC GCATTTTTCCCAGGCTTGAA	
Chr2	AGGGATGCCCATGCAGTCT CCTGTCATCAGTCCATTCTCCAT	

The specificity of all Real-Time-PCR primers has been checked by standard procedures including a melting curve plot at the end of the reaction showing one specific product.

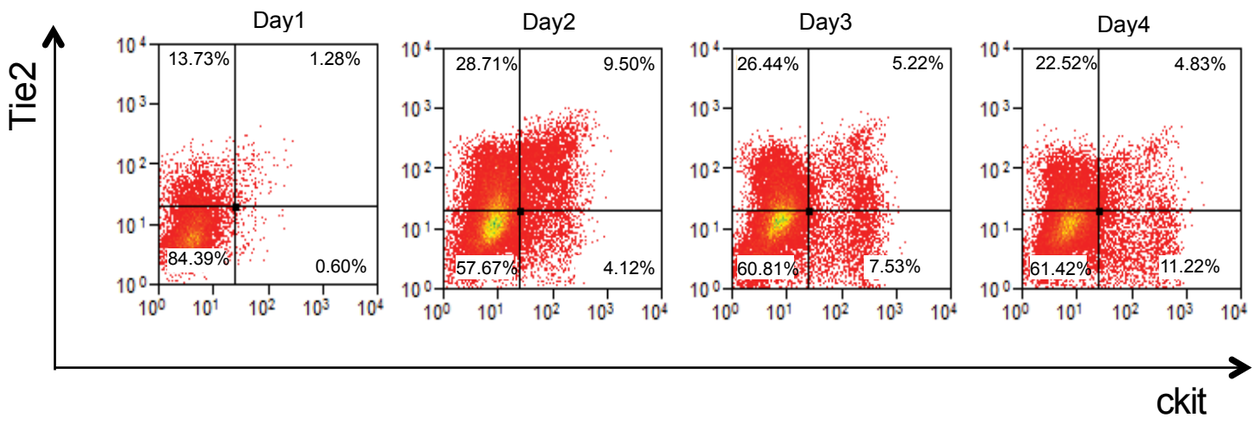
Supplementary Table 3:

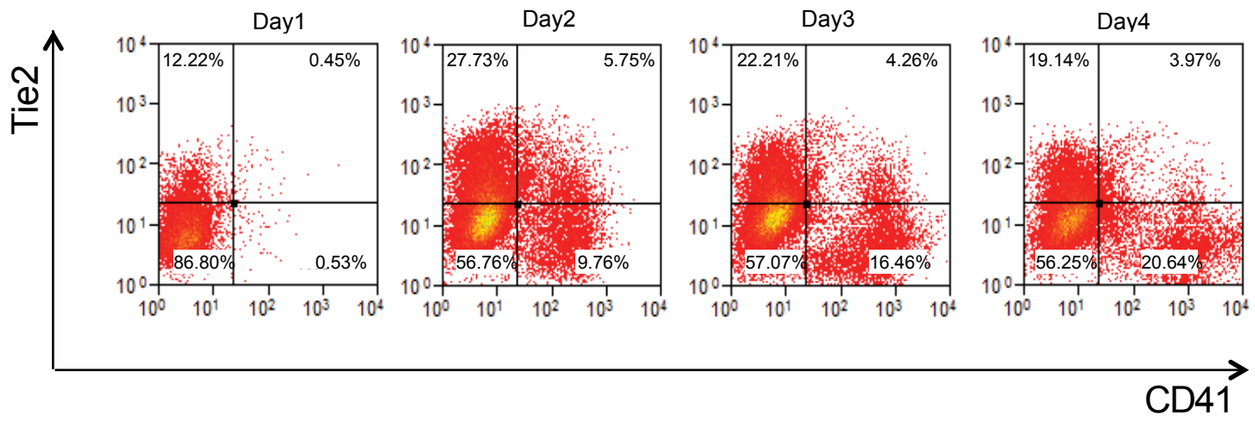
Genes bound by SCL/TAL1, FLI-1, C/EBPβ in the HE and their expression levels

Genes bound by SCL/TAL1 and FLI1 in c-kit+ cells

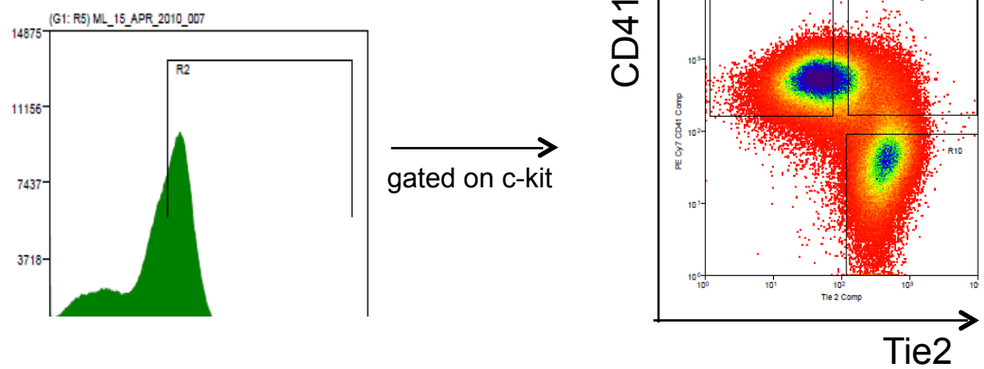
A

Blast culture differentiation of Brachyury ES cells:

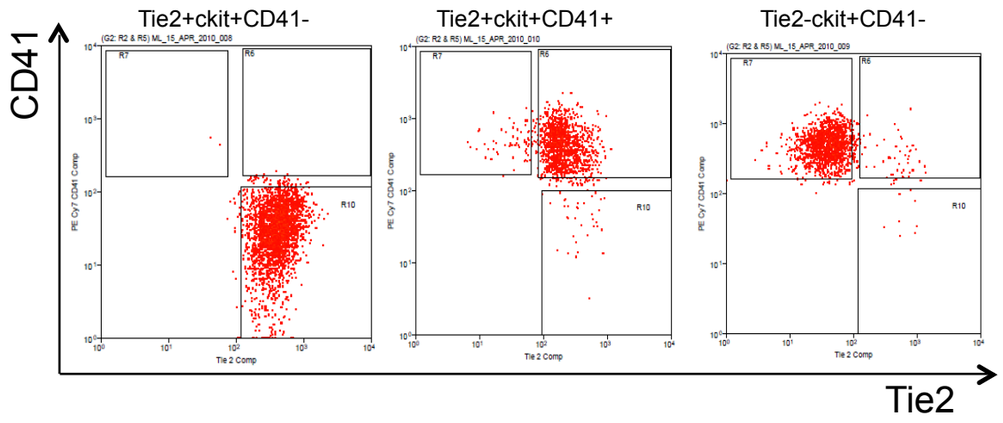




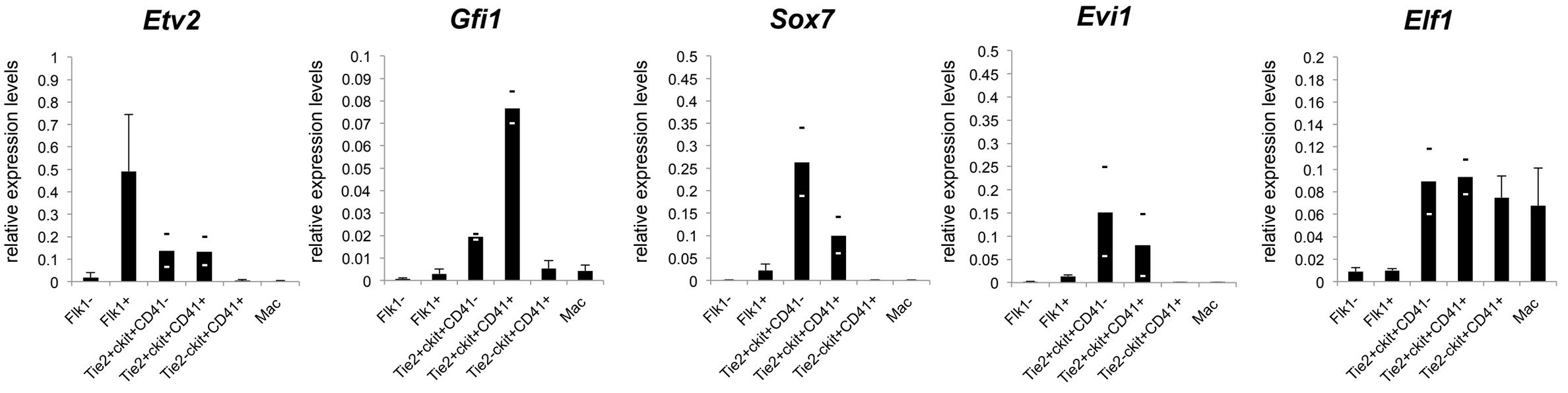
MoFlo setup (blast culture day3):



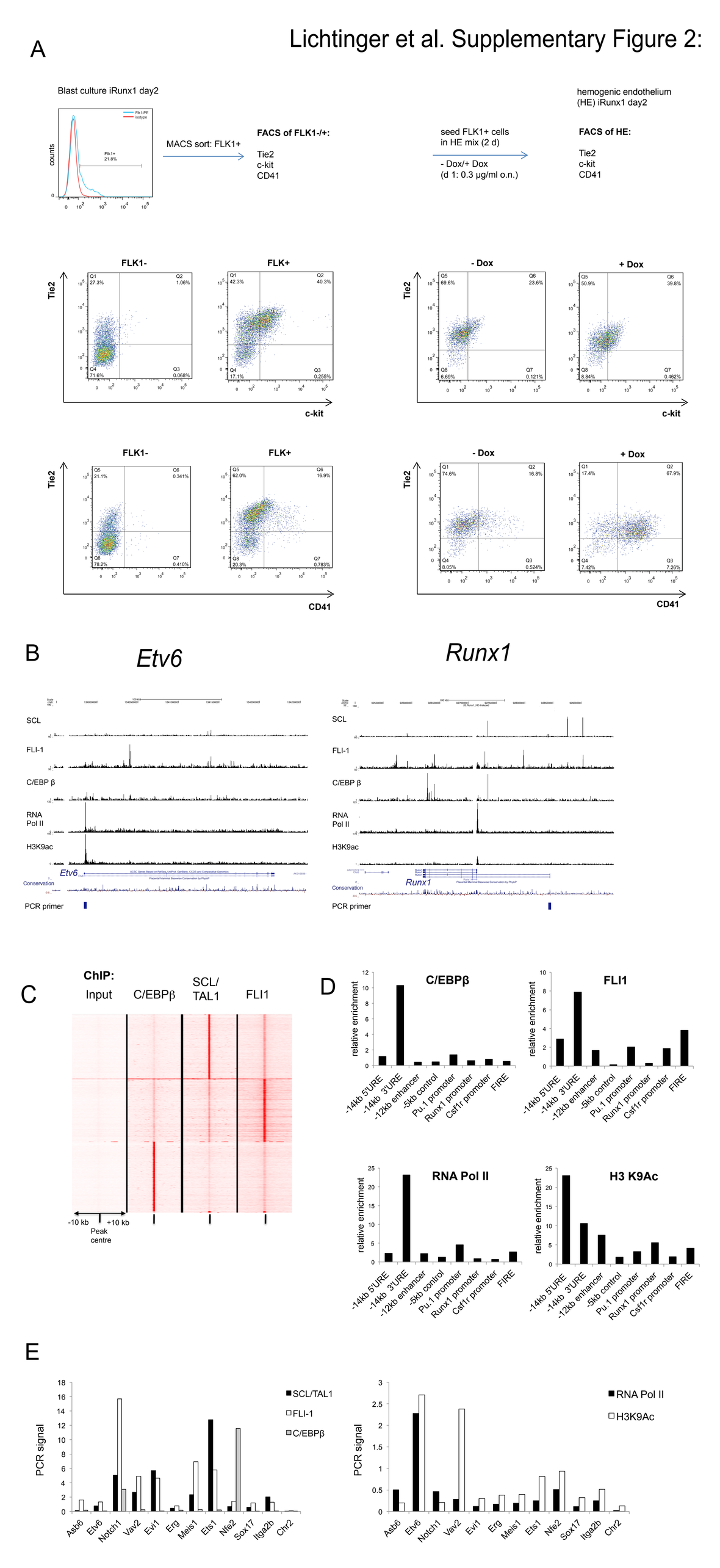
Purity Check:



B

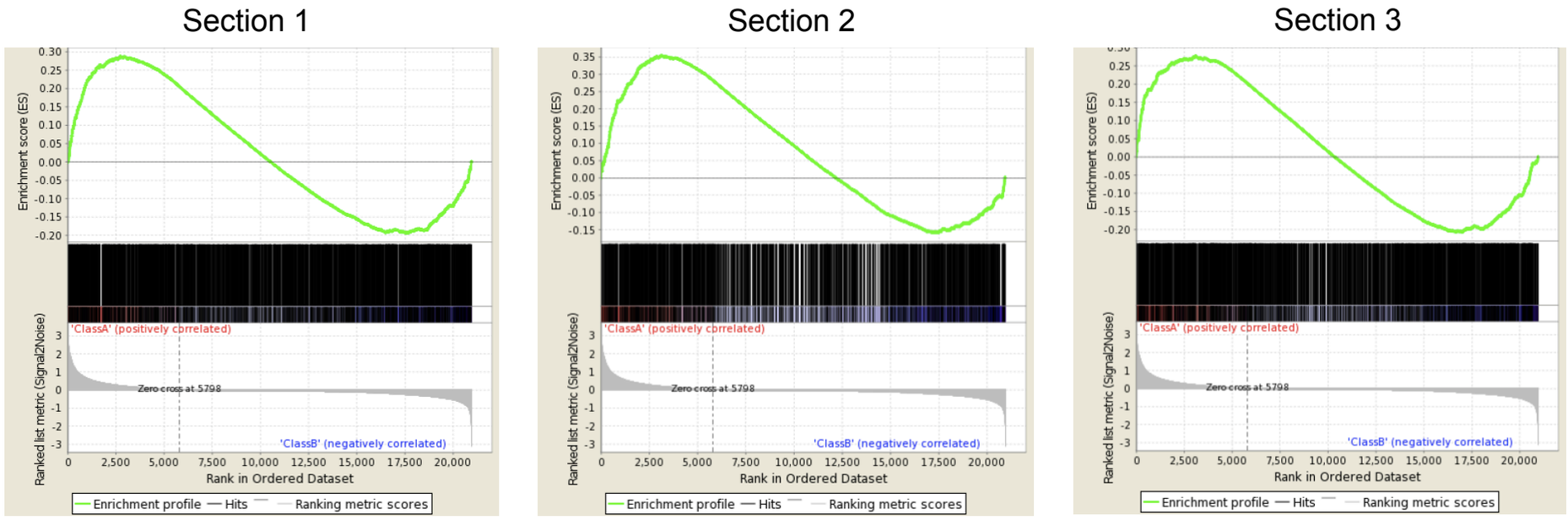


Lichtinger et al. Supplementary Figure 2:



Lichtinger et al. Supplementary Figure 3

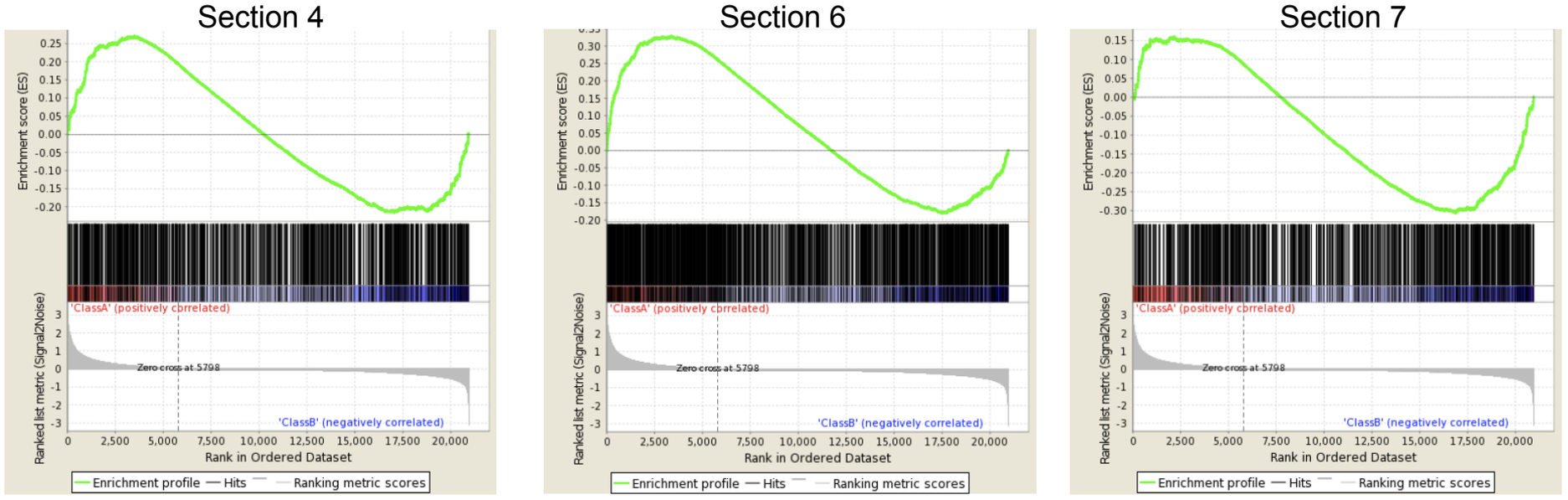
A



NES: 1.06, nominal p-Value: 0.2

NES: 1.28, nominal p-Value: 0.002

NES: 1.02, nominal p-Value: 0.405



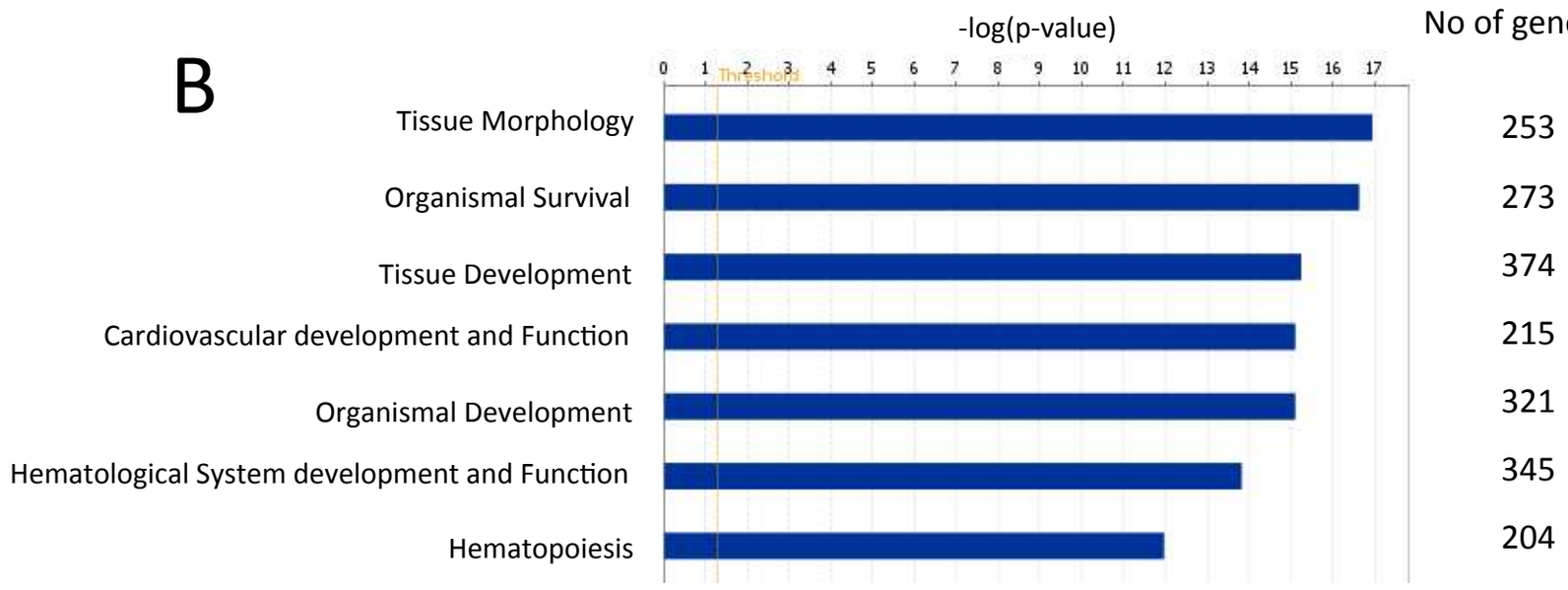
NES: 0.96, nominal p-Value: 0.608

NES: 1.19, nominal p-Value: 0.018

NES: -1.26, nominal p-Value: 0.0

HSCs ⟶ Mature blood cells

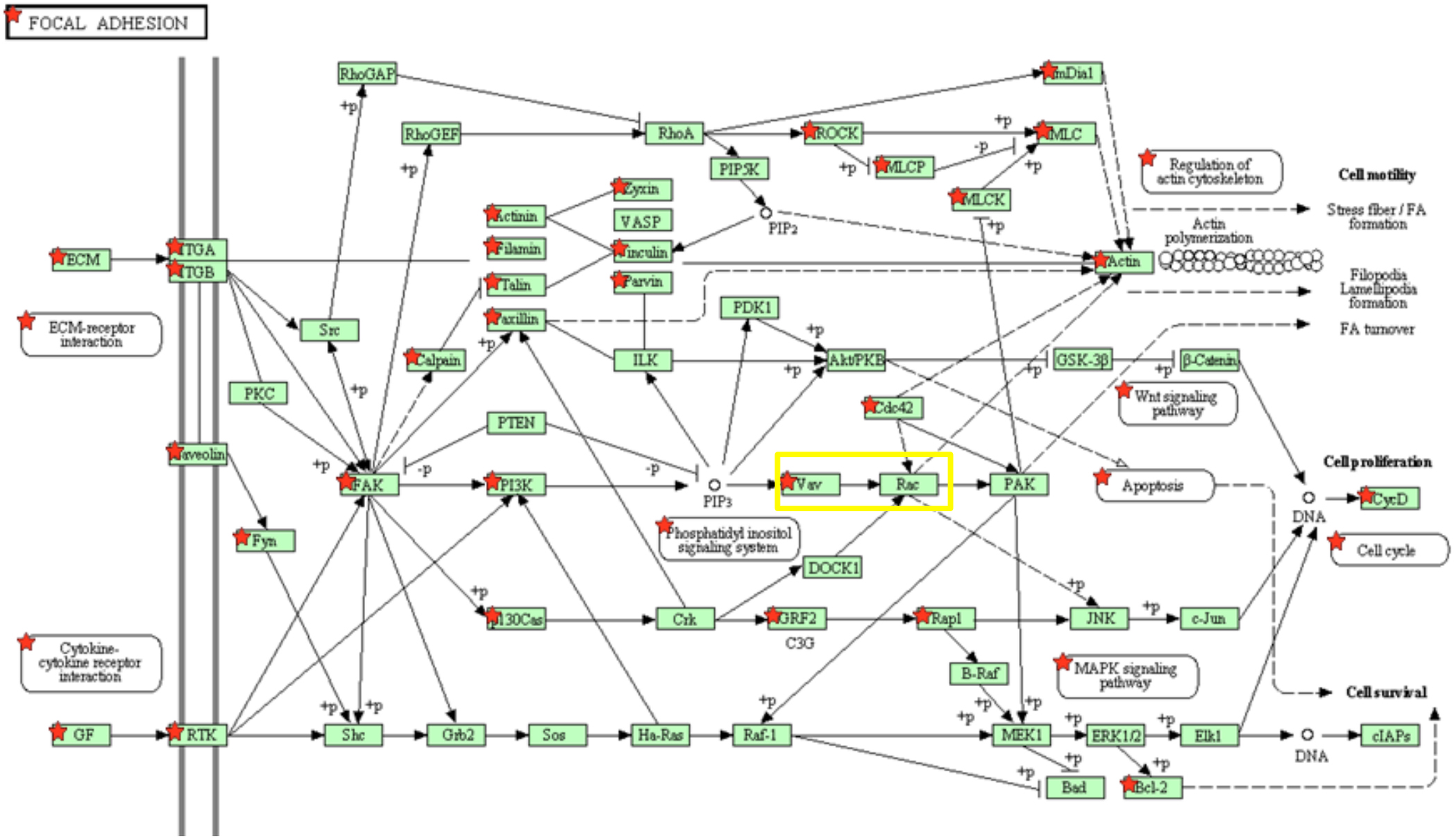
B



Physiological and systems development functions

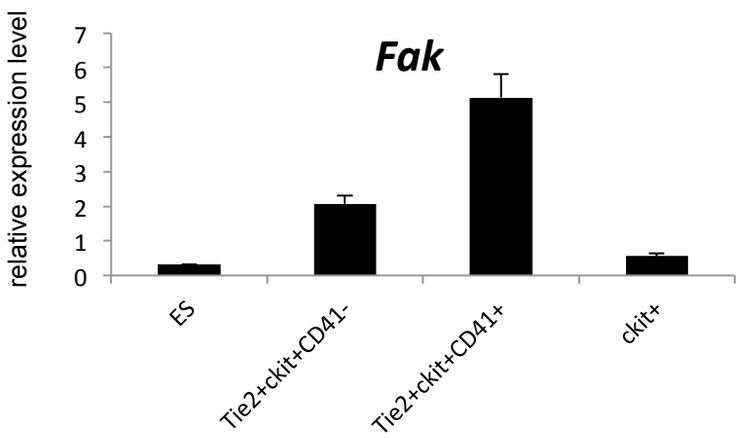
C

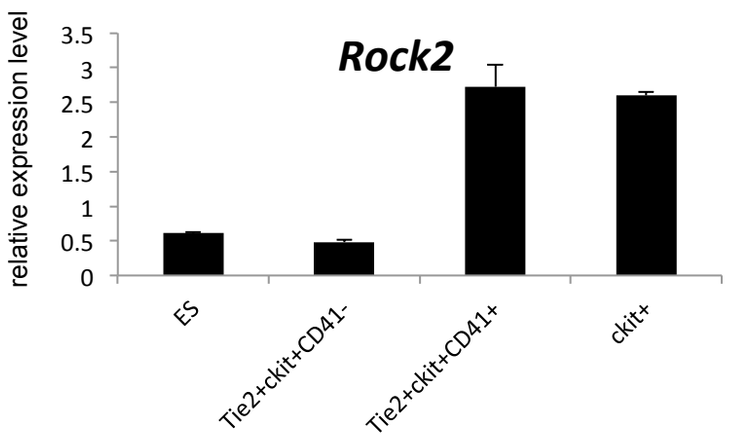
KEGG pathway analysis

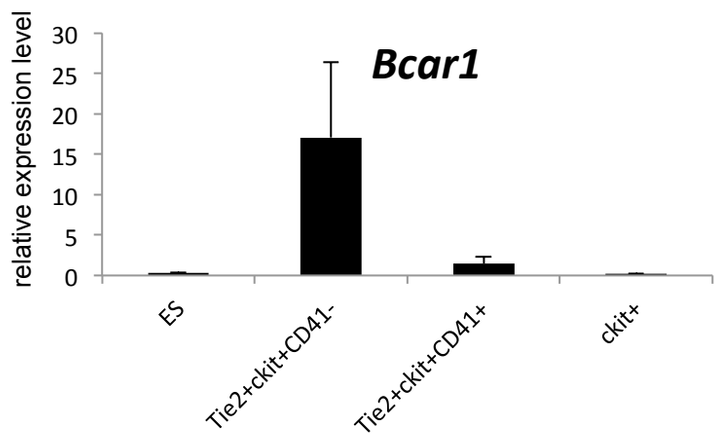


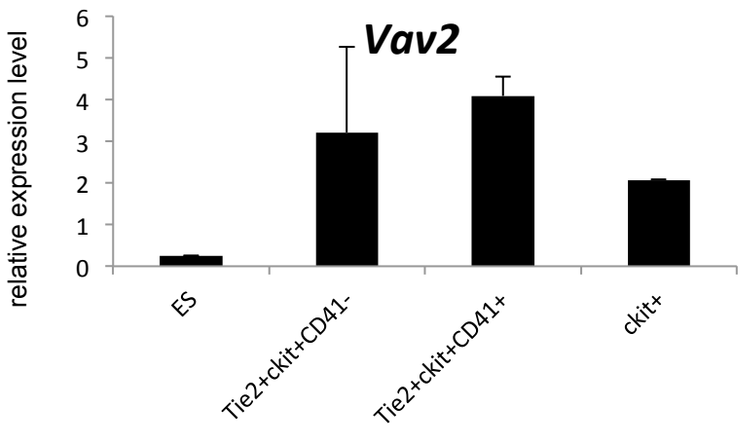
p value:
$5.4E^{-15}$

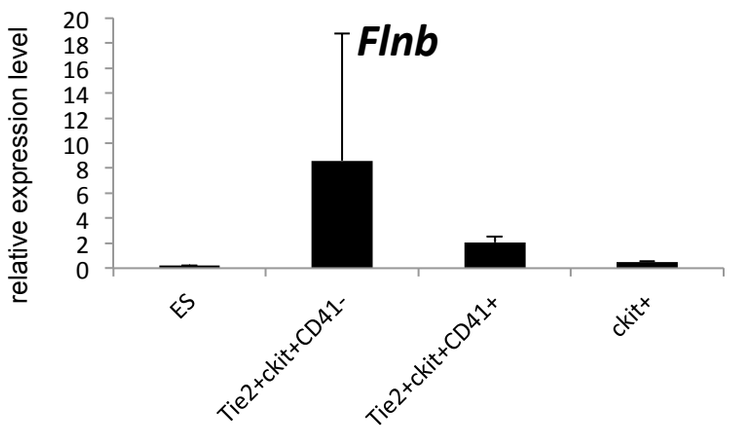
D











E

Pol II -DOX | H3K9Ac -DOX | Expression

FLI1 peak location

Promoter peaks 1926 Genes 1402

Intragenic peaks 4574 Genes 1318

Intergenic peaks 4123 Genes 1460

SCL/TAL1 peak location

Promoter peaks 1301 Genes 2345

Intragenic peaks 3222 Genes 1538

Intergenic peaks 3984 Genes 1355

CEBPβ peak location

Promoter peaks 1322 Genes 1602

Intragenic peaks 4429 Genes 1937

Intergenic peaks 4440 Genes 1697

tags (log2) 4 8

expression (log2) 6 12

Signals below threshold (log2 < 6)

SCL/TAL1

(449)

Promoter peaks

Intragenic peaks

Intergenic peaks

Pol II -DOX | H3K9Ac -DOX

FLI1

(405)

Promoter peaks

Intragenic peaks

Intergenic peaks

Pol II -DOX | H3K9Ac -DOX

CEBPβ

(451)

Promoter peaks

Intragenic peaks

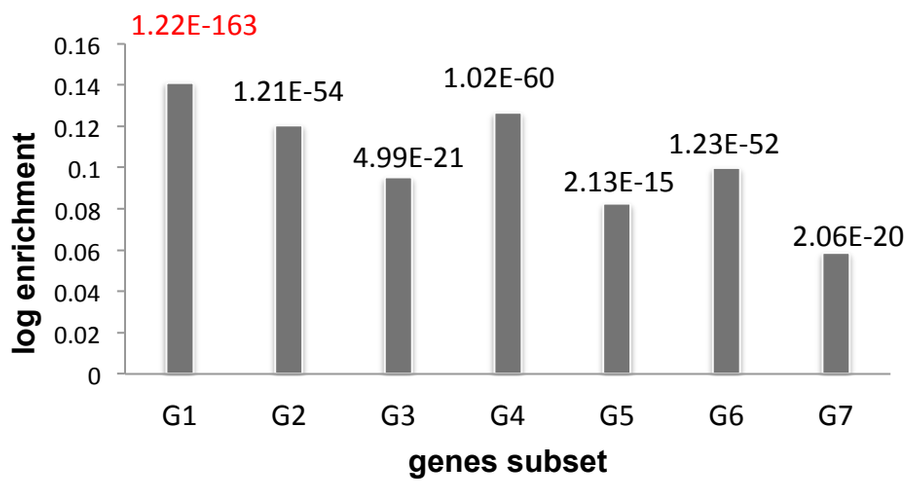
Intergenic peaks

Pol II -DOX | H3K9Ac -DOX

F

HIGH (Nh) expressed genes – log2 ≥6	13867
LOW (NI) non-expressed genes log2 <6	8182

genes subset	actual number (n_i)	BY CHANCE ((Nh*n_i)/(Nh+NI), {i=1 , ..., 7}	actual number in HIGH	actual number in LOW	p-value	log(enrichment)
G1	2300	1446.51	2002	298	1.22E-163	0.141143
G2	1171	736.46	972	199	1.21E-54	0.120516
G3	760	477.98	595	165	4.99E-21	0.09511
G4	1160	729.54	976	184	1.02E-60	0.126398
G5	747	469.80	568	179	2.13E-15	0.082434
G6	1739	1093.69	1376	363	1.23E-52	0.099725
G7	2064	1298.09	1486	578	2.06E-20	0.058715



G1 – FLI1, SCL/TAL1, C/EBPβ
G2 - FLI1, SCL/TAL1
G3 - FLI1 and C/EBPβ
G4 – C/EBPβ and SCL/TAL1
G5 - FLI1 only
G6 – SCL/TAL1 only
G7 – C/EBPβ only

Lichtinger et al., Supplementary Figure 4

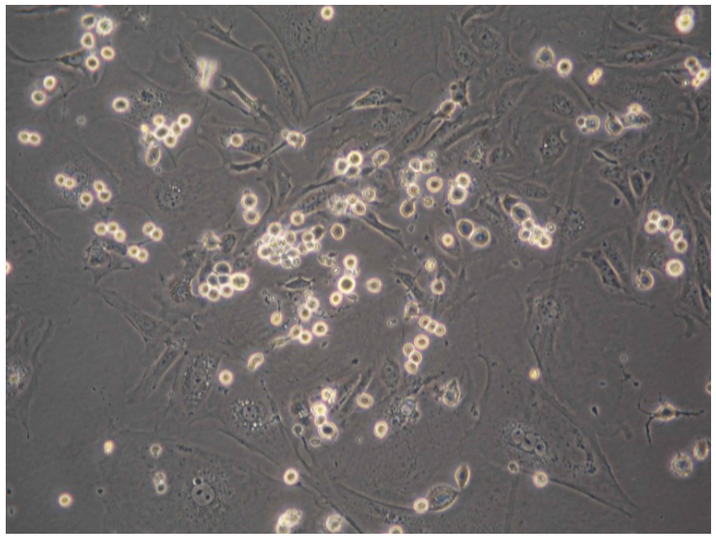
A

tgggcgcttc ctgttttctc aggc **Wild type**

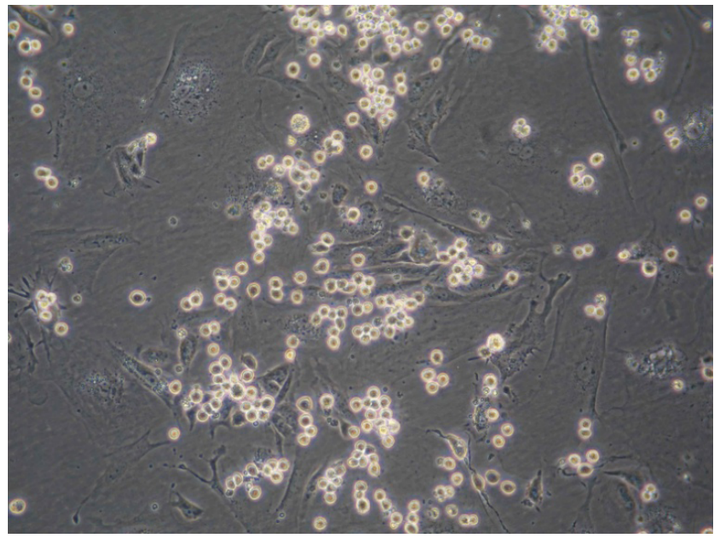
acccg**cgaag gacaaaa**gag tccg **Mutant**

B

Pu.1+/ki

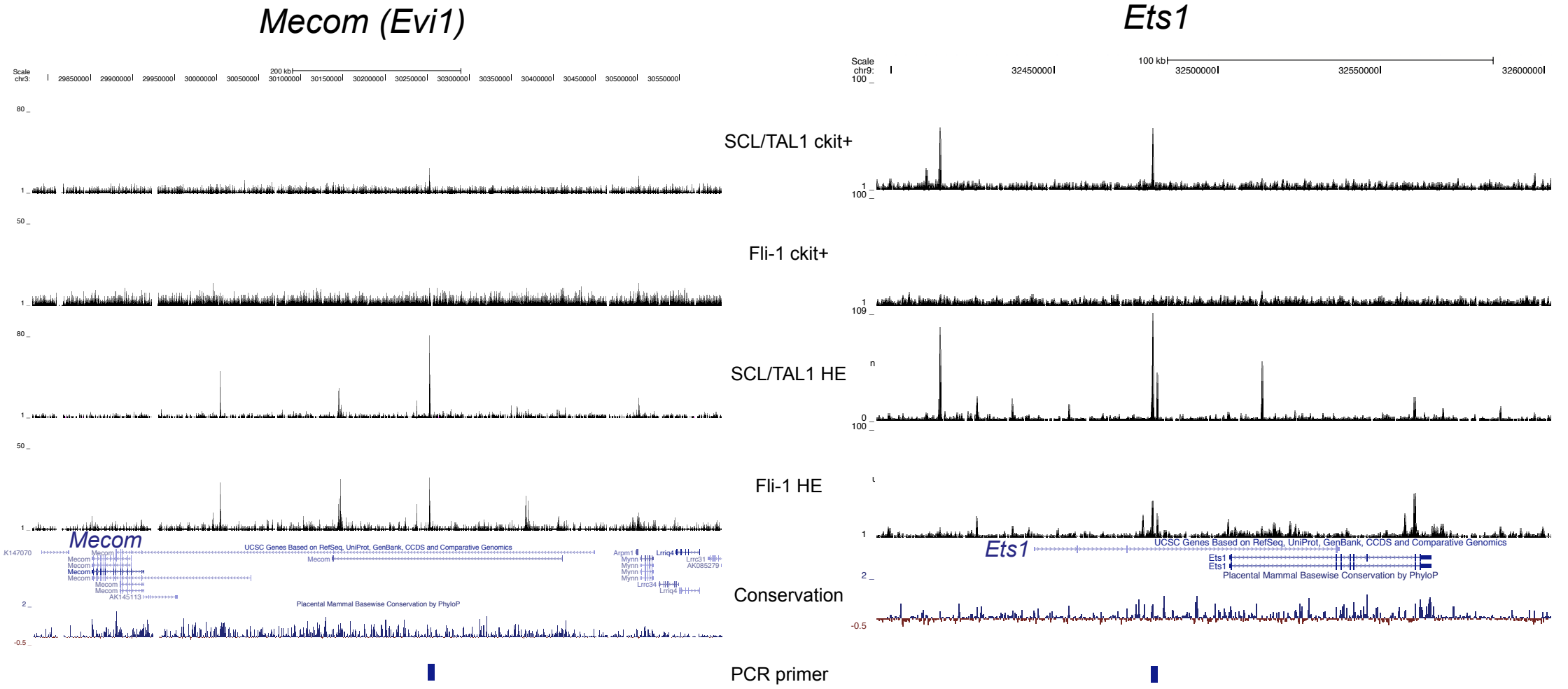


Pu.1ki/ki

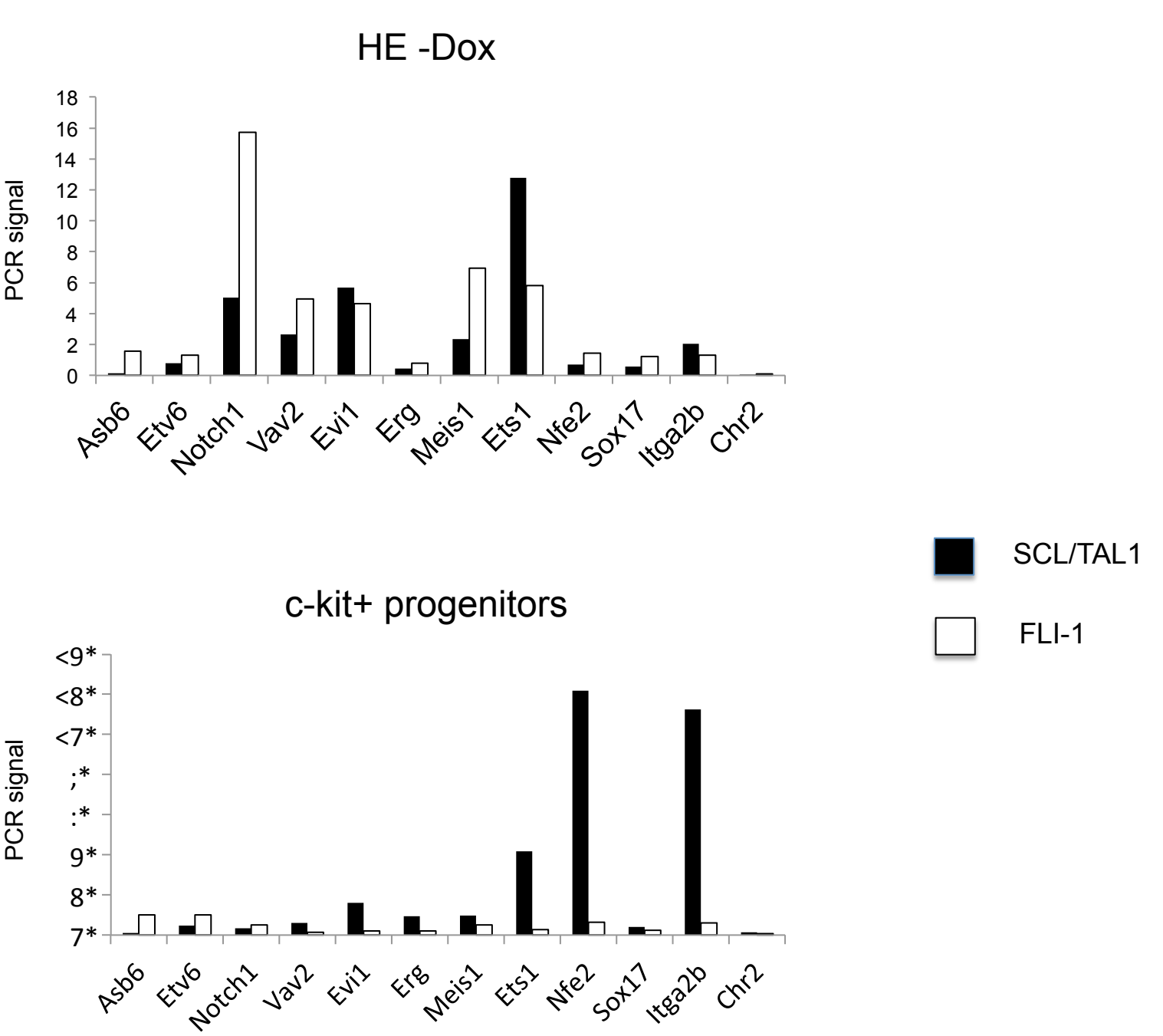


Lichtinger et al., Supplementary Figure 5

A



B

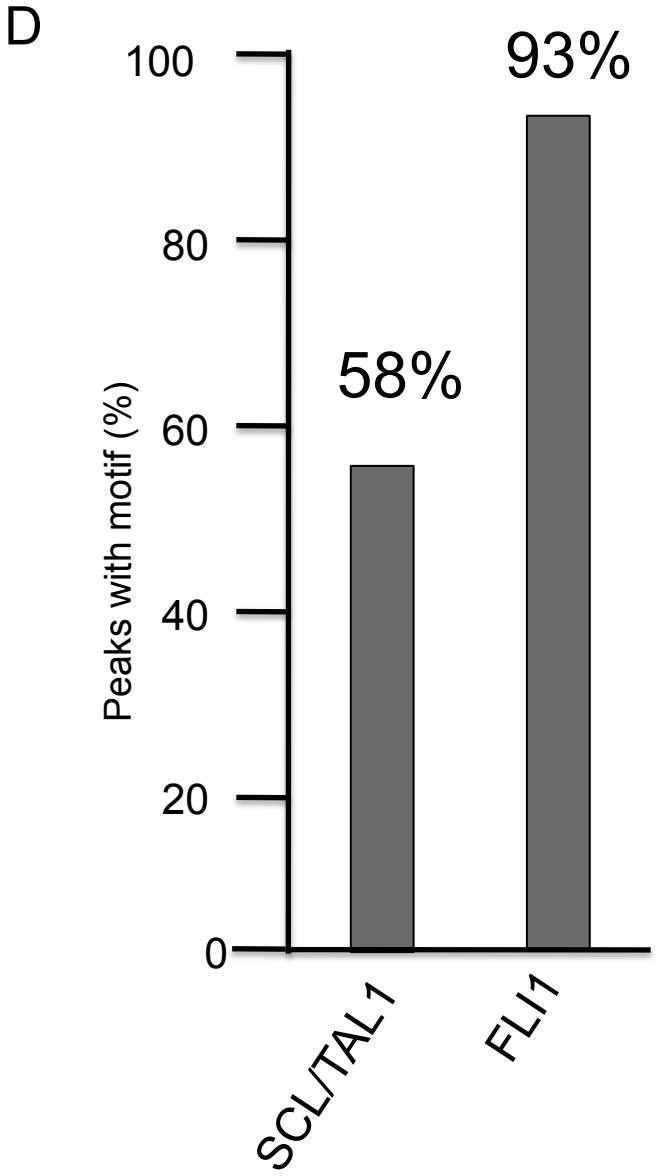


C

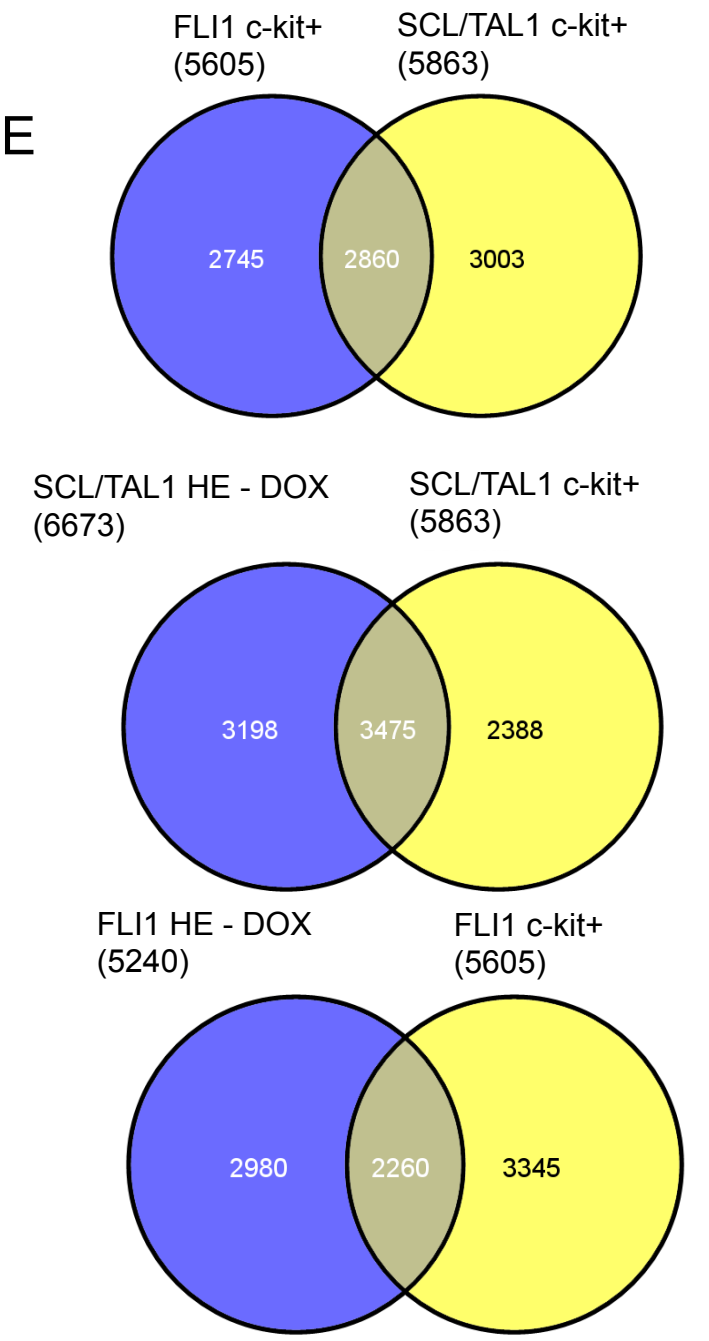
FLI-1 and SCL associated motifs in ckit+ cells

Section	Matrix	Motif	log p-Value	
1		ETS	-2.97e+3	FLI1 unique
2		ETS	-9.46e+2	FLI1 SCL/TAL1
2		ETS/Ebox	-4.41e+2	
2		ETS	-1.66e+2	
2		RUNX	--7.01e+1	
3		E-box	-1.84e+3	SCL/TAL1 unique
3		GATA	-1.82e+3	
3		ETS	-1.23e+3	
3		E-box	-1.18e+3	
3		RUNX	-6.75e+2	

D

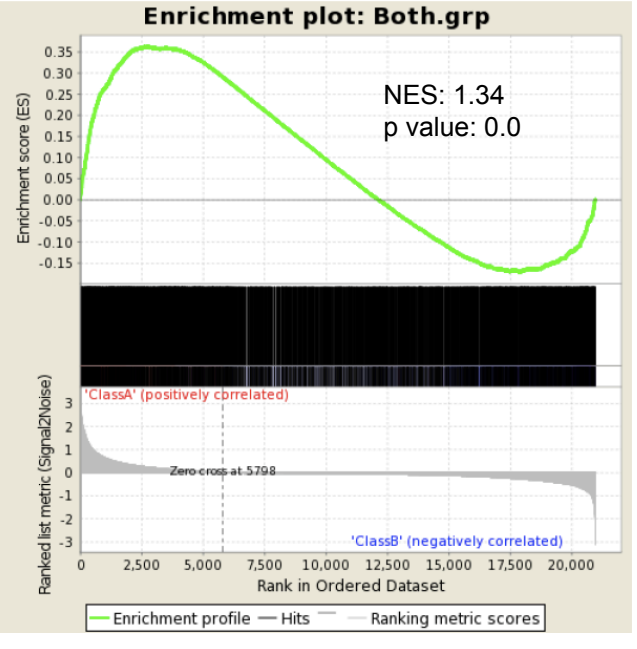


E

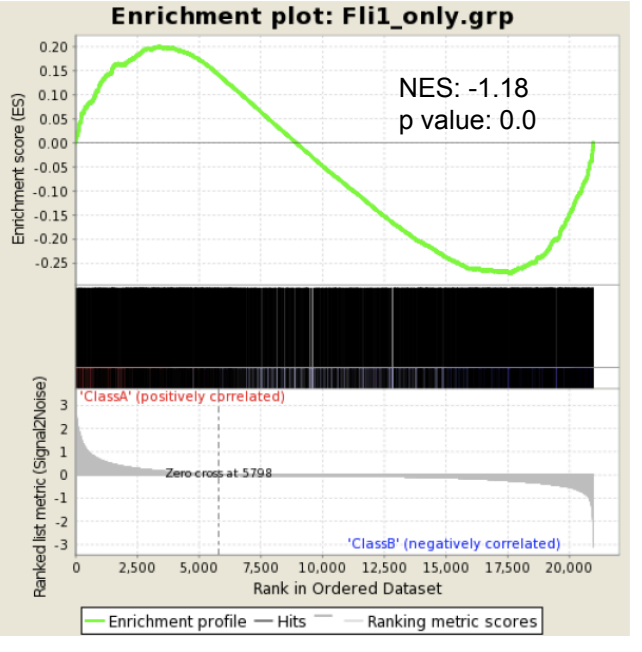


F

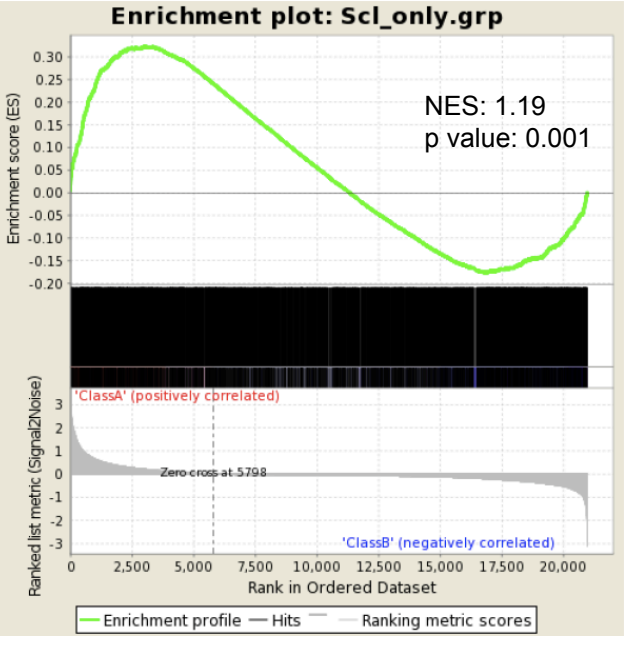
FLI1 + SCL/TAL1



FLI1 unique

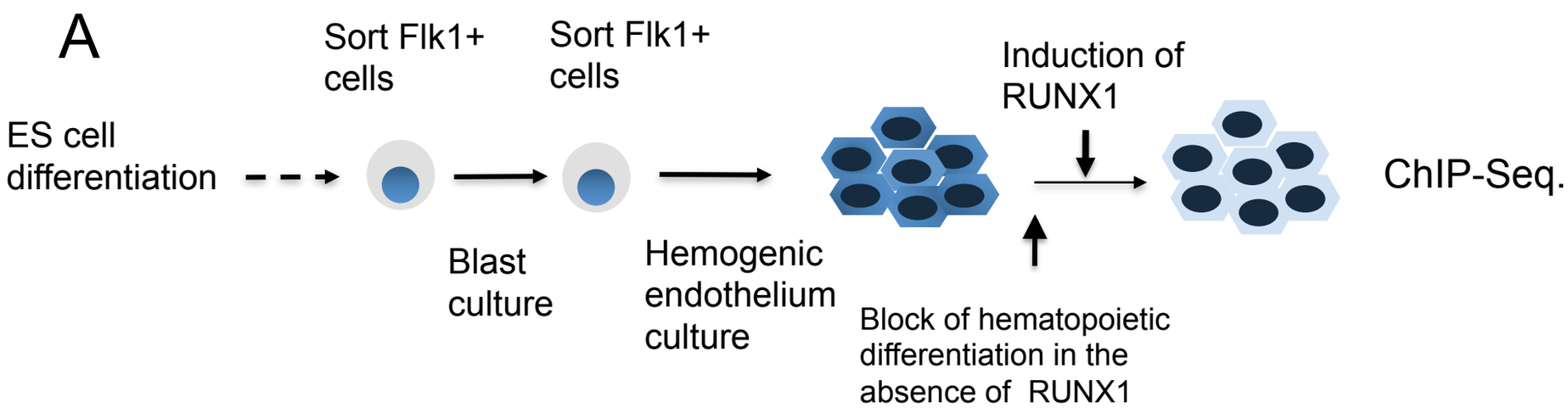


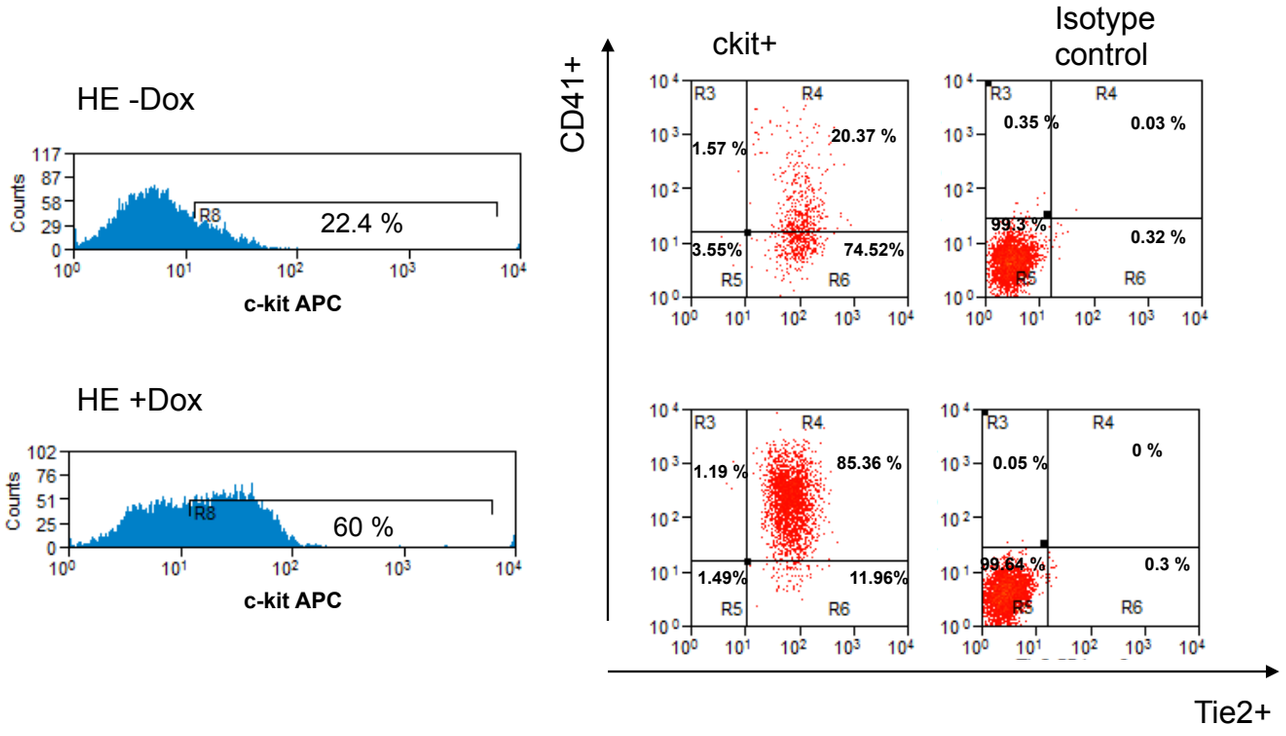
SCL/TAL1 unique



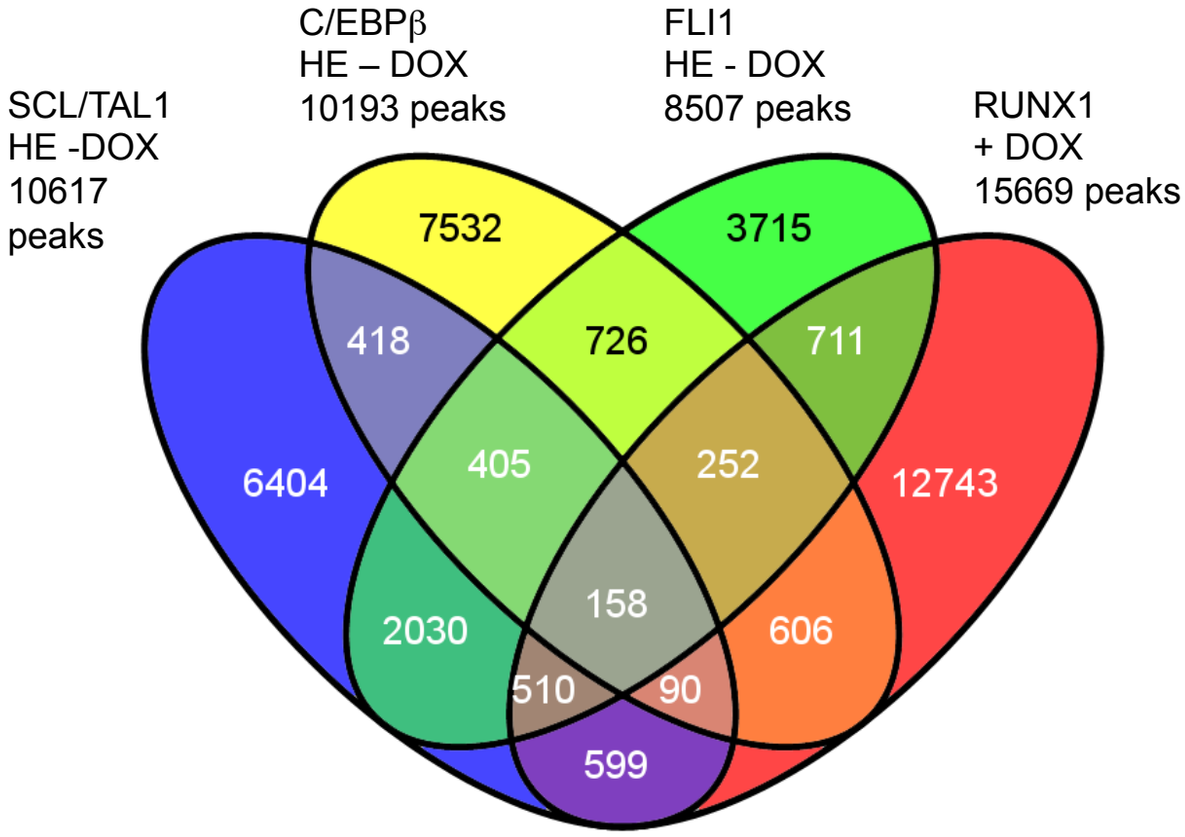
Lichtinger et al., Supplementary Figure 6

A

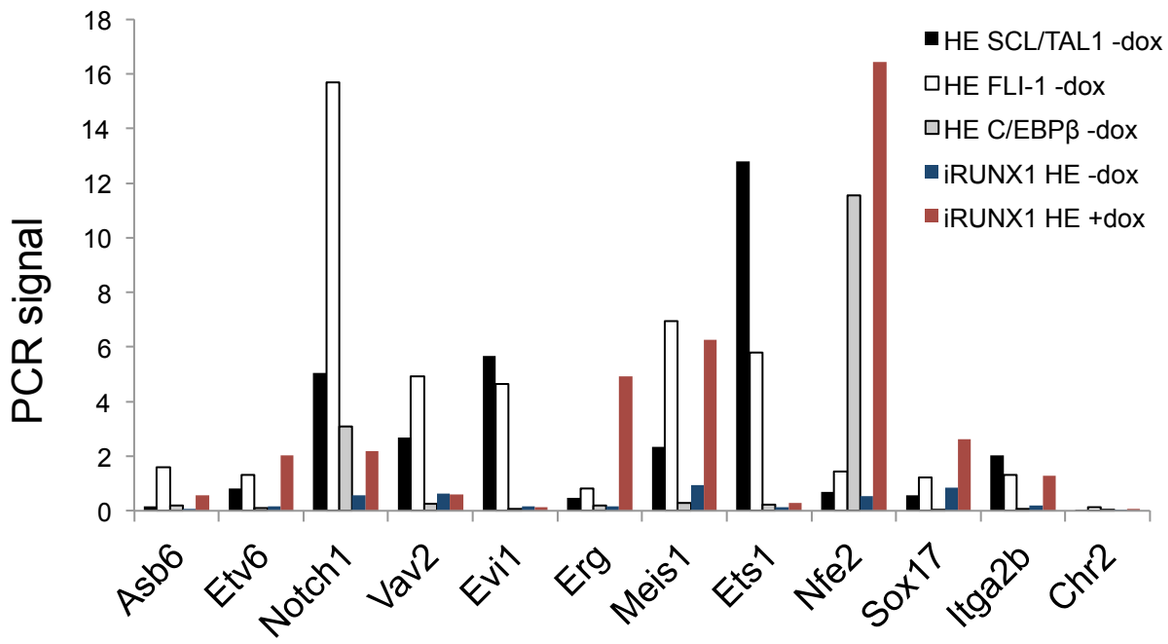


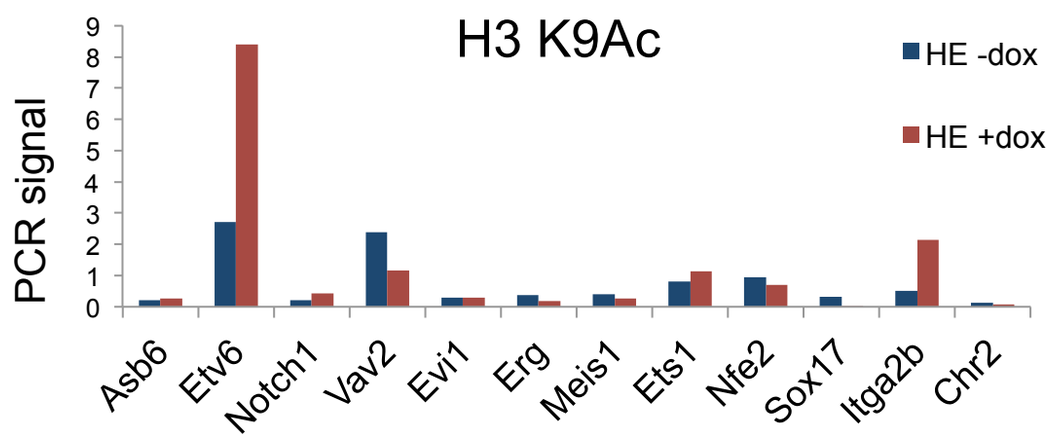


B

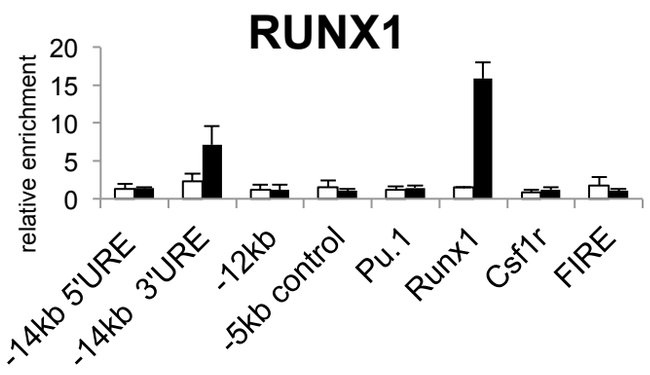


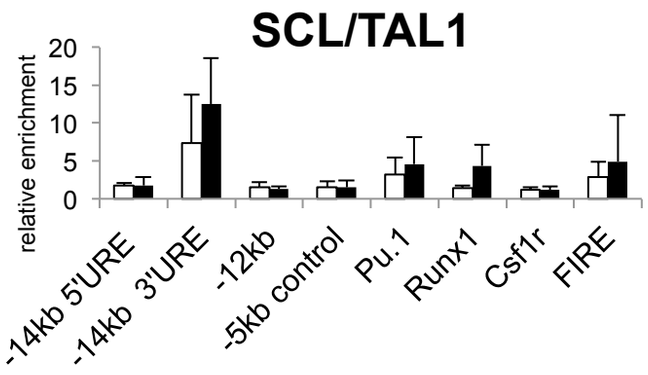
C

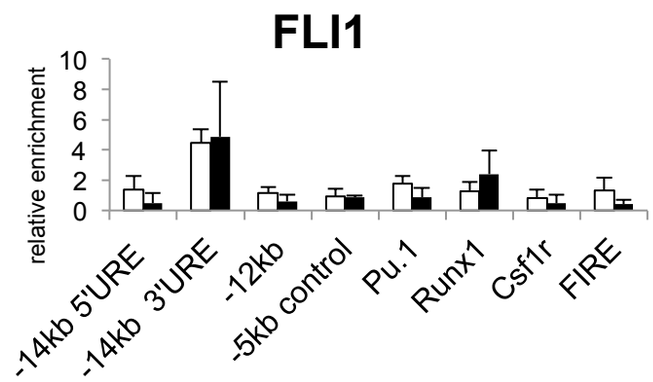

PCR signal
HE SCL/TAL1 -dox
HE FLI-1 -dox
HE C/EBPβ -dox
iRUNX1 HE -dox
iRUNX1 HE +dox
Asb6
Etv6
Notch1
Vav2
Evi1
Erg
Meis1
Ets1
Nfe2
Sox17
Itga2b
Chr2

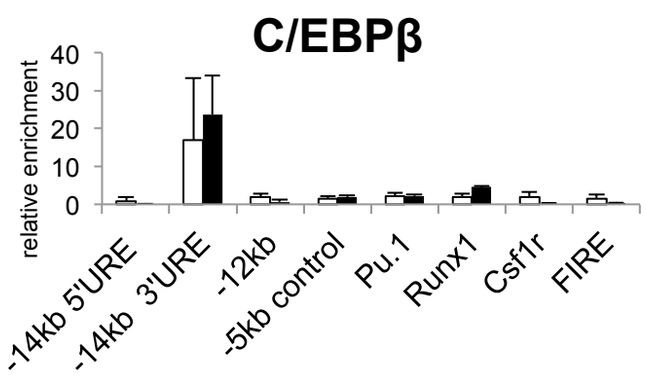

H3 K9Ac
PCR signal
HE -dox
HE +dox
Asb6
Etv6
Notch1
Vav2
Evi1
Erg
Meis1
Ets1
Nfe2
Sox17
Itga2b
Chr2

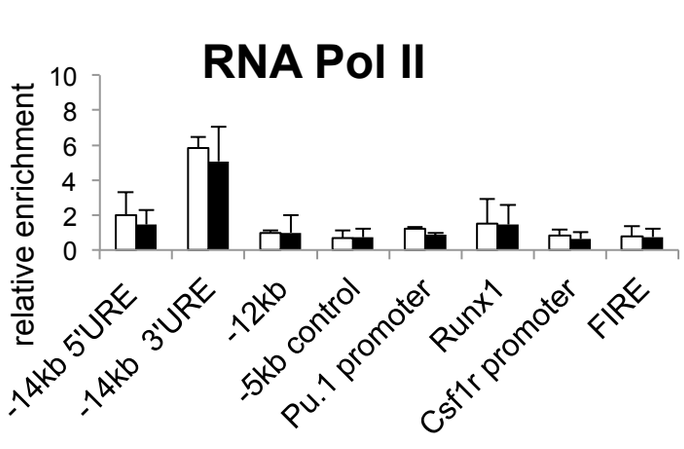
D

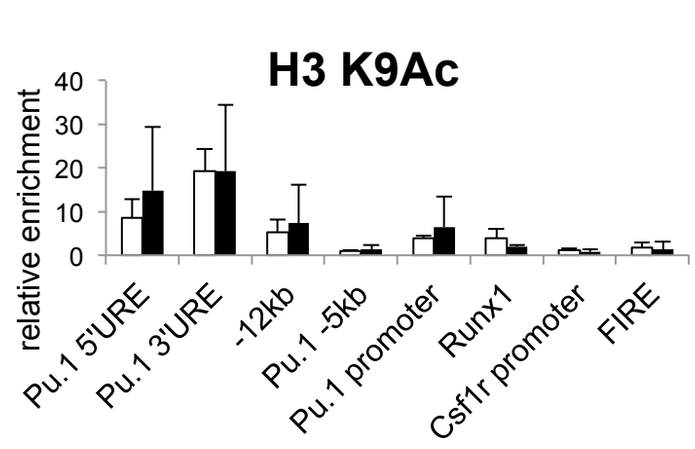

RUNX1
relative enrichment
-14kb 5'URE
-14kb 3'URE
-12kb
-5kb control
Pu.1
Runx1
Csf1r
FIRE

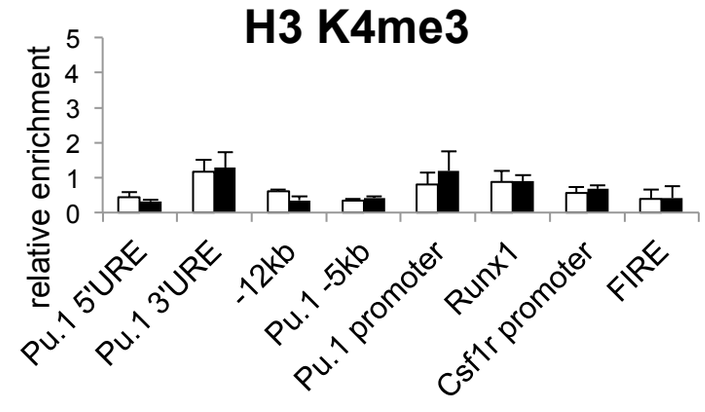

SCL/TAL1
relative enrichment
-14kb 5'URE
-14kb 3'URE
-12kb
-5kb control
Pu.1
Runx1
Csf1r
FIRE

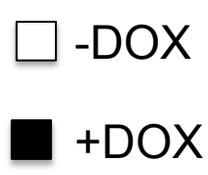

FLI1
relative enrichment
-14kb 5'URE
-14kb 3'URE
-12kb
-5kb control
Pu.1
Runx1
Csf1r
FIRE


C/EBPβ
relative enrichment
-14kb 5'URE
-14kb 3'URE
-12kb
-5kb control
Pu.1
Runx1
Csf1r
FIRE


RNA Pol II
relative enrichment
-14kb 5'URE
-14kb 3'URE
-12kb
-5kb control
Pu.1 promoter
Runx1
Csf1r promoter
FIRE


H3 K9Ac
relative enrichment
Pu.1 5'URE
Pu.1 3'URE
-12kb
Pu.1 -5kb
Pu.1 promoter
Runx1
Csf1r promoter
FIRE

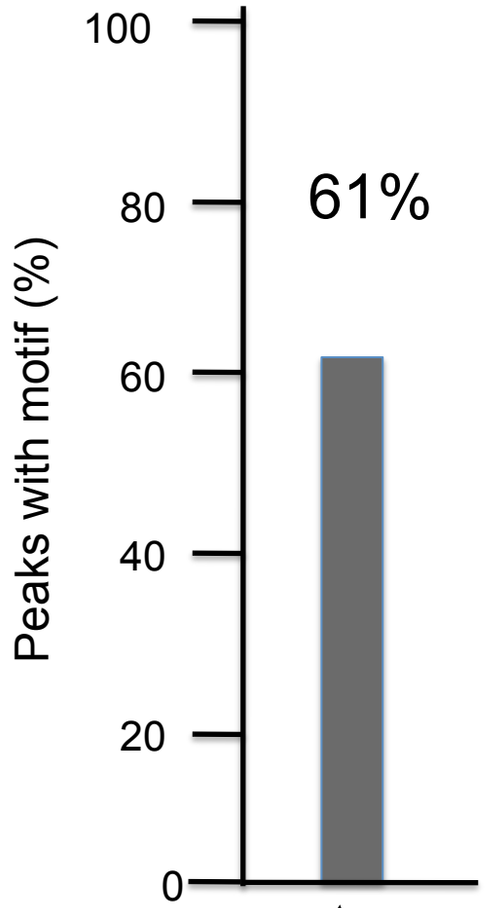

H3 K4me3
relative enrichment
Pu.1 5'URE
Pu.1 3'URE
-12kb
Pu.1 -5kb
Pu.1 promoter
Runx1
Csf1r promoter
FIRE


-DOX
+DOX

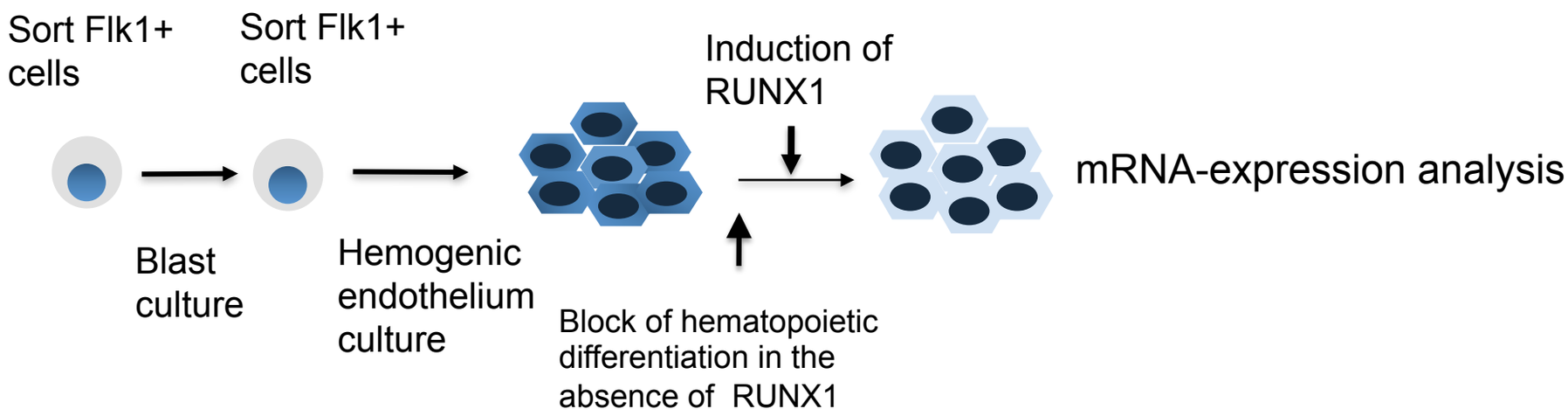
E

RUNX1 associated motifs

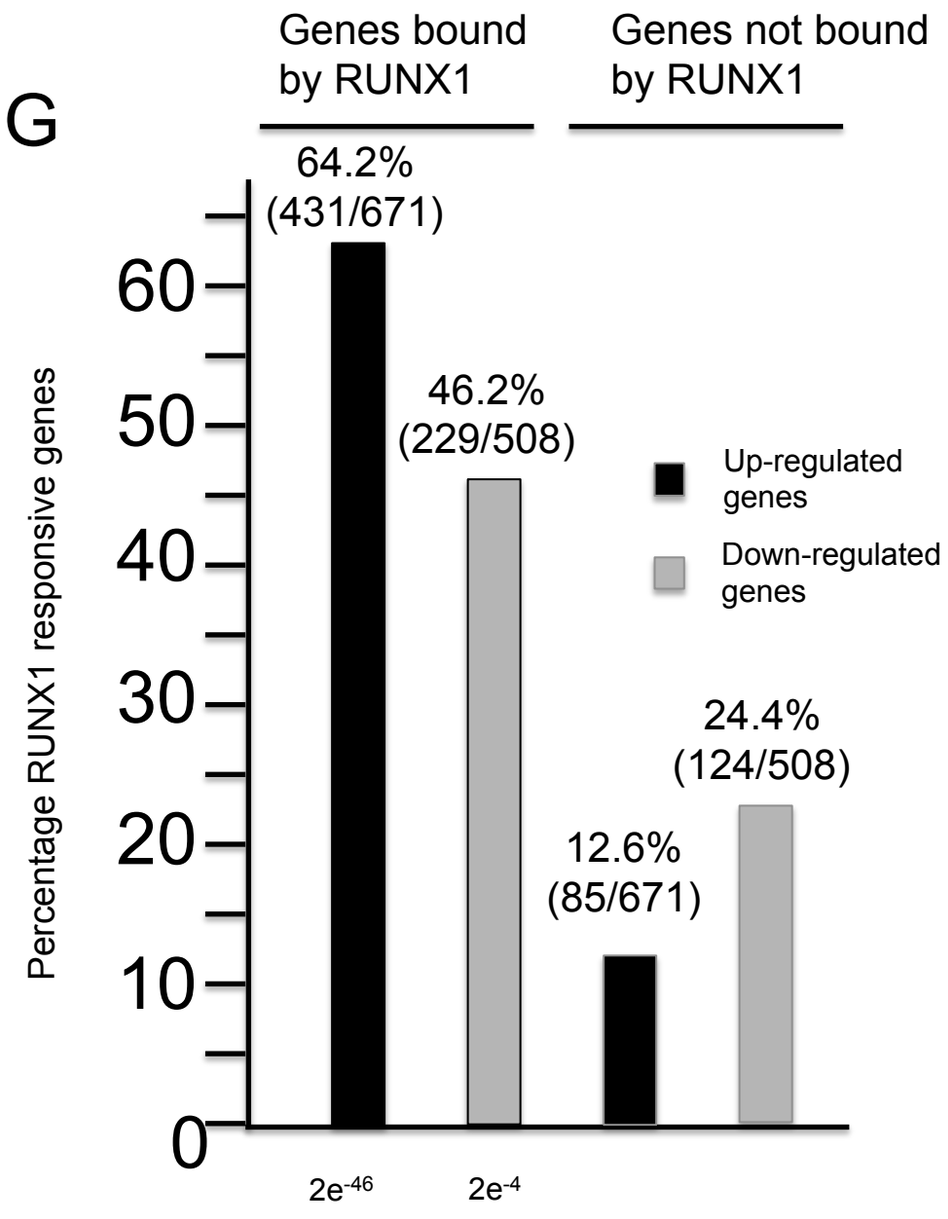
Matrix	Motif	log p-Value
	RUNX	-1.3e+4
	RUNX	-1.24e+4
	ETS	-1.04e+3
	NF-E2/AP1	-6.81e+2



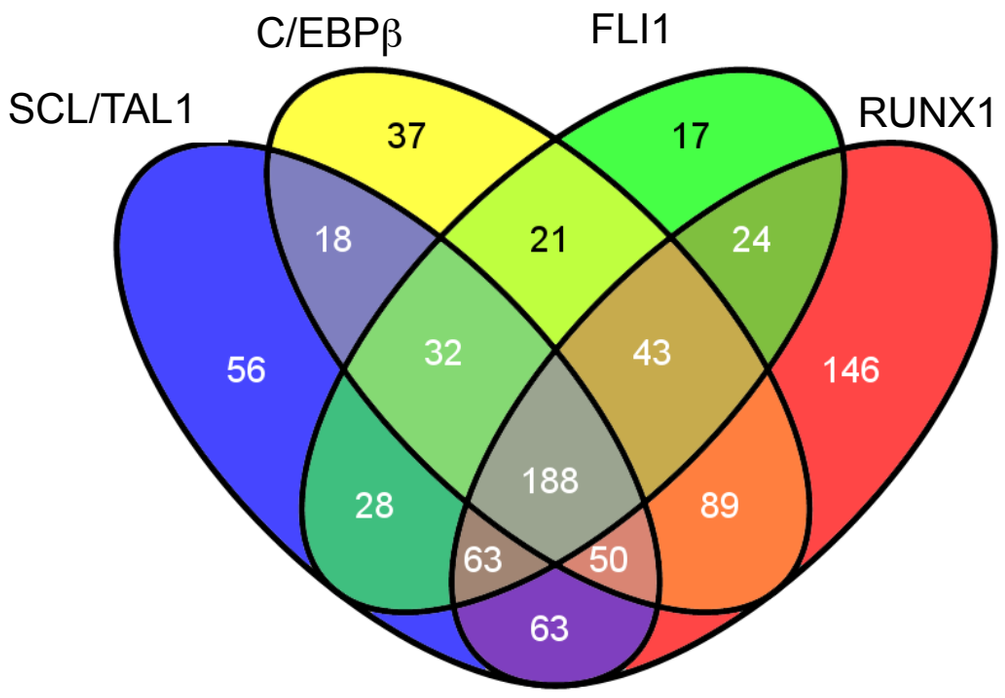
F

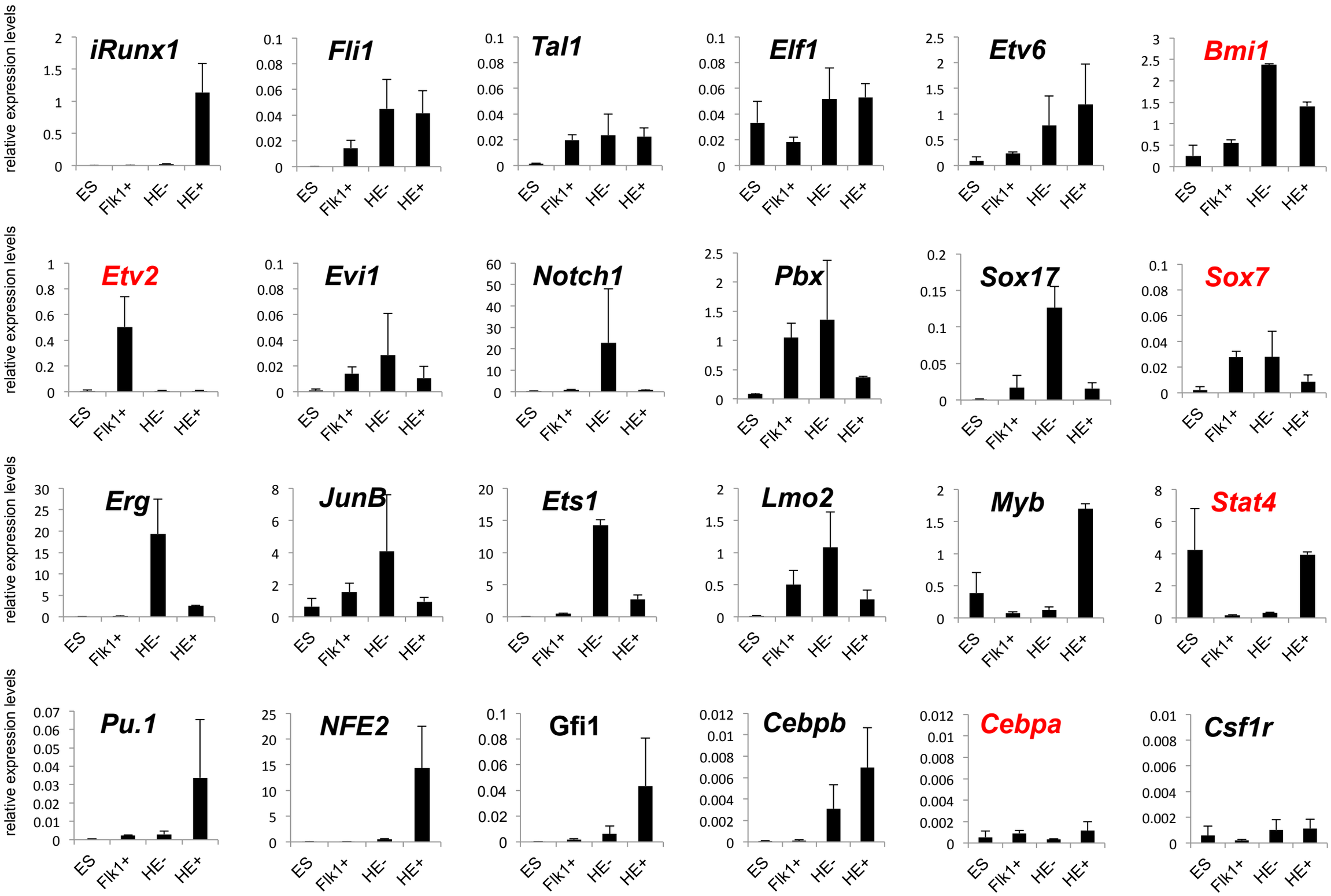


G



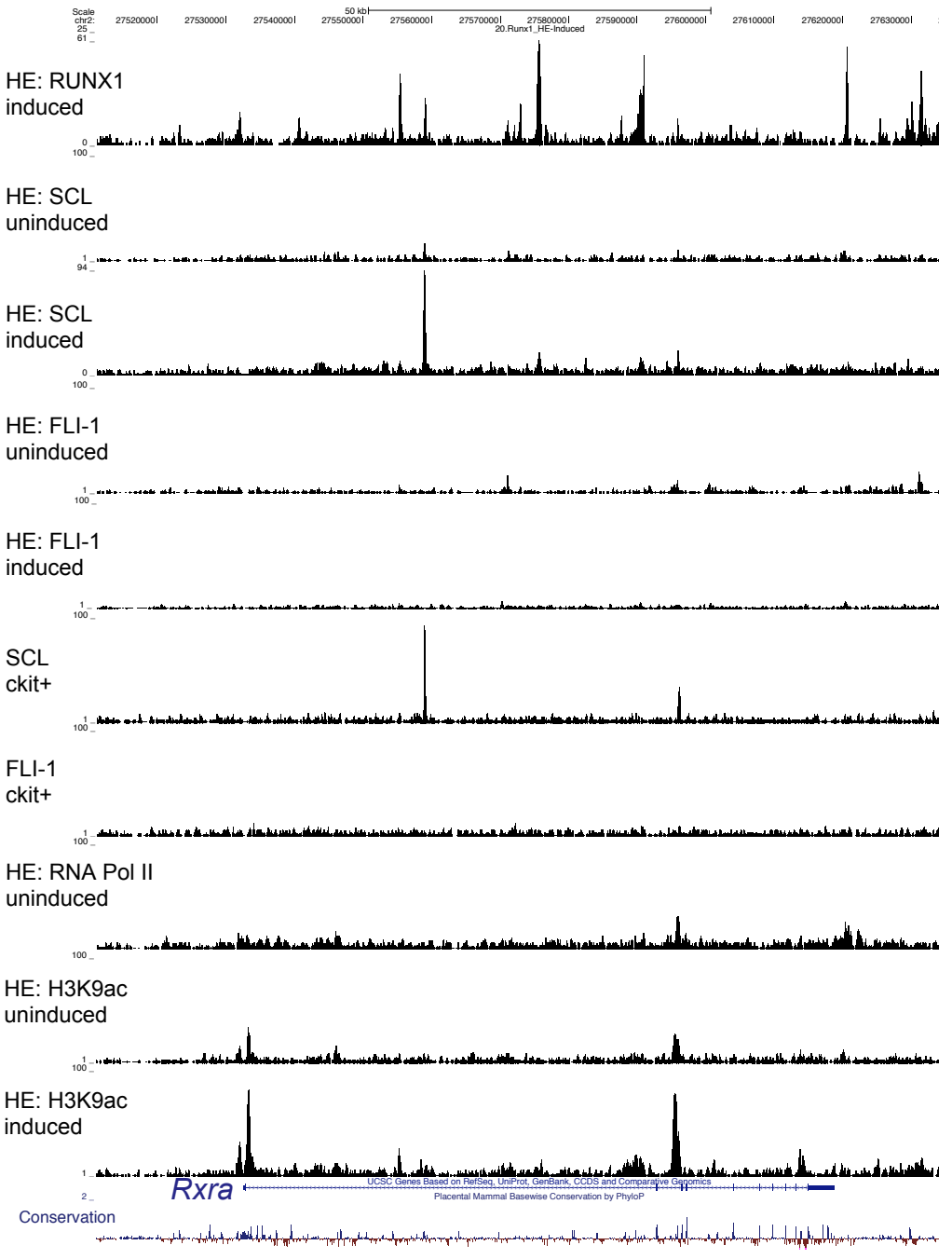
H

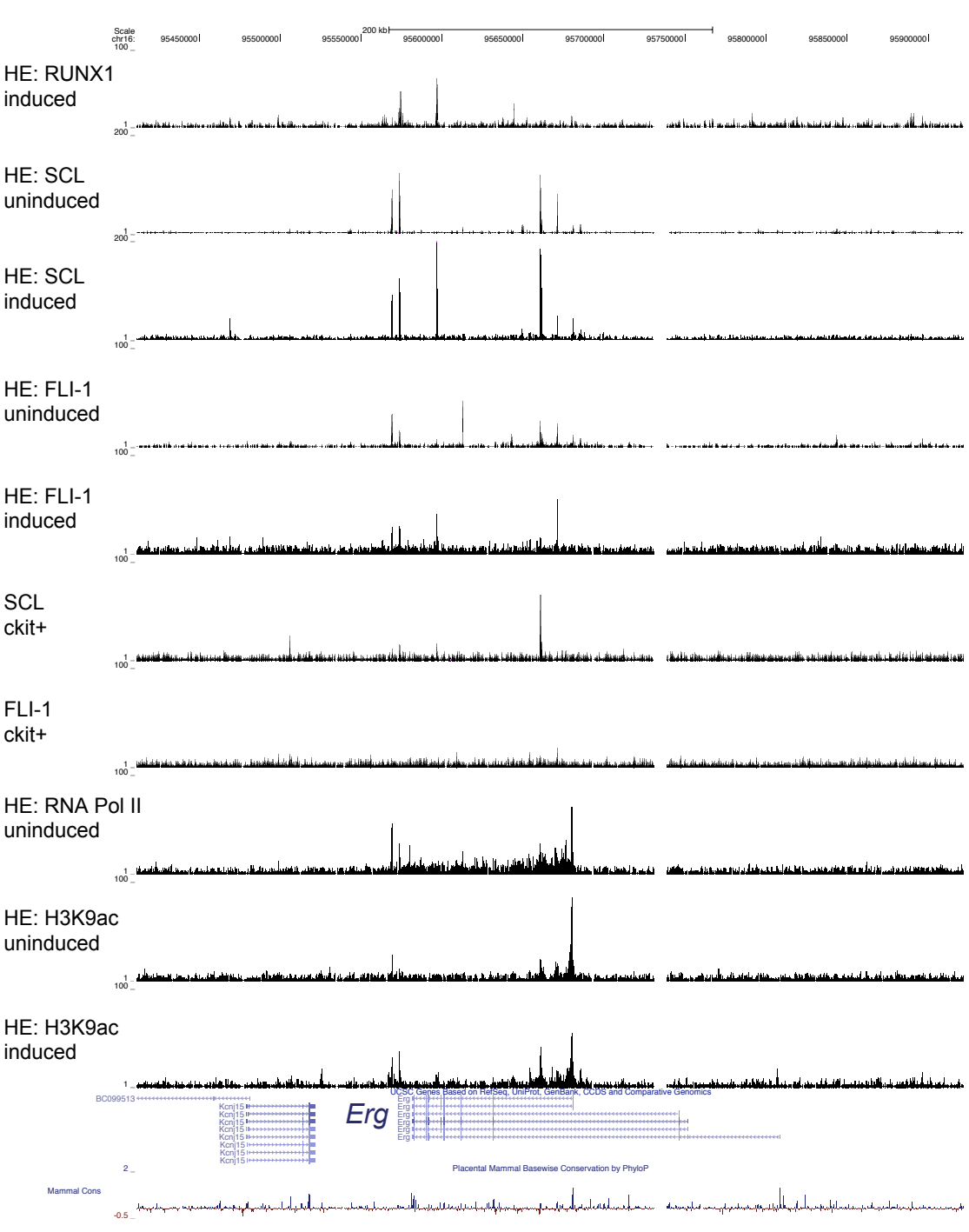



relative expression levels
iRunx1
Fli1
Tal1
Elf1
Etv6
Bmi1
Etv2
Evi1
Notch1
Pbx
Sox17
Sox7
Erg
JunB
Ets1
Lmo2
Myb
Stat4
Pu.1
NFE2
Gfi1
Cebpb
Cebpa
Csf1r
ES
Flk1+
HE-
HE+

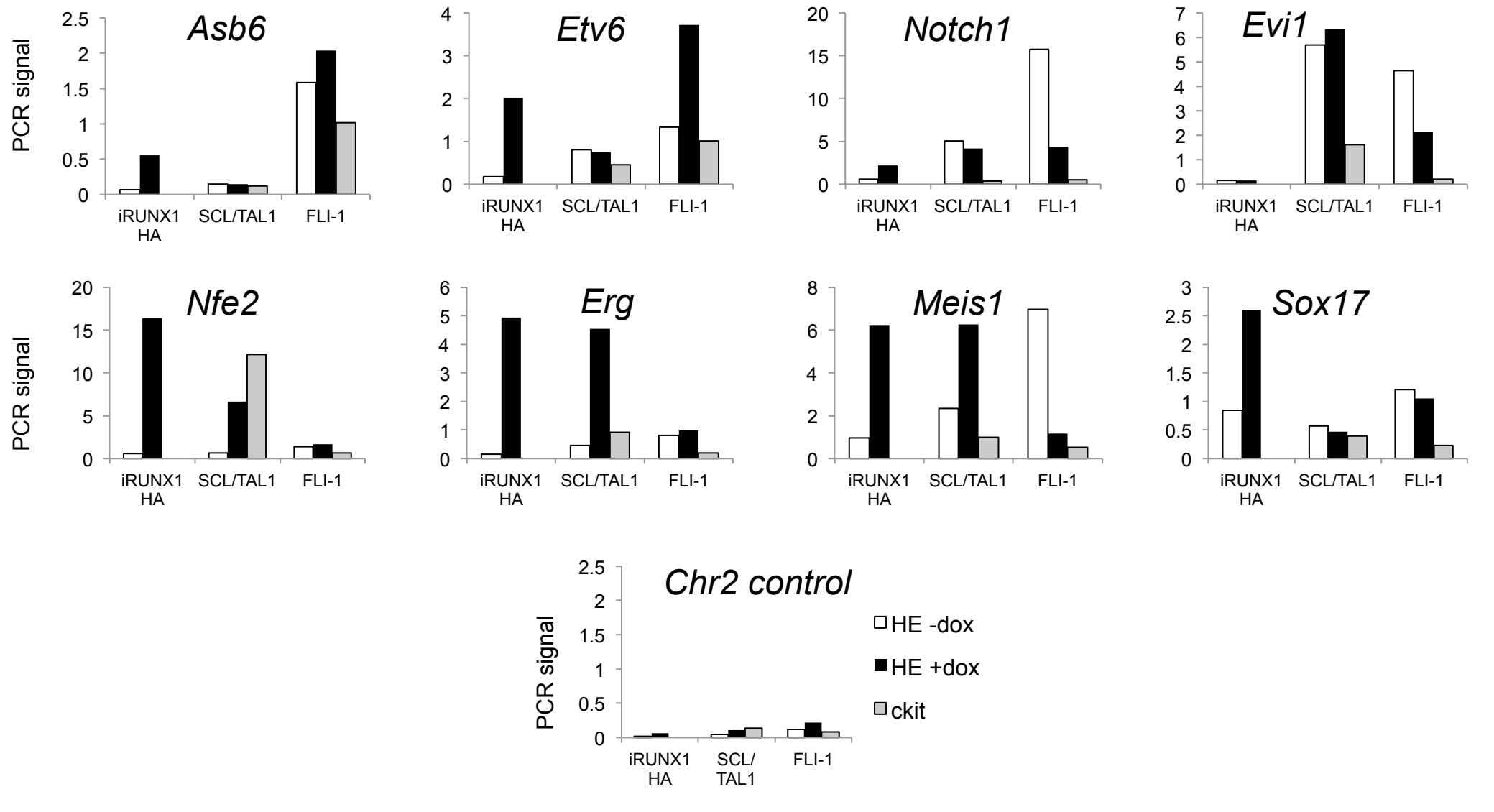
Lichtinger et al., Supplementary Figure 7

A

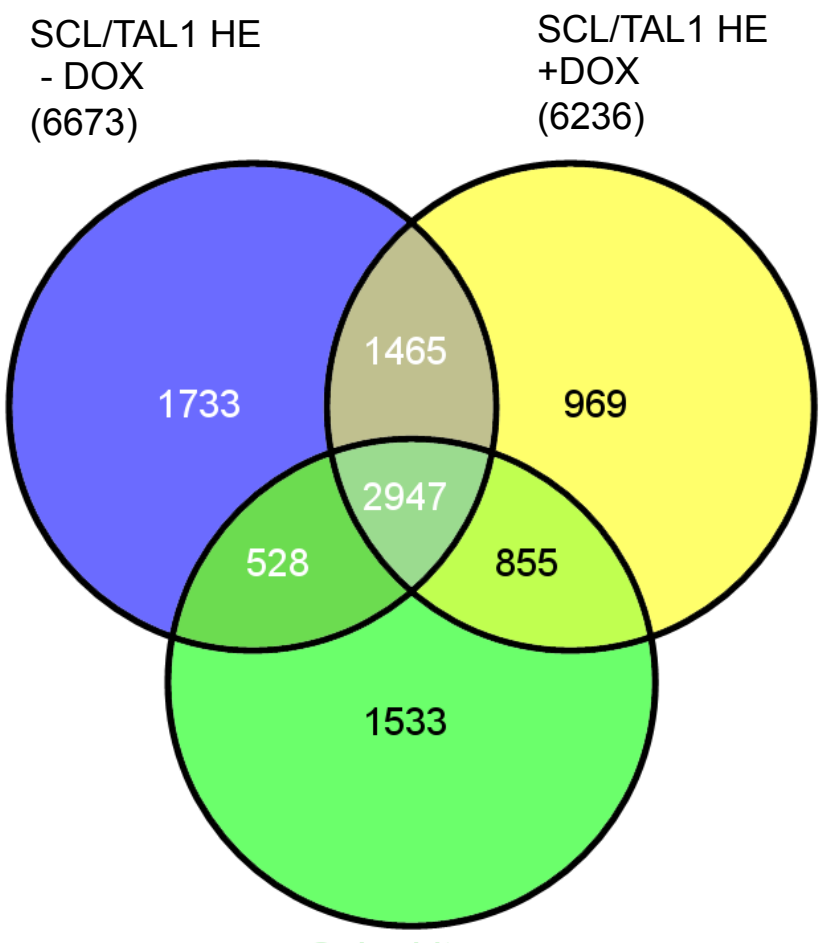




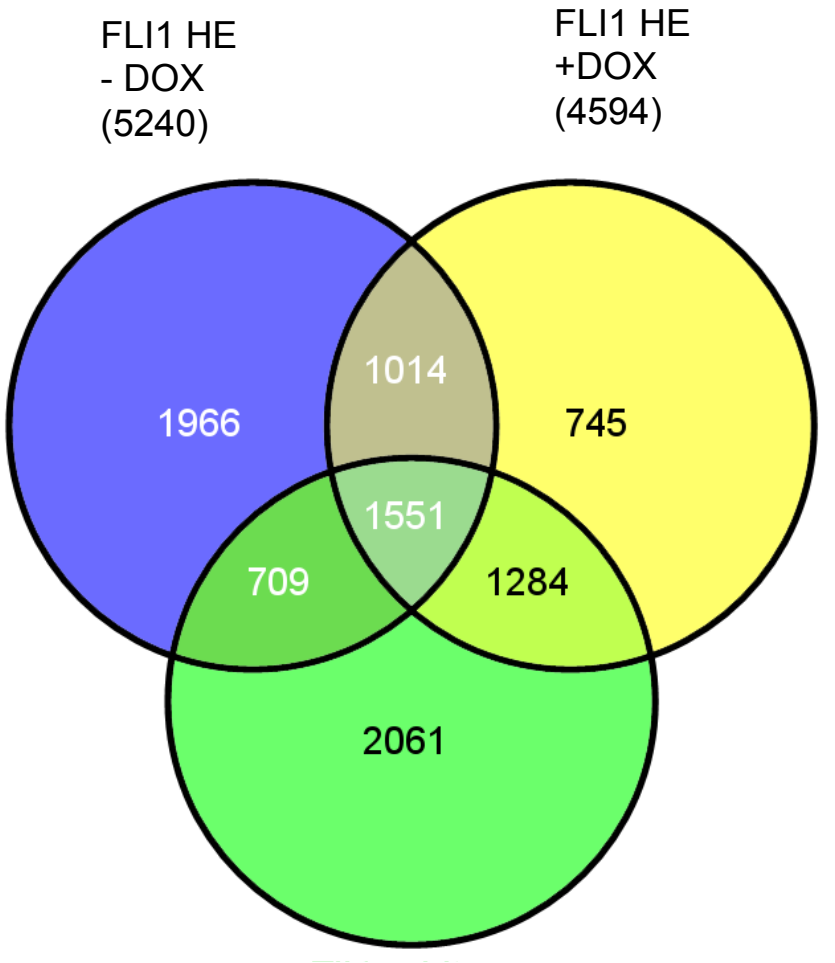
B



C



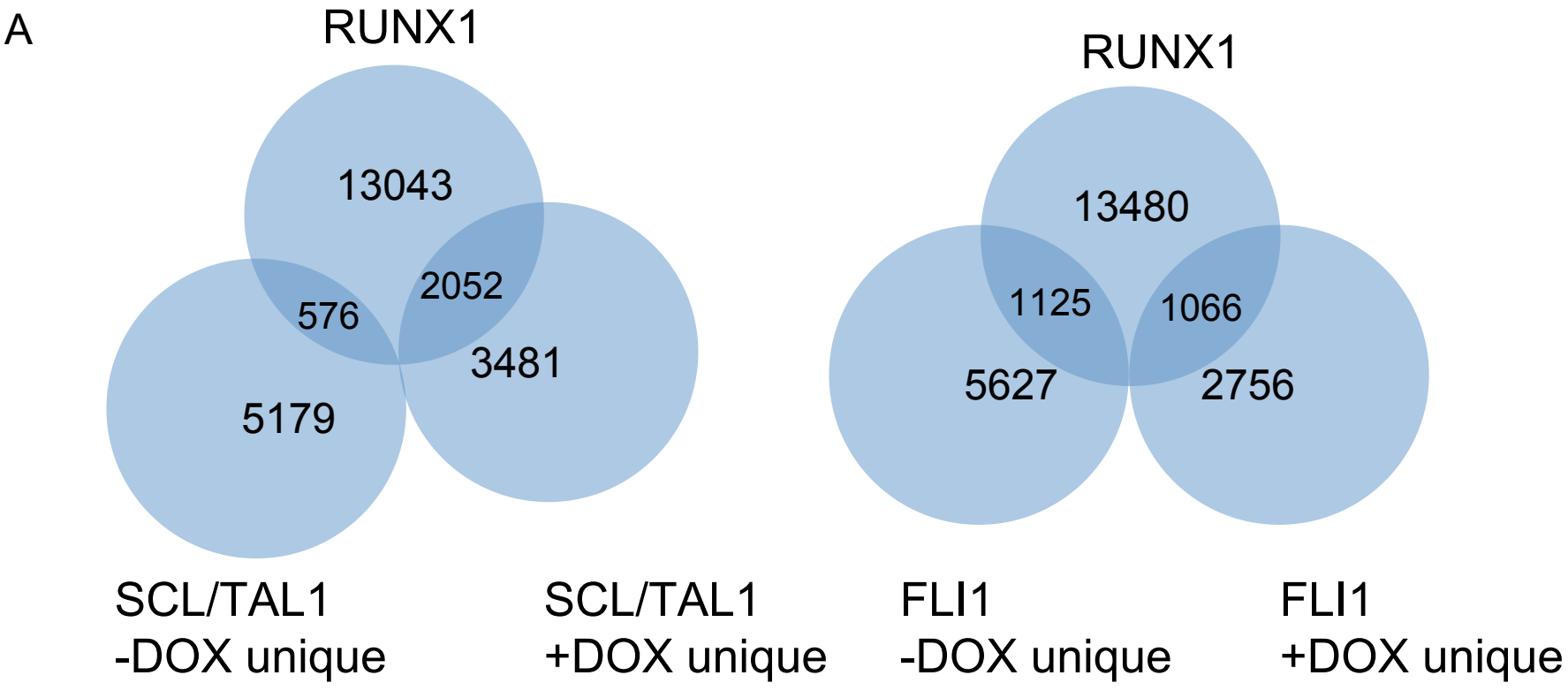
SCL/TAL1 c-kit+
(5863)



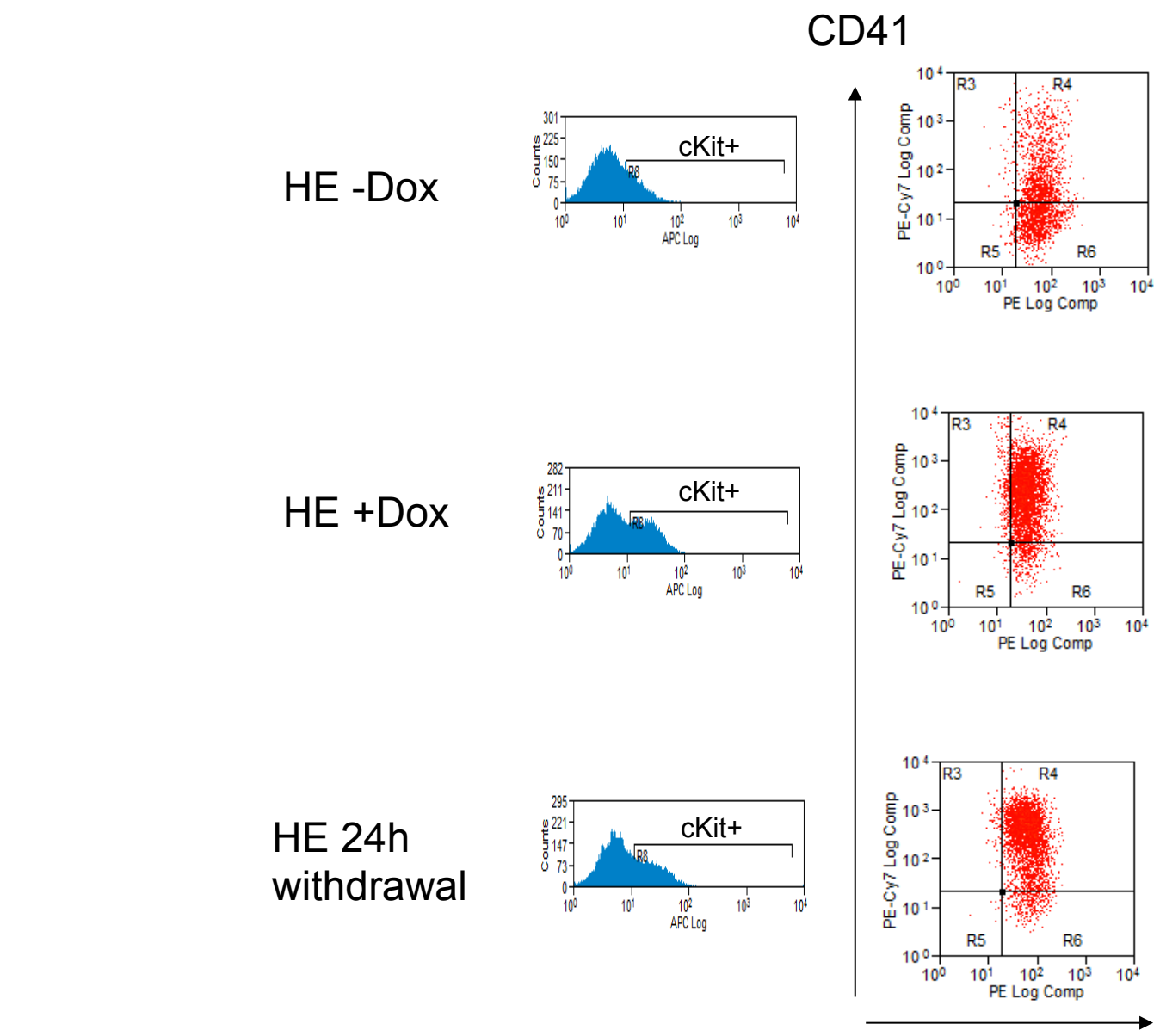
FLI1 c-kit+
(5605)

Supplementary Figure 8

A



B



C