Supplemental Figures for:

Human erythroleukemia genetics and transcriptomes identify master transcription factors as functional disease drivers

Alexandre Fagnan, Frederik Otzen Bagger, Maria-Riera Piqué-Borràs, Cathy Ignacimouttou, Alexis Caulier, Cécile K. Lopez, Elie Robert, Benjamin Uzan, Véronique Gelsi-Boyer, Zakia Aid, Cécile Thirant, Ute Moll, Samantha Tauchmann, Amina Kurtovic-Kozaric, Jaroslaw Maciejewski, Christine Dierks, Orietta Spinelli, Silvia Salmoiraghi, Thomas Pabst, Kazuya Shimoda, Virginie Deleuze, Hélène Lapillonne, Connor Sweeney, Véronique De Mas, Betty Leite, Zahra Khadri, Sébastien Malinge, Stéphane de Botton, Jean-Baptiste Micol, Benjamin Kile, Catherine L. Carmichael, Ilaria Iacobucci, Charles Mullighan, Martin Caroll, Peter Valent, Olivier A. Bernard, Eric Delabesse, Paresh Vyas, Daniel Birnbaum, Eduardo Anguita, Loïc Garçon, Eric Soler, Juerg Schwaller, Thomas Mercher

SUPPLEMENTAL FIGURES

Supplemental Figure 1.

Association with clinical AEL patient outcomes

(A) Patient distribution according to age (left): older adults (>60 years, n=34), adults (40-59 years, n=14), young adults (21-39 years, n=8) and pediatric patients (0-20 years, n=2) or according to WHO 2008 diagnosis (right): AML-M6a *de novo* (n=29), AML-M6a secondary (n=20), AML-M6b *de novo* (n=4) or AML-M6b (n=1).

(B) Variant allele frequency of recurrently mutated genes in AEL patients.

(C-F) Kaplan-Meier survival plots of AEL patients according to (C) genomic subgroup: Group1 (n=10), Group 2 (n=7), Group3 (n=6); (D) IPSS-R cytogenetic risk group: Very good/Good (n=24), Intermediate (n=4), Very poor/Poor (n=20); (E) WHO 2008 diagnosis: M6a *de novo* (n=29), M6b (n=5), M6a secondary (n=20); (F) age at diagnosis: Young (n=8), Adult (n=14), Older adult (n=34). p-value using Log rank Mantel-Cox test are indicated.

(G-H) Histogram representation of the number of mutations per patient according to genomic subgroups (G) in our AEL cohort [Group 1: n=12, Group 2: n=11, Group 3: n=10] and (H) in the Iacobucci *et al.* AEL cohort[33] [TP53: n=48, Others with epigenetic: n=25, Others without epigenetic: n=29]. The "TP53" and "Others" groups were extracted directly from Iacobucci *et al.*[33], while the sub-classification into "with epigenetic" and "without epigenetic" was a re-analysis of the data.

Supplemental Figure 2.

Hematopoietic lineage enrichment profile of AEL

(A) Heatmap representation of hematopoietic lineage enrichment score profile obtained using xCell software[34] on AEL RNAseq data.

(B) *SPI1* expression in AEL patients (normalized number of counts are represented). The red bar highlights sample that expressed more than 3-fold the highest expression observed in either normal human BFU-E or CFU-E.

Supplemental Figure 3.

Transcription factor activity inference

(A) Schematic representation of the generation of a transcription factor (TF) network using the ARACNe algorithm to compute a human HSPC network and the VIPER algorithm to compute the activity of TF based on our AEL patient transcriptome signatures.

(B) Heatmap representation of the top 50 most differentially transcription factors activated in AEL patient samples. Differential activated gene lists were established by PCA analysis using predicted activated gene matrix (previously computed using ARACNE and VIPER algorithm), then genes driving PCA dimensions were identified and ranked by contribution (using FactoMineR v1.41 and factoextra v1.0.5 R packages). Finally, heatmap of activated genes was obtained by plotting the top 50 most contributed genes from the first PCA dimension (using pheatmap v1.0.12 R package).

(C) PCA with regression lines from plot in Figure 3A with projection of AEL patient samples colored with SPI1, NFIA and TAL1 expression and predicted activity.

Supplemental Figure 4.

Fusion transcripts in human AEL

(A) Heatmap representation of patients carrying out-of-frame (green) or in-frame (red) fusion transcripts according to their molecular subgroup.

(B) Sequences of the breakpoint *YWHAE-EPO* and *HSD17B11-B4GALNT3* fusion transcripts. The number of reads supporting the fusion are indicated.

(C) Paired histogram representation of *YWHAE* and *EPO, HSD17B11* and *B4GALNT3, LINC01* and *DNMT3B* gene expression in AEL patients, normal human BFU-E and CFU-E. Patients with red bars indicate those presenting with the fusion transcripts.

Supplemental Figure 5.

ERG, ETO2, SKI and SPI1 impairs chromatin accessibility at GATA1 binding site and functionally impairs GATA1 activity

(A) Flow cytometry analysis of CD71 and Ter119 expression (top) and May Grunwald Giemsa stained cytospots (bottom) of mouse erythroblast transformed with either expression of *SKI, ERG, ETO2, SPI1* and *B4GALNT3* overexpression or with cooperation between *Tet2*$^{-/-}$ + *G1s.*

(B) Histogram representation of the foldchanges expression of human AEL samples (fold-changes computed by comparing impacted patient sample to the mean of non-impacted patient samples) and transduced mouse erythroblast (fold-changes computed by comparing cells transduced with the indicated factor to Ctrl) for *SKI*, *ERG*, *ETO2*, *SPI1*, *B4GALNT3* and *EPO* mRNA expression.

(C) Histogram representation of *SKI, ERG, ETO2, SPI1* and *B4GALNT3* mRNA expression measured by RT-qPCR in normal mouse erythroblast or erythroblast transformed with either expression of *SKI, ERG, ETO2, SPI1* and *B4GALNT3* overexpression (n=3 for all conditions). Expression levels were also compared to the erythroleukemia MEL or the Ba/F3 myelo-lymphoid cell line.

(D) Dot-plot representation of motifs analysis under ATAC-seq peaks of normal (Ctrl) or transformed erythroblast (ERG, ETO2, SKI and SPI1). Each dot represents the log(p.value) for a given motif and increased ranked in abscise by log(p.value)

(E) Heatmap representing the hierarchical clustering of ATAC-seq signal, performed using normal (Ctrl) and transformed erythroblast by either ERG, ETO2, SKI or SPI1, focused on specific open chromatin region of MEP (top), CFU-E (middle) and pro-erythroblast (proE; bottom). Heatmaps were focused on peaks centers with +/-5Kb.

(G) Quantitative *Gata1* mRNA expression measured by RT-qPCR in WT erythroblast (Ctrl) or erythroblast transformed with either *SKI, ERG, ETO2, SPI1* or *B4GALNT3* overexpression. Expression levels were also compared to the erythroleukemia MEL or the Ba/F3 myelo-lymphoid cell line (n=3 for all conditions).

(G) Flow-cytometry analysis histogram of Ter119 expression in G1E cells expressing empty control without (grey) or with (red) induction of GATA1 expression by doxycycline induction.

Supplemental Figure 6.

Transplantation of *in vitro*-transformed erythroblasts

(A) Kaplan-Meier plot of mice transplanted with mouse erythroblasts retrovirally expressing *SKI*, *ERG, ETO2, SPI1or B4GALNT3*, expanded over 2 months *in vitro*. n=4 for all groups. Only erythroblast expressing ERG, ETO2 or SKI were able to engraft in mice and promote disease.

(B) Flow cytometry analysis of CD71 and Ter119 (top) or CD11b and Gr1 (bottom) surface expression, gated on transgene-expressing cells (GFP^+), in mice injected with 10^6 mouse erythroblasts expressing *SKI, ERG* or *ETO2*.

(C) Flow cytometry analysis of myeloid cells ($CD11b^+Gr1^+$), erythroid progenitors ($CD71^+$ $Ter119^+$), B cells ($B220^+$ $CD19^+$) and T cells ($CD4^+CD8^+$) gated for $CD45.2^+$ cells in the BM of immunodeficient recipients ($CD45.1^+$) engrafted with $CD45.2^+$ erythroblasts from *Tet2*$^{-/-}$+*G1s* mice.

(D) Histopathological analysis of BM and spleen of mice engrafted with normal erythroblasts or transformed erythroblast either expressing SKI, *ERG* or *ETO2* or from *Tet2*$^{-/-}$ + *G1s* mice, previously expanded *in vitro* over 2 months. Staining were performed using H&E.

(E) Histopathology analysis of Bone marrow (BM, left) and spleen (rigth) of $CD45.1^+$ mice engrafted with transformed erythroblast expressing an empty vector (top) or SKI (bottom).

Supplemental Figure 7.

Modeling AEL-associated oncogenic cooperation in mice

(A) Flow cytometry analysis of myeloid cells ($CD11b^+Gr1^1$), erythroid progenitors ($CD71^+$ $Ter119^+$) and erythroid differentiation (forward scatter area: FSC-A, CD44) in BM of WT, *Tet2*$^{-/-}$+*Gata1*$^{+/Y}$ (*Tet2*$^{-/-}$), *Tet2*$^{+/+}$+*Gata1*$^{Gata1s/Y}$ *(Gata1s)* or *Tet2*$^{-/-}$+*Gata1*$^{Gata1s/Y}$ (*Tet2*$^{-/-}$+*Gata1s)* male transgenic mice at 8 months.

(B) Histogram representation of spleen weight (left) and spleen photograph (right) of recipient mice engrafted with HSPC from *Tet2*$^{-/-}$ (n=5) and *Tet2*$^{-/-}$+*Gata1s* (n=4) mice.

(C) Histopathology of BM and spleen sections from recipient mice engrafted with HSPC from *Tet2*$^{-/-}$+*Gata1s* mice.

(D) Kaplan-Meier plot of lethally irradiated mice engrafted with HSPC cells from either WT or TP53^{R248Q} mice transduced with ERG or empty vector (Ctrl). WT: n=5, TP53^{R248Q}: n=6.

(E) GFP+ myeloid cells (Gr1^{+}CD11b^{+}), erythroid progenitors (CD71^{+} Ter119^{+}), B cells (B220^{+}) and T cells (CD4^{+}CD8^{+}) in recipients of WT or TP53^{R248Q} HSPC transduced with *ERG* or empty vector (*Ctrl*) at pre-leukemic or leukemic state.

(F) Histopathology of spleen and liver sections from secondary recipient mice engrafted with GFP^{+} erythroblast from primary recipient mice previously engrafted with HSPC from *TP53*^{R248Q} expressing ERG.

Supplemental Figure 8.

SKI overexpression in HSPC leads to a myeloid/erythroid disease in mice

(A) GFP^{+} cells (%) in BM, and Kaplan-Meier plot of lethally irradiated mice transplanted with 0.5x10^{6} lineage marker-depleted HSPC transduced with either murine *SKI* cDNA (n=5) or empty vector (*Ctrl*; n=5).

(B) Peripheral white blood cells (WBC), red blood cells (RBC), platelets (PLT), lymphocytes (Lym), monocytes (Mono), neutrophils and granulocytes (N/Gr) counts. Mean+/-SD is indicated (n=4).

(C) GFP+ myeloid cells (CD11b^{+}Gr1^{+}), erythroid progenitors (CD71^{+} Ter119^{+}), B cells (B220^{+}CD19^{+}) and T cells (CD4^{+}CD8^{+}) gated in the BM of mice transplanted with SKI or empty vector (Ctrl) transduced cells.

(D) GFP+ (%) myeloid cells (CD11b^{+}Gr1^{+}), erythroid progenitors (CD71^{+}Ter119^{+}), B cells (B220^{+}CD19^{+}) and T cells (CD4^{+}CD8^{+}) in the BM or spleen of mice transplanted with *SKI* or control (Ctrl) transduced cells. Mean+/-SD is indicated (Ctrl: n=5; SKI: n=5).

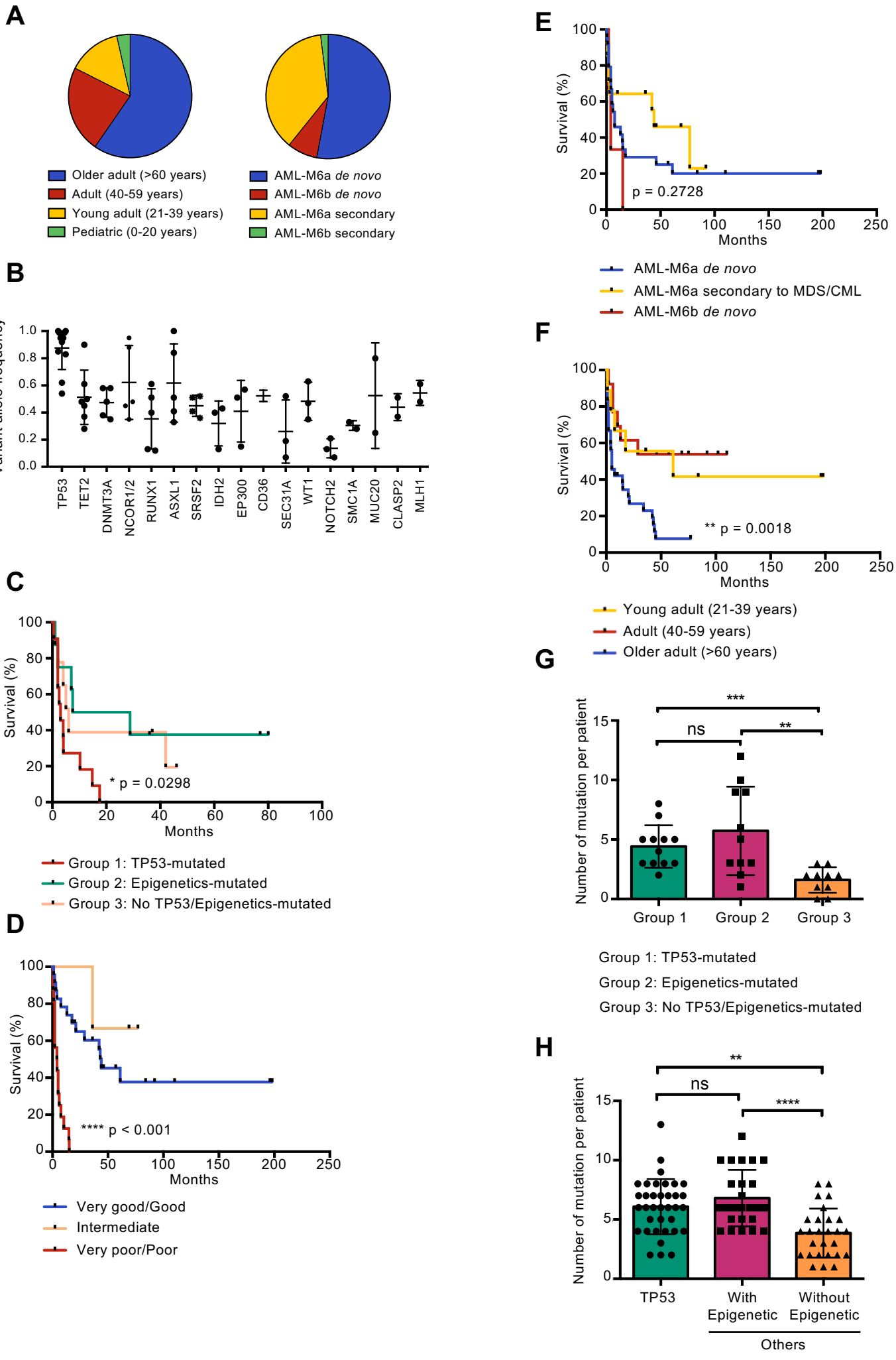
(E) Number of GFP^+ and negative myeloid cells ($CD11b^+Gr1^+$), erythroid progenitors ($CD71^+Ter119^+$), B cells ($B220^+CD19^+$) and T cells ($CD4^+CD8^+$) in BM or spleen of primary recipients. Mean+/-SD is indicated (Ctrl: n=5; SKI: n=5).

(F) Flow cytometry analysis of GFP^+ cells in peripheral blood: platelets (PLT), red blood cells (RBC) and white blood cells (WBC) of primary engrafted mice

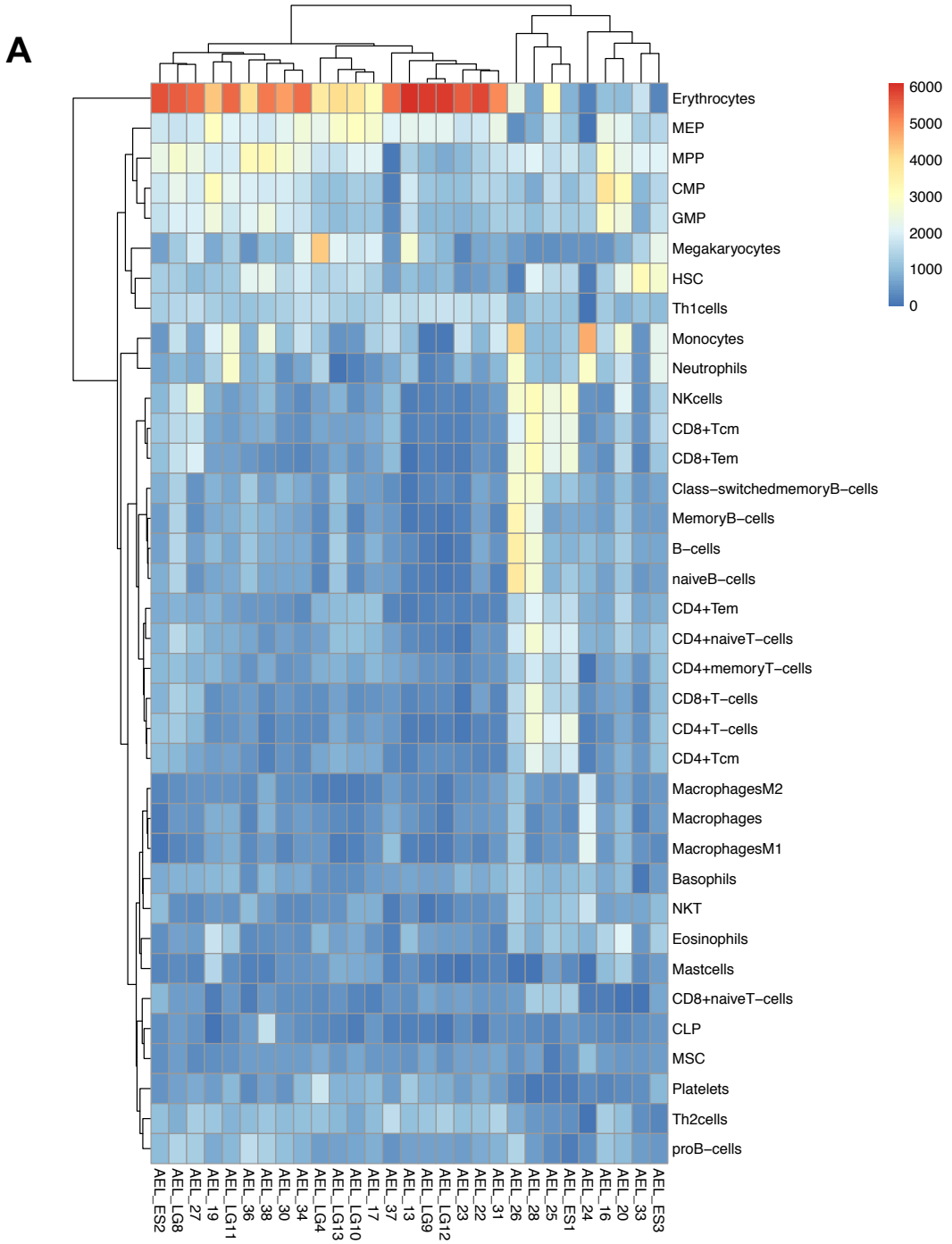
(G) GFP^+ (%) cells in peripheral blood (WBC, RBC, PLT) and BM in recipient mice engrafted with HSC, MEP and GMP expressing *SKI* or empty vector (*Ctrl*), 3 weeks post-engraftment. Mean+/-SD is indicated (HSC: n=5, GMP: n=5, MEP: n=4).

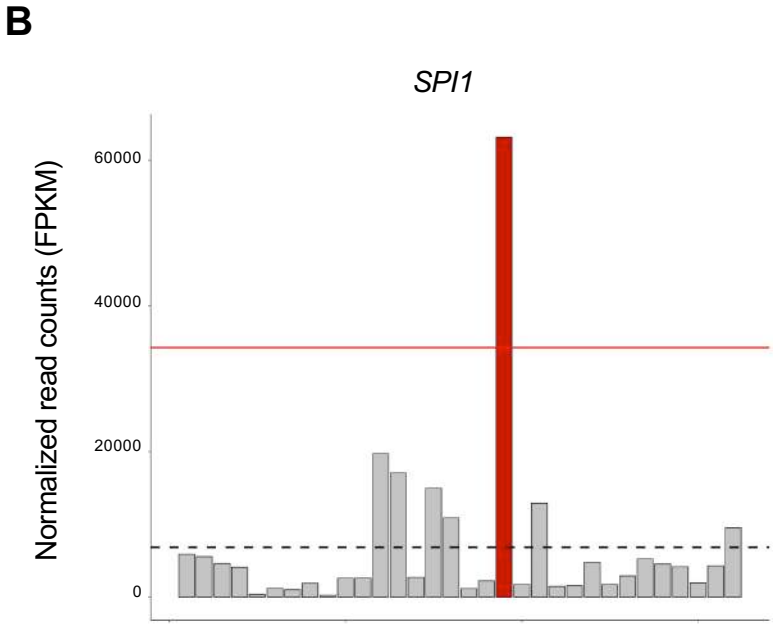
(H) Peripheral white blood cells (WBC), red blood cells (RBC) and platelet (PLT) counts in mice transplanted with HSC or MEP expressing *SKI* or empty vector (*Ctrl*), at time of sacrifice. Mean+/-SD is indicated (HSC: n=5; MEP: n=5).

(I) Histopathology of BM and spleen of MEP expressing *SKI* engrafted mice stained by HE or an GATA1 antibody.

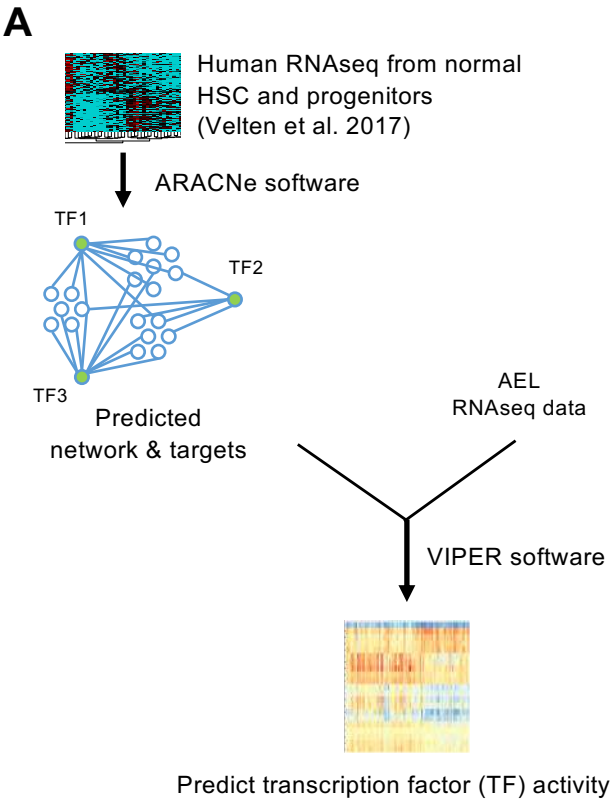


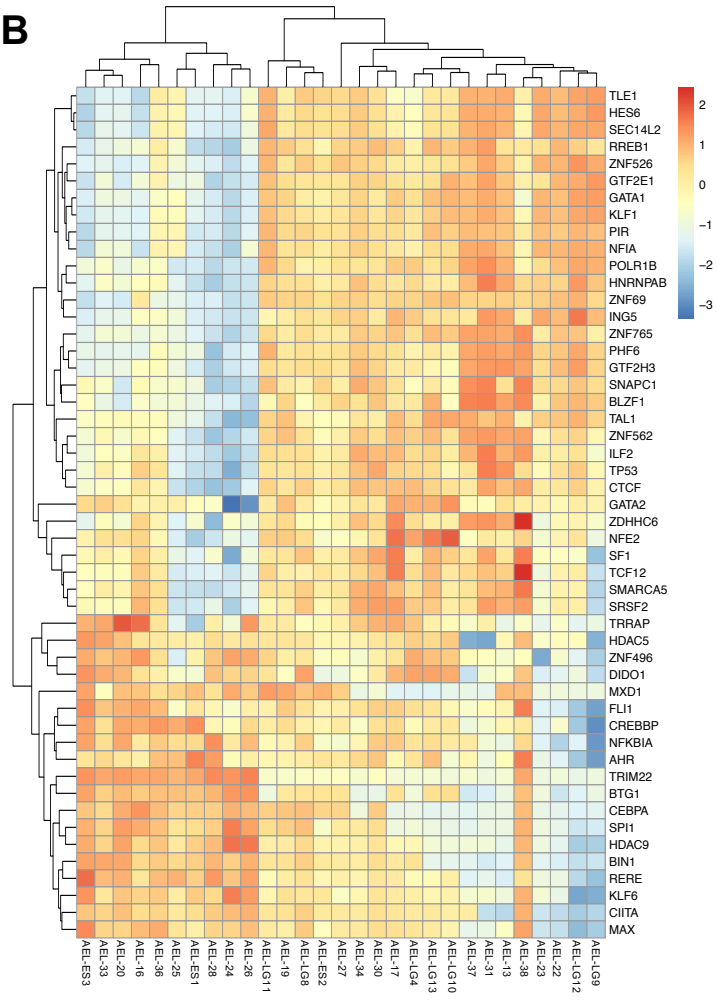
Supplemental Figure 1



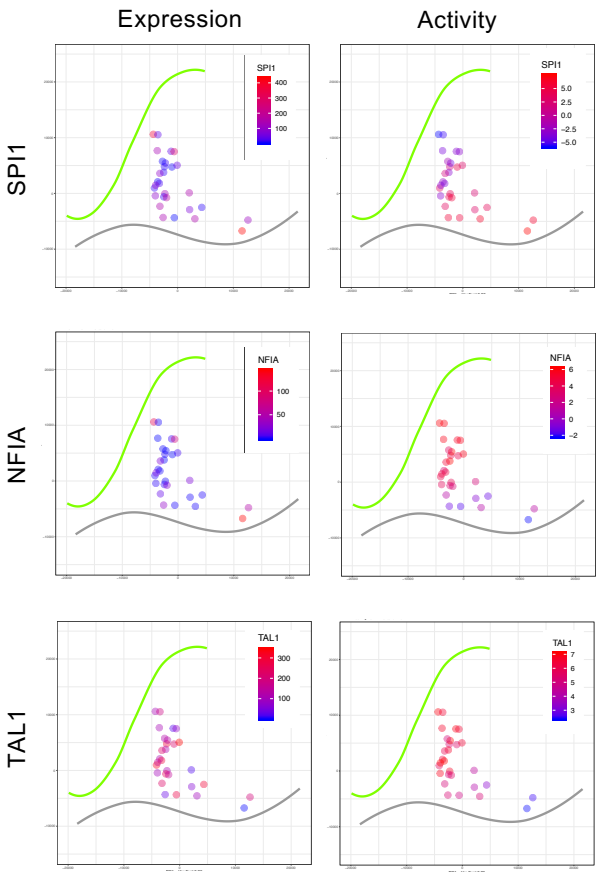


Supplemental Figure 2

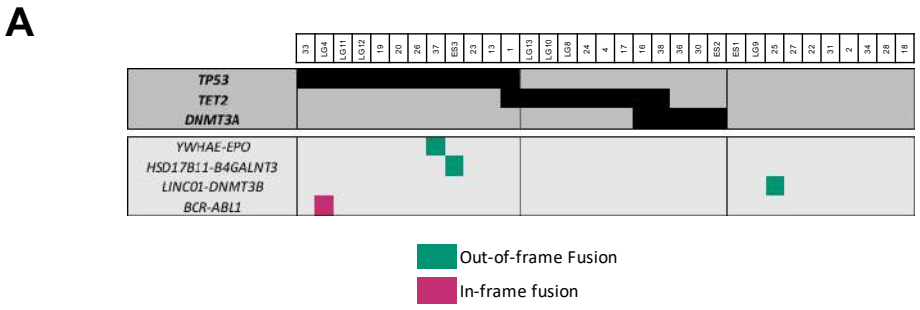


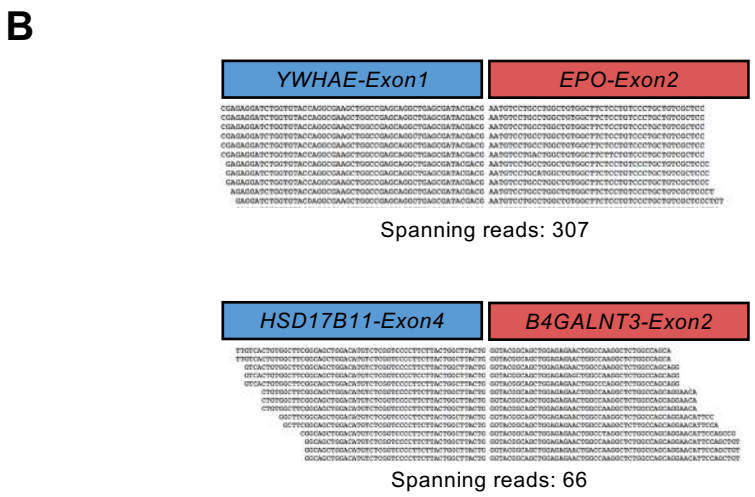


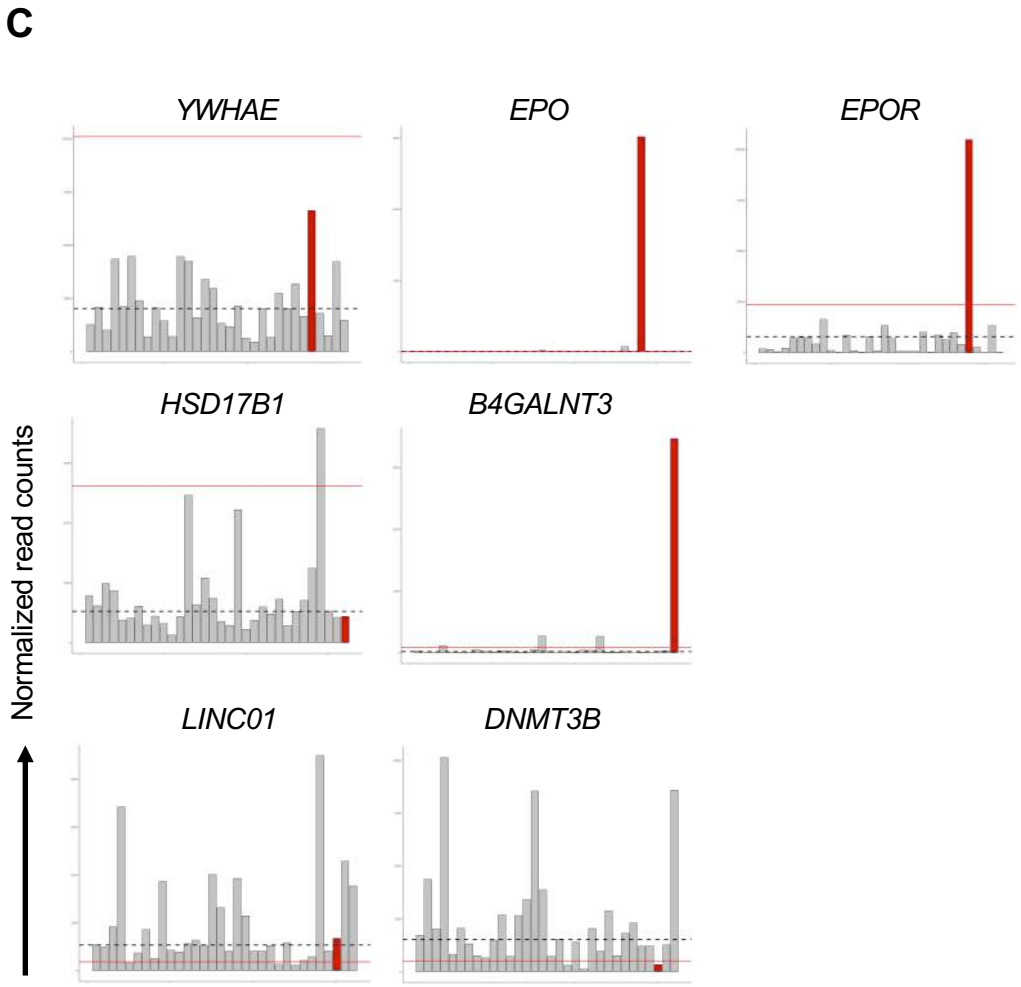
50 most differentially activated TF



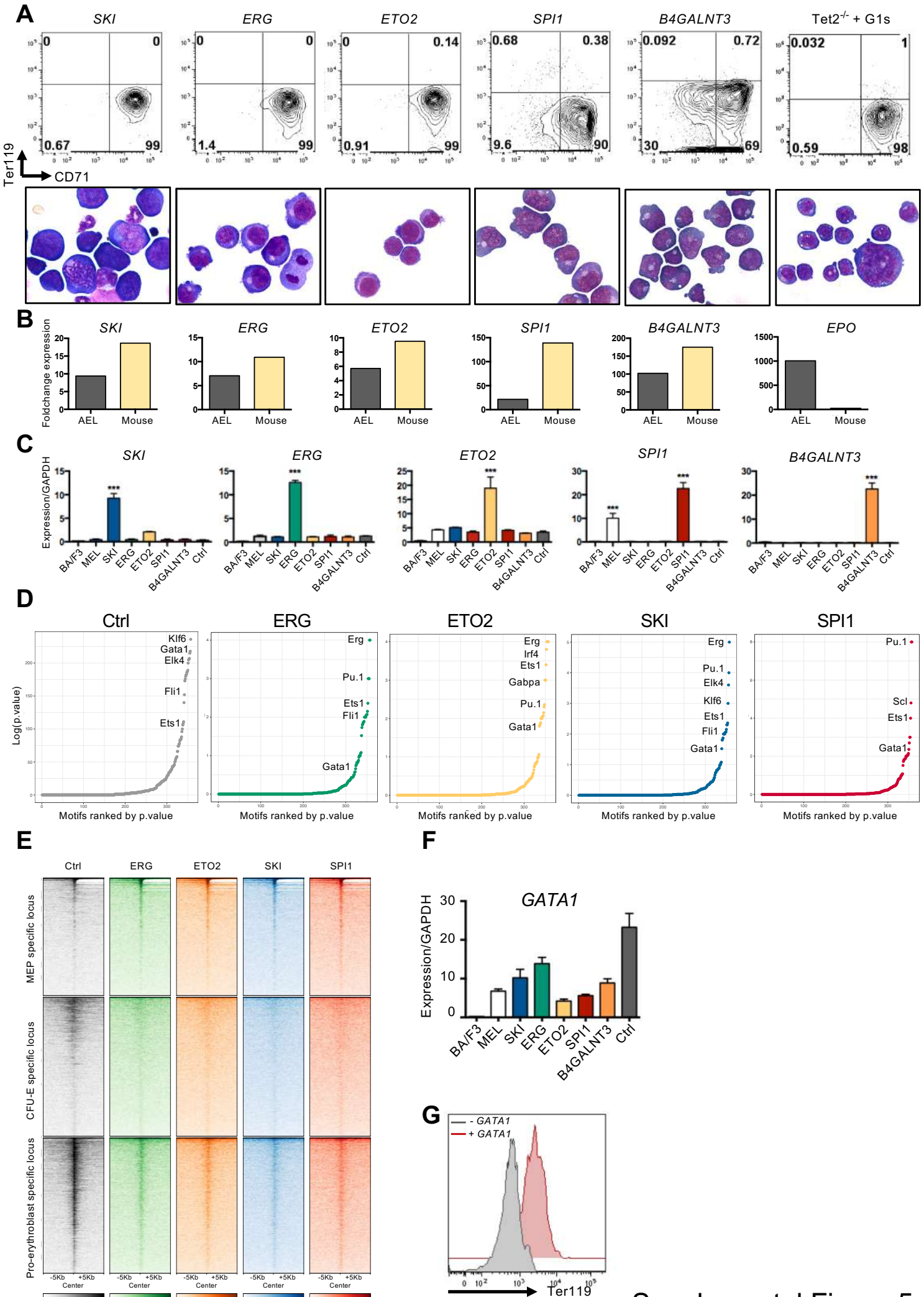
Supplemental Figure 3



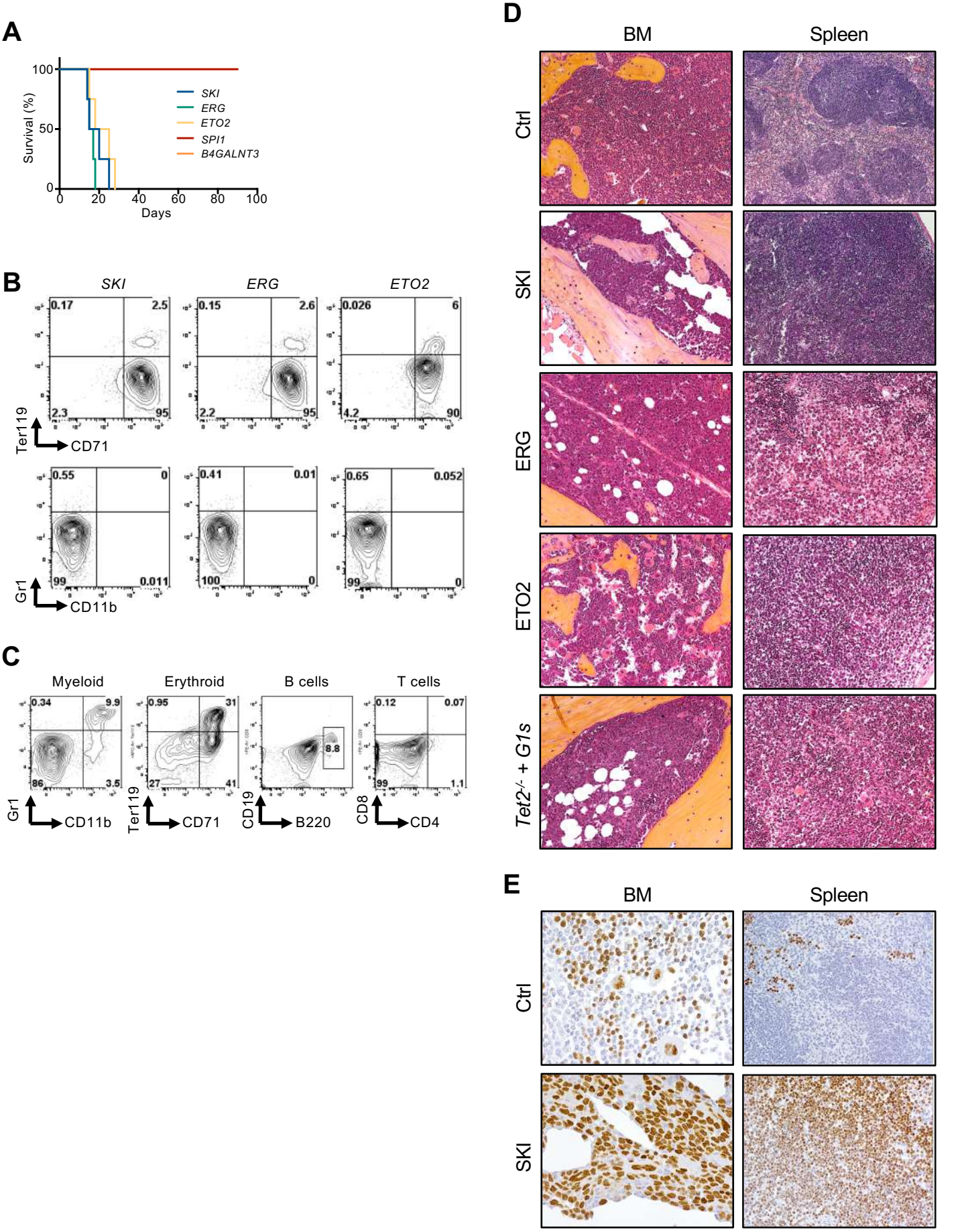




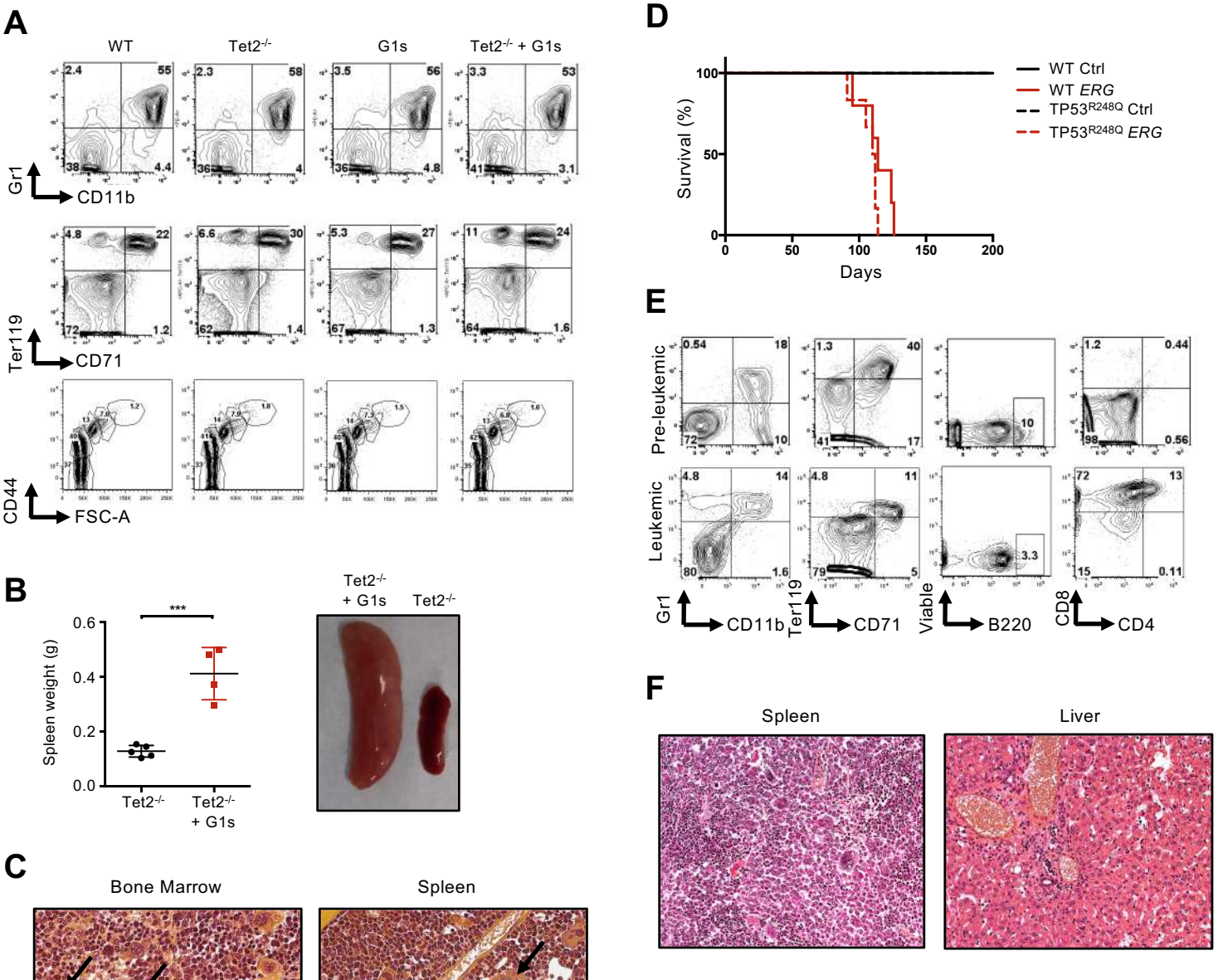
Supplemental Figure 4



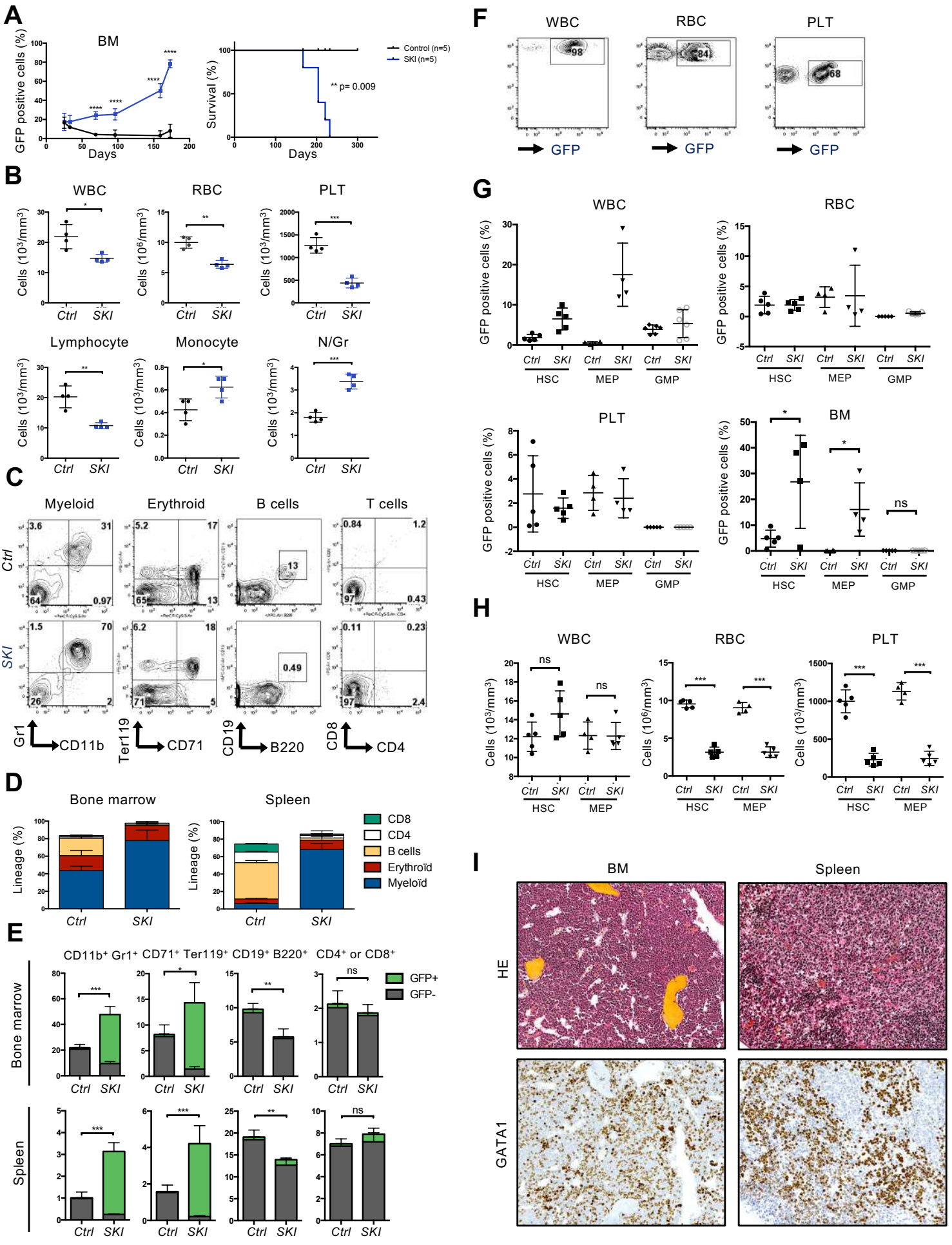
Supplemental Figure 5



Supplemental Figure 6



Supplemental Figure 7



Supplemental Figure 8