SUPPLEMENTAL MATERIALS AND METHODS

Cell culture

HEK293T cells were cultured in DMEM containing 10% bovine calf serum, 10 mM glutamine, and penicillin (100 IU)/streptomycin (100 µg/ml). CRISPR-edited K562 cells were published previously[1] and cultured in RPMI with 10% fetal bovine serum, 10 mM glutamine, and penicillin (100 IU/ml)/streptomycin (100 µg/ml). For proliferation assays, K562 cells were seeded in duplicate in three independent experiments. Relative growth was evaluated by Trypan Blue exclusion.

Flow cytometry

Primary cells were collected from mice, treated in ammonium-chloride-potassium (ACK) buffer (150 mM NH_4Cl, 10 mM $KHCO_3$, 0.1 mM Na_2-EDTA), and stained. The following antibodies were used: PerCP-Cy5.5 or PE-Cy5-conjugated lineage antibodies (CD3, CD4, CD8a, B220, CD19, Gr1, CD11b, and Ter119), APC or PE-Cy5-conjugated Sca-1, PE-Cy7 or APC-conjugated c-Kit, PE-conjugated CD48, biotin-conjugated CD150 with streptavidin-conjugated APC-Cy7, PE-conjugated FcγR, FITC-conjugated CD34, and PerCP-Cy5.5 or PE-conjugated IL7Rα (all from Biolegend). For the apoptosis assay, Annexin V-APC (BioLegend) and 7AAD (BioLegend) were used. $AnnexinV^+7AAD^-$ cells were considered apoptotic. Cells were analyzed on a BD FACS Canto flow cytometer equipped with standard lasers and filters (Becton, Dickinson and Company (BD); Franklin Lakes, NJ). Data were analyzed on the BD FACS Diva Software and FlowJo (BD).

RNA-sequencing and data analysis

Total RNA from mouse LK ($Lin^{-}cKit^{+}$) cells and human K562 cells were extracted with Trizol reagent (Thermo Fisher Scientific). PolyA-selected RNA was used for library construction using the TruSeq RNA Sample Preparation Kit (Illumina) according to the manufacturer's protocol. Libraries were sequenced on a HiSeq4000 (Illumina; San Diego, CA) at the UCSD Institute for Genomic Medicine (IGM) core facility (mouse samples) or on a NovaSeq6000 (Illumina) at the UCSF core facility (K562 samples). RNA-seq reads from LK and K562 cells were mapped to the GRCm38 and GRCh38 reference genomes, respectively, using the STAR software package[2]. DEseq2 was used to perform differential gene expression analysis[3]. In K562 cells, differentially expressed genes (DEGs) were defined as those with fold change > 2 or < 0.5 and adjusted p-value < 0.05. Altered splicing events in K562 cells were identified using rMATS[4] by requiring at least 5% inclusion ratio difference (-cstat) with a false discovery rate (FDR) < 0.05. To obtain a comparable number of DEGs and altered splicing events in LK cells vs K562 cells, we used modified cutoffs. DEGs in LK cells were defined as those with fold change > 1.5 or < 0.667 and adjusted p-value < 0.05, and altered splicing events were selected if the inclusion ratio difference > 5% and FDR < 0.01.

Gene list enrichment analyses were all performed using the clusterProfiler R package[5]. The top 30 enriched GO terms were visualized with the enrichplot R package to connect functionally related GO terms into clusters. The line width is positively correlated with the number of shared genes by two connected GO terms.

The frequencies of individual k-mers (k = 4, 5 or 6) were counted in repressed and promoted exons for quantitative comparison. A Wilcoxon test p-value cutoff of 0.01 was used to define k-mers showing significant enrichment in repressed or promoted exons.

Western blotting

Protein samples were denatured in 1× SDS loading buffer (10% glycerol, 2% SDS, 10 mM DTT, and 50 mM Tris-HCl [pH 6.8]). Protein concentrations were adjusted, and samples were loaded on SDS polyacrylamide gels after adding bromophenol blue (0.05%). Primary anti-RUNX1 antibody (4434S; Cell Signaling) and anti-β-actin antibody (A1978; Sigma-Aldrich) were used. Signals from fluorophore-conjugated secondary antibodies were detected with the LI-COR Odyssey Classic Infrared Imaging System (LI-COR Biosciences; Lincoln, NE).

Lentivirus preparation and infection

MISSION pLKO.1-shRNA-puro lentiviral vector constructs targeting human RUNX1 (TRCN0000013659 and TRCN0000013660) were obtained from the Functional Genomics Center at La Jolla Institute for Immunology. For virus production, HEK293T cells were transfected with pLKO.1 lentiviral vectors, psPAX2, and pMD2.G using polyethylenimine (Polysciences Inc.) for 10 hours and then the medium was changed from DMEM to RPMI. 48 hours later, supernatant was collected, filtered through a 0.45μm filter, and added to K562 cells, along with 1% HEPES and 0.1% polybrene (final concentration 4 μg/ml). The cells were spinoculated at 1200g for 3 hours at 32°C. Infections were performed twice on consecutive days.

Reverse transcription and polymerase chain reaction (RT-PCR)

Total RNA was extracted using Trizol reagent (Thermo Fisher Scientific) according to the manufacturer's instructions. RT reactions were carried out by using the First Strand cDNA Synthesis Kit (MCLab). PCR was performed using KOD Hot Start DNA Polymerase (EMD Millipore). Primers used for PCR are as follows:

BRIP1(F): 5’-GGAAACCAGCAGATGAGGGCGT-3’

BRIP1(R): 5’-TTTCTGTGGCGAAAAGGAGTTT-3’

FAN1(F):5’-GGCTCTTTCAACGTAAATTAAGCTGG-3’

FAN1(R): 5’-CTCCATTCGGCCGAGGTTGACC-3’

NABP1(F): 5’-TCCATGTGGAAAGGATGTCTGACA-3’

NABP1 (R): 5’-CCCTTTGTTCTGCTGTCCTCGA-3’

TBRG4(F): 5’-TCGGAGCCACTAATGAACCGCC-3’

TBRG4 (R): 5’-CGCGCAAAAGCCAGAAGTACGC-3’

PFKM-1(F): 5’-CTGGGGAAGCTTCTACTTCC-3’

PFKM-1 (R): 5’-CTTGGGCATCTCCACCAGAG-3’

PFKM-2(F): 5’-TGAGGGTGCAATTGACAAGA-3’

PFKM-2 (R): 5’-GCCTTGGTCACATCTTTGGT-3’

EXOSC9 (F): 5’-TGGAACTGCCCAAATTGGAGAGGG-3’

EXOSC9 (R): 5’-TCTTTGGATTCTTGTCTGGTTCCA-3’

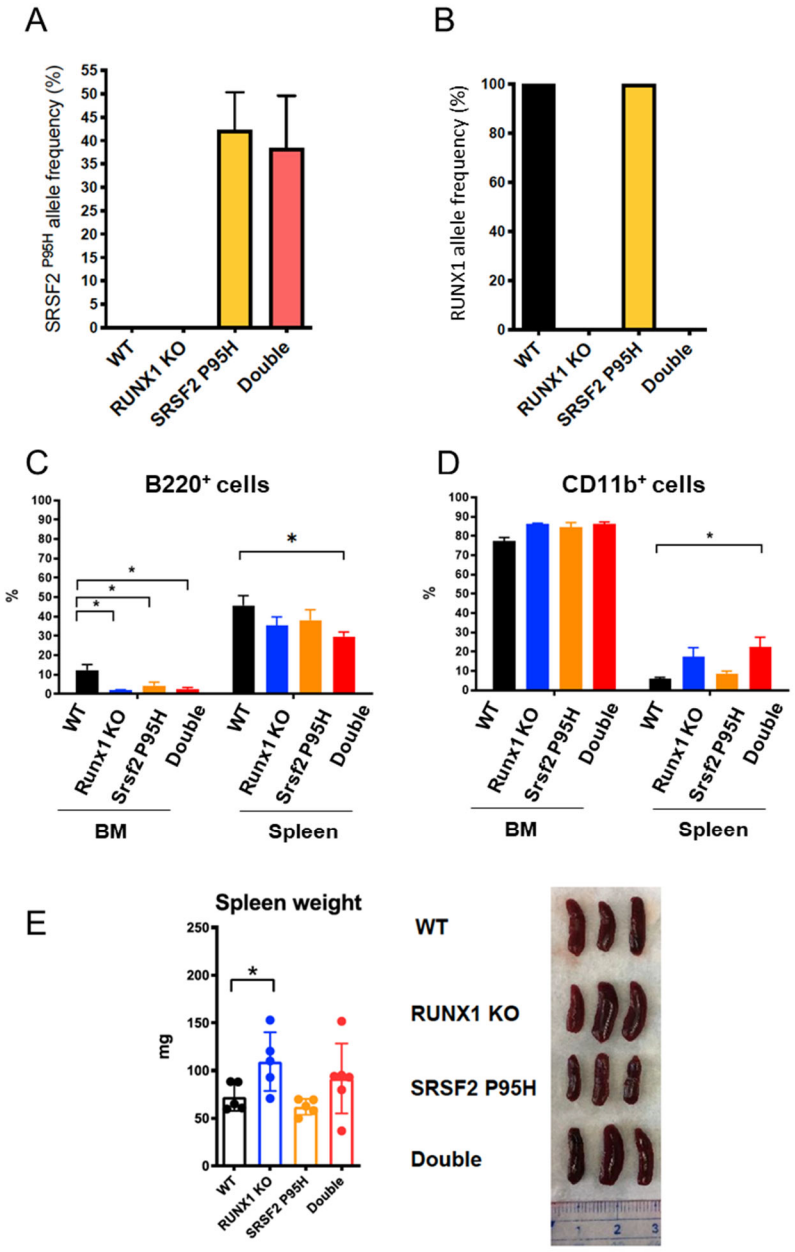
AKAP8L (F): 5’-GCTACGAGGGCTATGGCTATGGC-3’

AKAP8L (R): 5’-TCTCCACCTGAGCCGTACACGC-3’

Statistics

Results from repeated experiments were represented as mean ± standard error of the mean (SEM) or standard deviation (SD). Student’s t-tests, one-way ANOVAs, or Tukey post hoc tests were performed as indicated in specific figure legends to determine the statistical significance of the data collected. Whenever asterisks are used to indicate the statistical significance: * $p < 0.05$; ** $p < 0.01$; and *** $p < 0.001$.

SUPPLEMENTAL FIGURES



Supplementary figure 1. Characterization of single and double mutant mice

(A) Srsf2 P95H allele frequency in mouse bone marrow LK cells quantified by RNA-seq four weeks after pIpC activation.

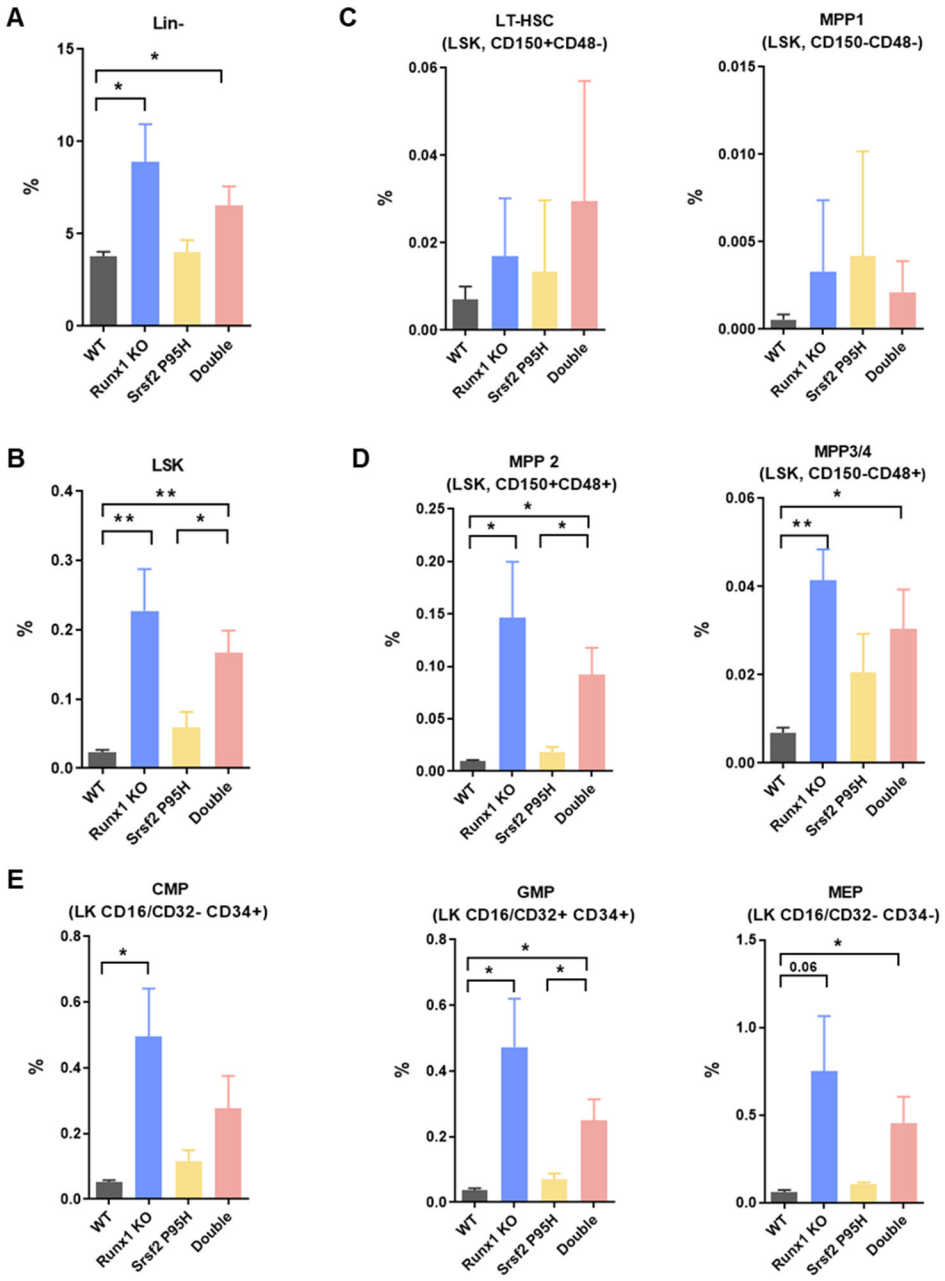
(B) Runx1 allele frequency in mouse bone marrow LK cells quantified by RNA-seq (exon 4 excision) 4 weeks after pIpC activation.

(C) Frequencies of B cells ($B220^{+}$) in the bone marrow (BM) and spleens of mice at 20 weeks post-transplantation (WT, n = 7; $Srsf2^{P95H/+}$, n = 8; Runx1 KO, n = 6; Double mutant, n = 7).

(D) Frequencies of myeloid cells ($Cd11b^{+}$) in the bone marrow (BM) and spleens of mice at 20 weeks post-transplantation (WT, n = 7; $Srsf2^{P95H/+}$, n = 8; Runx1 KO, n = 6; Double mutant, n = 7).

(E) Weights and a representative photo of spleens (SP) from mice transplanted with each genotype at 20 weeks post-bone marrow transplantation (BMT).

Data are mean +/- the standard error of the mean (SEM). Significance was determined by one-way ANOVA with Tukey post hoc test. * indicates $p < 0.05$, ** indicates $p < 0.01$



Supplementary figure 2. Characterization of hematopoietic stem and progenitor cells in spleens of single and double mutant mice

Percentages of hematopoietic stem and progenitor populations from mouse spleens analyzed by flow cytometry 20 weeks post-BMT (WT, n = 5; $Srsf2^{P95H/+}$, n = 5; Runx1 KO, n = 5; Double mutant, n = 6).

(A) $Lineage^-$ (Lin^-) cells.

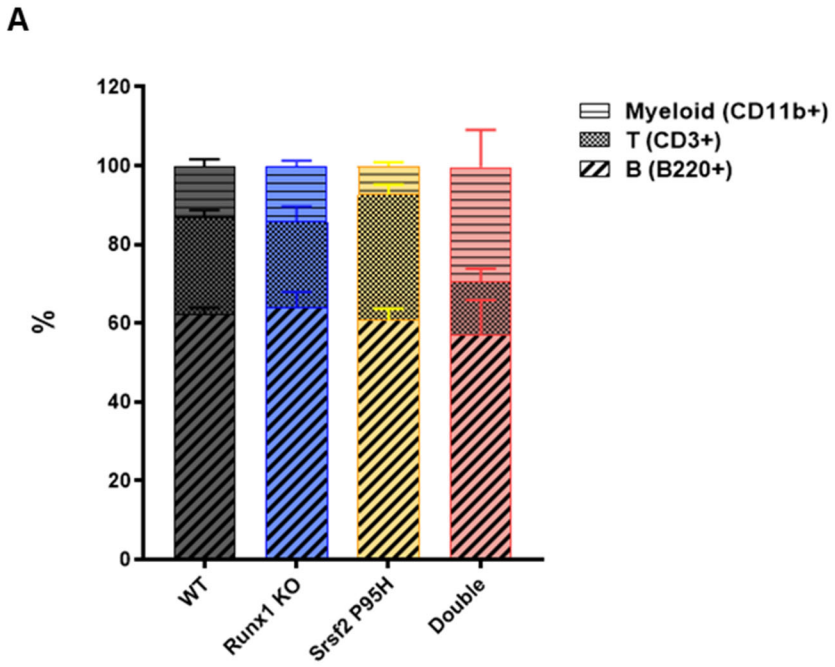
(B) $Lin^-Sca\text{-}1^+c\text{-}Kit^+$ (LSK) cells.

(C) Long-term hematopoietic stem cells (LT-HSC, LSK $CD150^{+}CD48^{-}$) and multipotent progenitor 1 cells (MPP1, LSK $CD150^{-}CD48^{-}$).

(D) Multipotent progenitor 2 (MPP2, LSK $CD150^{+}CD48^{+}$) and multipotent progenitor 3/4 cells (MPP3/4, LSK $CD150^{-}CD48^{+}$).

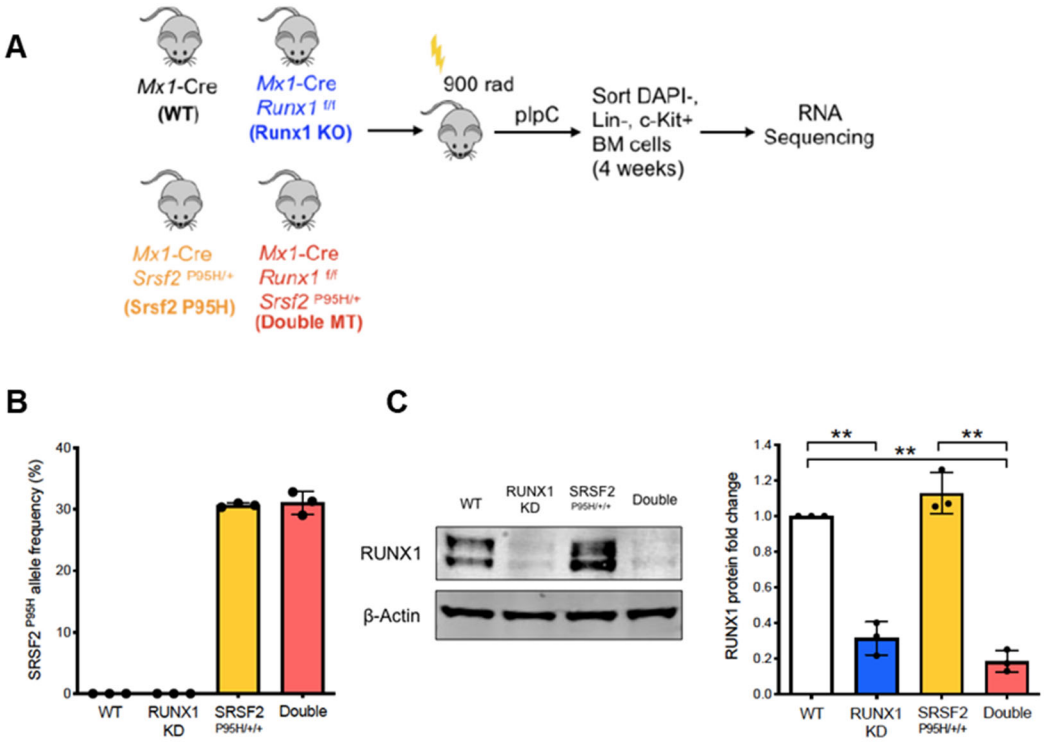
(E) Common myeloid progenitors (CMP, $Lin^{-}Sca\text{-}1^{-}c\text{-}Kit^{+}CD16/CD32^{-}CD34^{+}$), granulocyte-monocyte progenitors (GMP, $Lin^{-}Sca\text{-}1^{-}c\text{-}Kit^{+}CD16/CD32^{+}CD34^{+}$), and megakaryocyte-erythrocyte progenitors (MEP, $Lin^{-}Sca\text{-}1^{-}c\text{-}Kit^{+}CD16/CD32^{-}CD34^{-}$).

Data are mean +/- the standard error of the mean (SEM). Significance was determined by one-way ANOVA with Tukey post hoc test. * indicates $p < 0.05$, ** indicates $p < 0.01$.



Supplementary figure 3. Composition of CD45.2^{+} competitor cells in competitive bone marrow transplantation experiments

(A) Percentages of T cells (CD3^{+}), myeloid cells (CD11b^{+}), and B cells (B220^{+}) in CD45.2^{+} peripheral blood cells of mice 12 weeks post-transplantation (WT, n = 8; Srsf2$^{P95H/+}$, n = 8; Runx1 KO, n = 8; double mutant, n = 8).

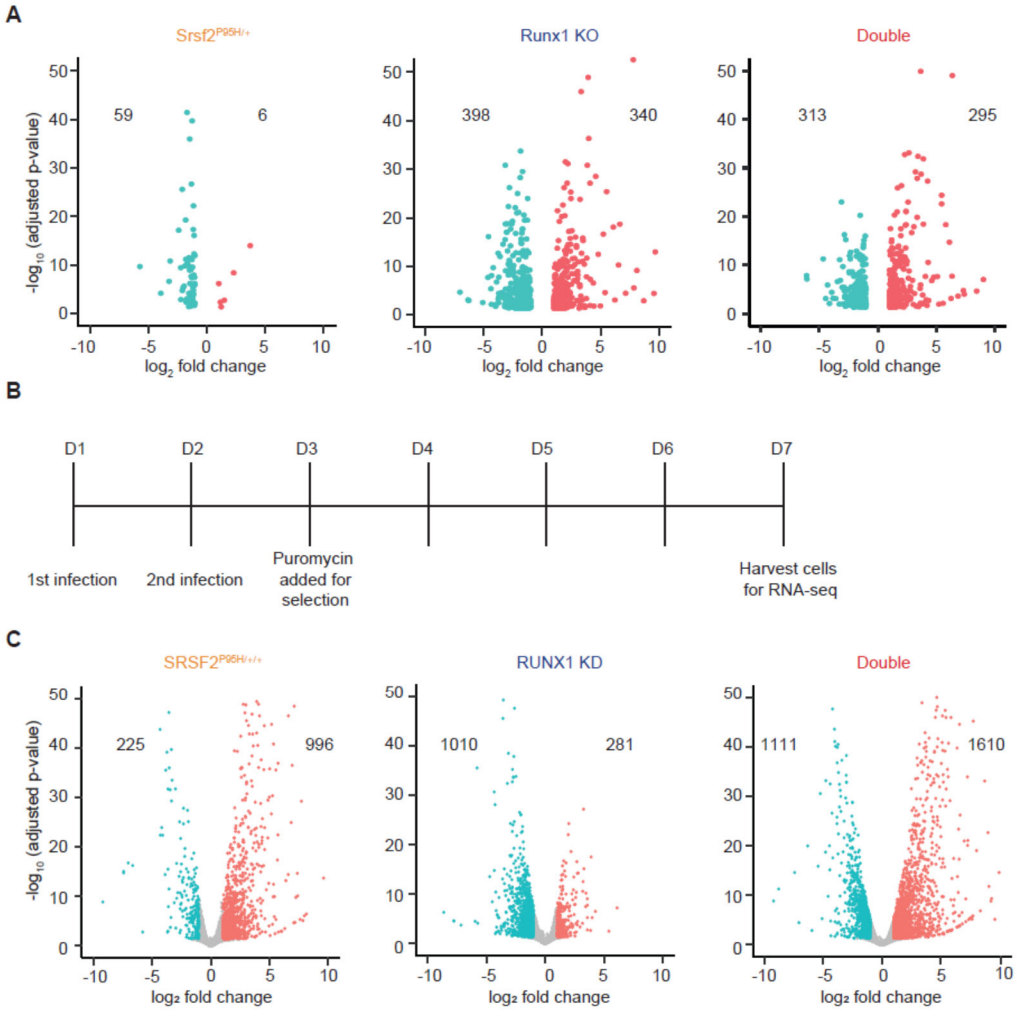


Supplementary figure 4. Mouse models and K562 cells used for RNA sequencing analyses

(A) A schematic diagram depicting the preparation of murine Lin$^-$c-Kit$^+$ (LK) cells for RNA-seq. Bone marrow cells from the indicated donors were transplanted into recipients. Four weeks after pIpC injection to activate deletion of Runx1 and/or expression of Srsf2P95H mutant in recipient mice, RNA samples were prepared from sorted LK bone marrow cells and used for RNA-seq. The composition of myeloid progenitors among the LK cells was not checked at this early 4-week timepoint. By 16-weeks post-pIpC injection, we observed a difference in the distribution of myeloid progenitor cells (Fig. 2E). It is possible that early alterations to these populations may be present in the LK cell collected and thus reflected in the RNA-seq results.

(B) Expression of SRSF2 P95H alleles as a percentage of all SRSF2 mRNAs expressed in K562 cells of the indicated genotypes: parental K562 (WT), RUNX1 knockdown (RUNX1 KD), SRSF2 P95H single copy knock-in (P95H/+/+), and SRSF2 P95H knock-in with RUNX1 knockdown (Double).

(C) Western blot showing the knockdown efficiency of RUNX1 in K562 cells of the indicated genotypes. To quantitate depletion of RUNX1 in three biological replicates, the relative protein levels of RUNX1 were first normalized to the β-Actin loading control and then quantified relative to the protein level present in the WT controls (set as 1).

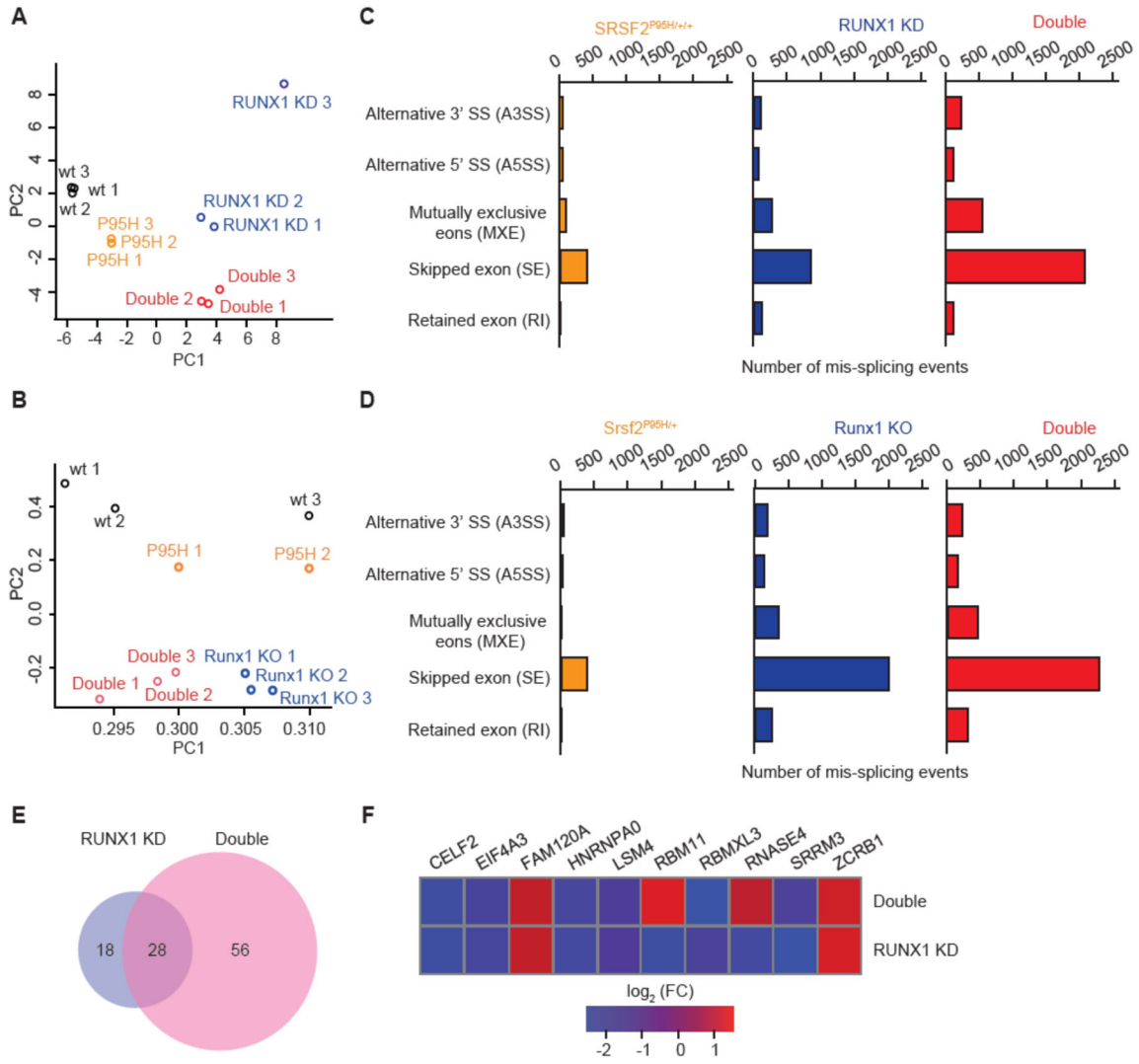


Supplementary figure 5. Global gene expression dysregulated by single and double mutants

(A) Volcano plots of significantly altered gene expression changes in Srsf2$^{P95H/+}$, Runx1 KO, and Double mutant mouse LK cells compared to WT cells.

(B) Timeline of RUNX1 shRNA knockdown in K562 cells and RNA collection for RNA-seq.

(C) Volcano plots of significantly altered gene expression changes in SRSF2$^{P95H/+/+}$, RUNX1 KD, and Double mutant K562 cells compared to WT K562 cells.



Supplementary figure 6. RUNX1 deficiency dramatically impacts global RNA splicing

(A) Unsupervised principal component analysis of the significantly altered splice events in three biological replicates per genotype of the isogenic K562 cell lines. Significant splicing events were identified by the rMATS pipeline with FDR < 0.05 and |PSI| > 0.05.

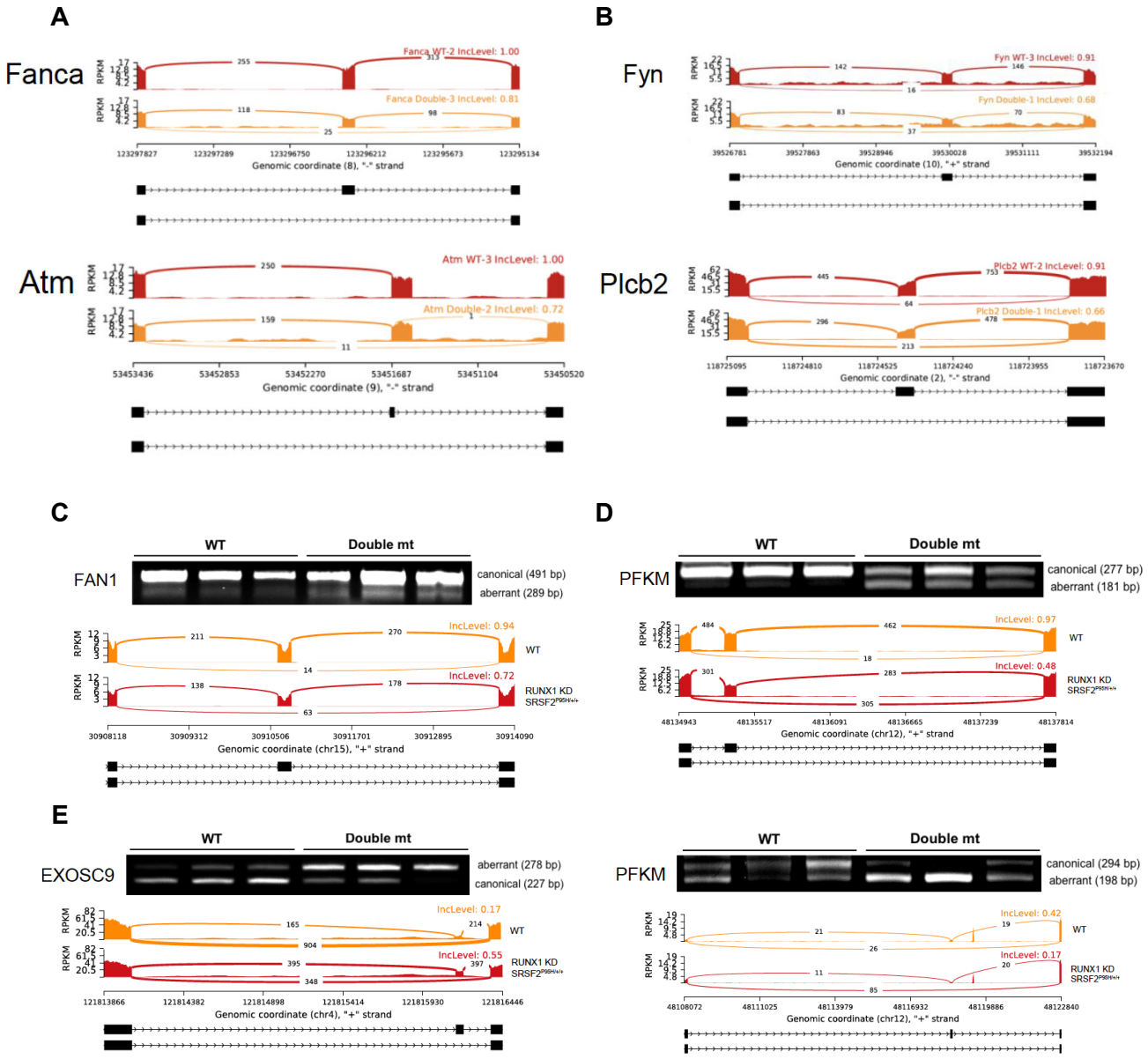
(B) Unsupervised principal component analysis of the significantly altered splice events in biological replicates per genotype of the murine LK cells (WT, n = 3; $Srsf2^{P95H/+}$, n = 2; Runx1 KO, n = 3; Double mutant, n = 3). Significant splicing events were identified by the rMATS pipeline with FDR < 0.05 and |PSI| > 0.05.

(C) The number and types of alternative splicing events in single/double mutant K562 cells compared to WT cells.

(D) The number and types of alternative splicing events in single/double mutant murine LK cells compared to WT cells.

(E) Numbers and overlap of dysregulated genes encoding RNA binding proteins in RUNX1 KD and double mutant K562 cells.

(F) Heatmap showing the change in expression (log2 fold change) of 10 of the 28 commonly differentially expressed RBP genes in (E) that encode RBPs known to regulate RNA splicing.



Supplementary figure 7. Representative altered splicing events

(A-B) Sashimi plots illustrating the altered cassette exon usage of Fanca and Atm (A), Fyn and Plcb2 (B) in WT and Double mutant murine LK cells.

(C-E) RT-PCR validation and sashimi plots depicting the abnormal spicing of FAN1 (C), PFKM (D), and EXOSC9 (E) in WT and Double mutant K562 cells.

SUPPLEMENTAL REFERENCES

1. Pollyea DA, Harris C, Rabe JL, et al. Myelodysplastic syndrome-associated spliceosome gene mutations enhance innate immune signaling. *Haematologica*. 2019;104(9):e388-e392.
2. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.
3. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
4. Shen S, Park JW, Lu ZX, et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S A*. 2014;111(51):E5593-5601.
5. Yu GC, Wang LG, Han YY, He QY. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omics-a Journal of Integrative Biology*. 2012;16(5):284-287.