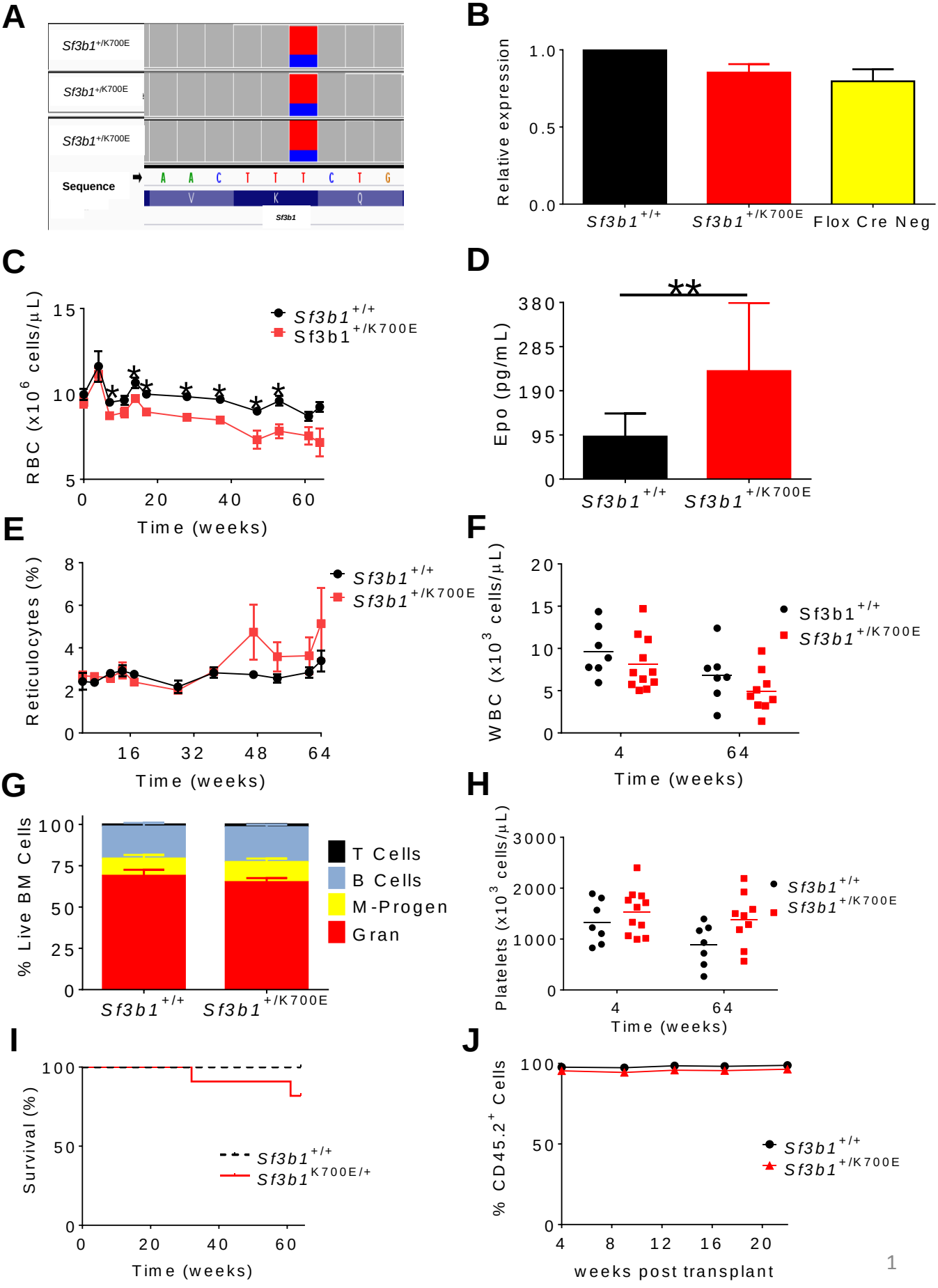
Supplemental Data



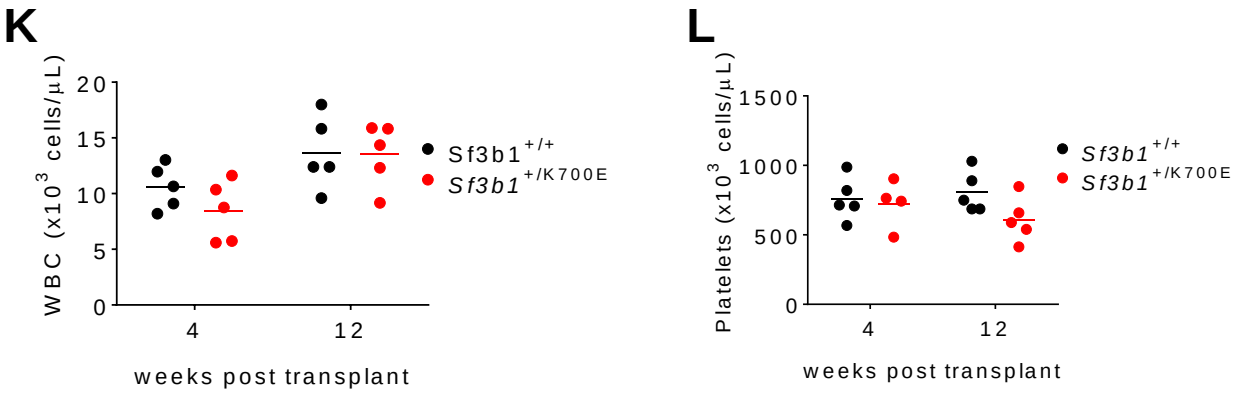
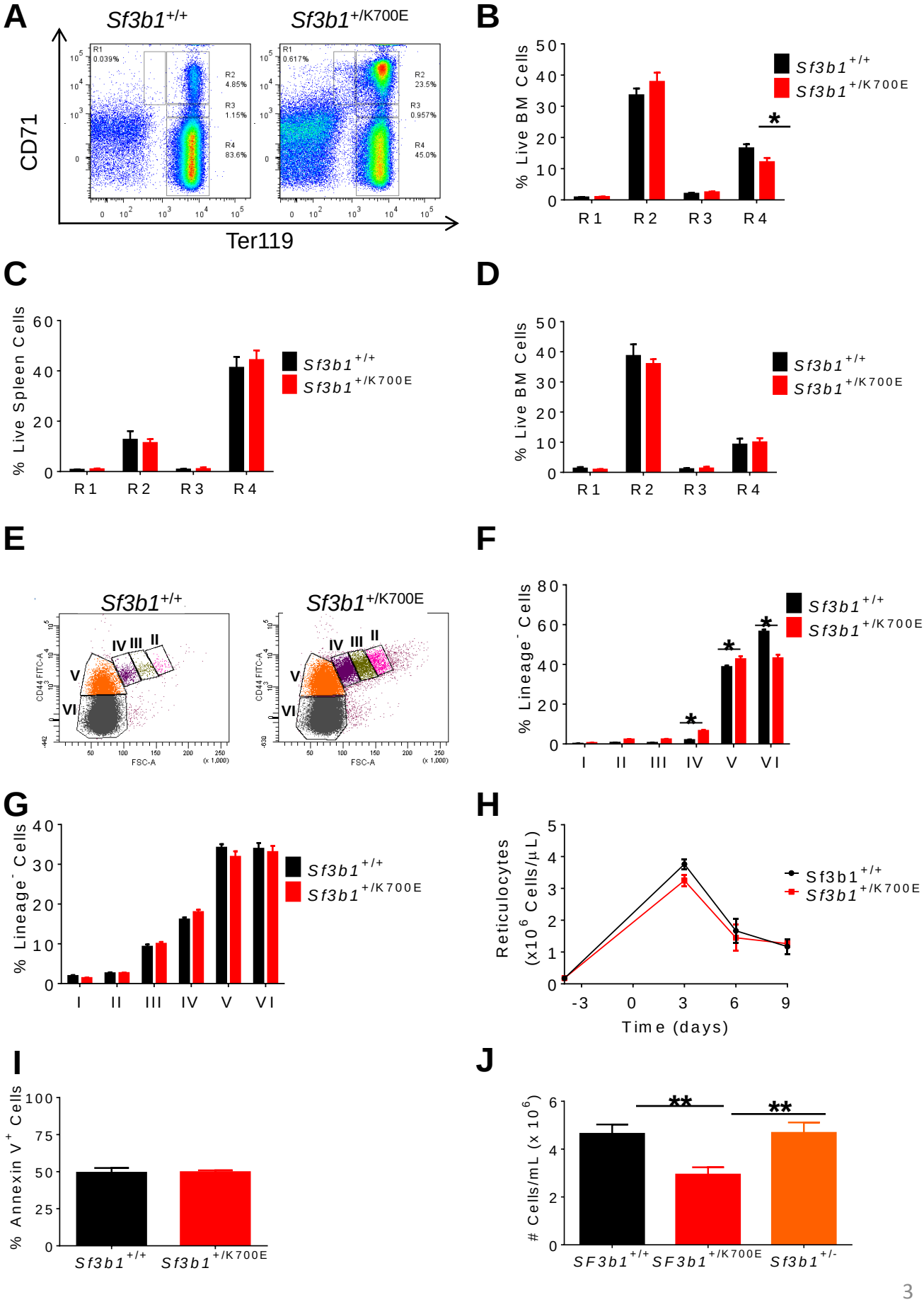


Figure S1, related to Figure 1. Heterozygous conditional knock-in of *Sf3b1*^{K700E} specifically results in a progressive decrease in red blood cells. (A) Sequence fragment density from RNA sequencing performed on 3 Sf3b1$^{+/K700E}$ lineage negative, c-kit positive (LK) myeloid progenitor samples, four weeks after pIpC treatment. Red denotes wild-type (adenine) and blue denotes mutant (guanine) nucleotide reads. Mutant allele frequency ranged from 27 – 32 percent. (B) Total Sf3b1 mRNA relative to *Sf3b1*$^{+/+}$ mice for *Sf3b1*$^{+/+}$, *Sf3b1*$^{+/K700E}$, and Flox Cre negative (*Sf3b1*$^{+/K700E}$ *Mx1*-Cre negative) mice (n = 3 mice per group, 8 weeks post-pIpC). Samples were normalized to Gapdh. (C) Red blood cell count over the course of 64 weeks post-pIpC (n = 9 *Sf3b1*$^{+/+}$ and 11 *Sf3b1*$^{+/K700E}$ mice). (D) Plasma erythropoietin levels measured 60 weeks after pIpC induction (n = 14 *Sf3b1*$^{+/+}$and 16 *Sf3b1*$^{+/K700E}$ mice). (E) Reticulocyte percentage over the course of 64 weeks post-pIpC (n = 9 *Sf3b1*$^{+/+}$ and 11 *Sf3b1*$^{+/K700E}$ mice). (F) White blood cell counts 4 weeks and 64 weeks after pIpC induction (n = 9 Sf3b1$^{+/+}$ and 11 Sf3b1$^{+/K700E}$ mice). (G) Percentage of mature white blood cells in the bone marrow of 9 *Sf3b1*$^{+/+}$and 11 *Sf3b1*$^{+/K700E}$ mice, 64 weeks post-pIpC. Gran: Granulocytes: Gr1^{hi} CD11b^{+}. M-progen: Myeloid progenitors (Gr1^{lo} CD11b^{+}), B cells B220^{+}, T cells CD3^{+}. (H) Platelet counts for *Sf3b1*$^{+/+}$ and *Sf3b1*$^{+/K700E}$ mice, 4 weeks and 64 weeks after pIpC induction (n = 9 Sf3b1$^{+/+}$ and 11 Sf3b1$^{+/K700E}$ mice). (I) Kaplan-Meier survival curve for *Sf3b1*$^{+/K700E}$ mice and *Sf3b1*$^{+/+}$ littermate controls (n = 9 *Sf3b1*$^{+/+}$ and 11 *Sf3b1*$^{+/K700E}$ mice). (J) Percentage of CD45.2 donor chimerism in the peripheral blood over the course of 20 weeks after noncompetitive transplantation assay. One million unfractionated bone marrow cells from CD45.2^{+} *Sf3b1*$^{+/+}$ or *Sf3b1*$^{+/K700E}$ mice were transplanted into lethally irradiated (10.5 Gy) CD45.1^{+} B6.SJL recipients (n = 5 mice per group). (K - L) White blood cell (K) and platelet (L) counts for *Sf3b1*$^{+/+}$ and *Sf3b1*$^{+/K700E}$ recipients, 4 weeks and 12 weeks after noncompetitive transplantation (n = 5 mice per group; horizontal lines indicate the mean). Data presented as mean ± SEM. * $p < 0.05$; ** $p < 0.001$. Horizontal line indicates the mean for the scatter dot plots.


A
Sf3b1+/+
Sf3b1+/K700E
CD71
Ter119
R1 0.039%
R2 4.85%
R3 1.15%
R4 83.6%
R1 0.617%
R2 23.5%
R3 0.957%
R4 45.0%
B
% Live BM Cells
Sf3b1+/+
Sf3b1+/K700E
R1 R2 R3 R4
C
% Live Spleen Cells
D
% Live BM Cells
E
Sf3b1+/+
Sf3b1+/K700E
CD44 FITC-A
FSC-A
I II III IV V VI
F
% Lineage- Cells
G
% Lineage- Cells
H
Reticulocytes (x10^6 Cells/μL)
Time (days)
I
% Annexin V+ Cells
J
Cells/mL (x 10^6)
SF3b1+/+ SF3b1+/K700E Sf3b1+/-

K

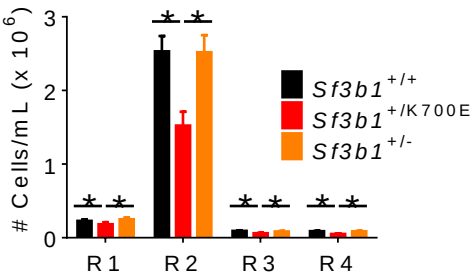


Figure S2, related to Figure 2. Sf3b1^{K700E} causes a block in terminal erythroid maturation. (A) Representative flow cytometry plot for CD71 (transferrin receptor) and Ter119 staining of erythroid precursors from the spleen of a *Sf3b1*$^{+/+}$ and a *Sf3b1*$^{+/K700E}$ mouse, as indicated. (B) Analysis of terminal erythroid maturation in the bone marrow of *Sf3b1*$^{+/+}$ and *Sf3b1*$^{+/K700E}$ mice, 64 weeks after pIpC treatment (n = 9 *Sf3b1*$^{+/+}$and 11 *Sf3b1*$^{+/K700E}$ mice). (C - D) Analysis of terminal erythroid maturation in the spleens (C) and bone marrow (D) of *Sf3b1*$^{+/+}$ and *Sf3b1*$^{+/K700E}$ mice, 12 weeks post pIpC treatment (n = 5 *Sf3b1*$^{+/+}$and 6 *Sf3b1*$^{+/K700E}$ mice). (E) Representative flow cytometry plot using CD44 staining and forward scatter to characterize erythroid maturation in *Sf3b1*$^{+/+}$ and *Sf3b1*$^{+/K700E}$ mice. Erythroid maturation can be quantified from the stage II basophilic erythroblast to the stage VI mature red blood cell, as indicated. (F - G) Percentage cells in each stage of terminal erythroid maturation in the spleen (F) and bone marrow (G) of *Sf3b1*$^{+/+}$ and *Sf3b1*$^{+/K700E}$ mice, 52 weeks post pIpC treatment (n = 6 *Sf3b1*$^{+/+}$and 7 *Sf3b1*$^{+/K700E}$ mice). (H) Absolute reticulocyte count [(% reticulocytes/100) x red blood cell count] measured 4 days before and 3, 6, and 9 days after the first injection of *Sf3b1*$^{+/+}$and *Sf3b1*$^{+/K700E}$ animals with phenylhydrazine (PHZ; n = 11 *Sf3b1*$^{+/+}$ and 10 *Sf3b1*$^{+/K700E}$ mice). (I) Percentage of Annexin V positive cells 48 hr after c-kit^{+} progenitor cells were plated in erythroid differentiation medium. Triplicate samples were run on 3 mice per group. (J) Absolute number of *Sf3b1*$^{+/+}$, *Sf3b1*$^{+/K700E}$, or *Sf3b1*$^{+/-}$ cells in culture 48 hr after 2.0 x 10^5 bone marrow c-Kit^{+} progenitor cells were plated in erythroid differentiation medium. Triplicate samples were analyzed from each of 3 mice per group. (K) Analysis of the total number of cells in each stage of terminal erythroid maturation 48 hr after bone marrow c-Kit^{+} progenitor cells from *Sf3b1*$^{+/+}$, *Sf3b1*$^{+/K700E}$, or *Sf3b1*$^{+/-}$ mice were plated in erythroid differentiation medium. Triplicate samples were analyzed from each of 3 mice per group. Data presented as mean ± SEM. * $p < 0.05$; ** $p < 0.01$.

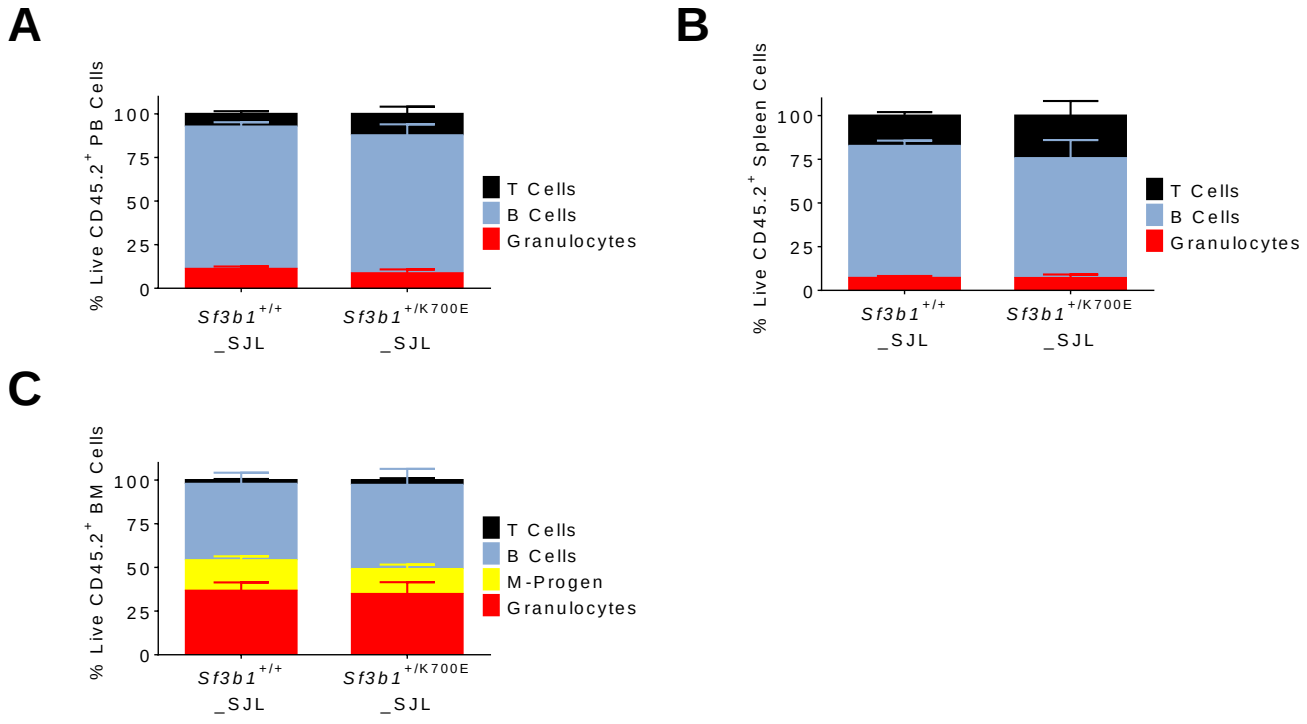
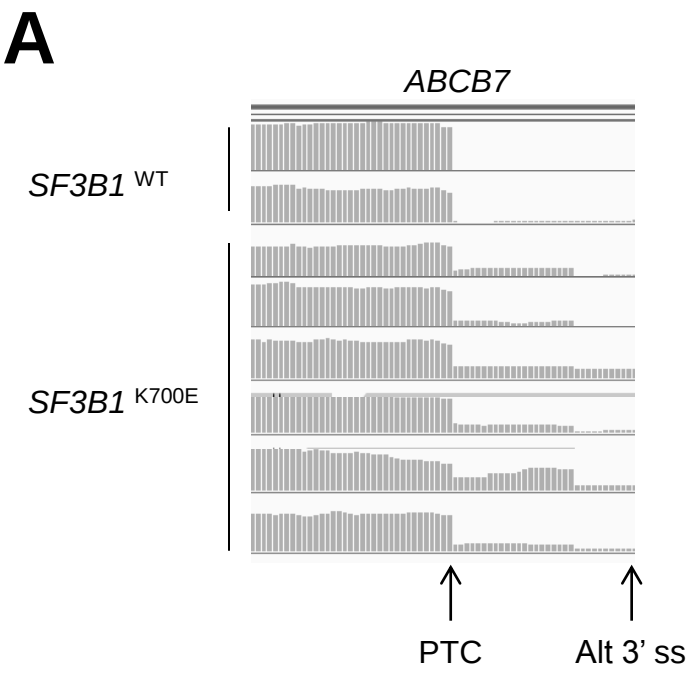
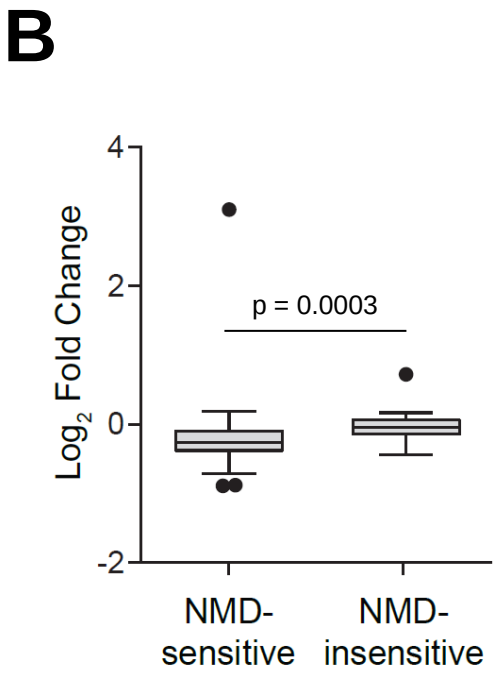


Figure S3, related to Figure 3. Heterozygous conditional knock-in of *Sf3b1*^{K700E} does not alter the distribution of mature white blood cell lineages in competitive transplantation recipients. (A - C) Percentage of mature white blood cells in the peripheral blood (A), spleen (B) and bone marrow (C) of recipient mice, 20 weeks after competitive repopulation assay. One million unfractionated bone marrow cells from CD45.2$^+$ *Sf3b1*$^{+/+}$ or *Sf3b1*$^{+/K700E}$ mice were mixed 1:1 with 1 x 10^6 CD45.1$^+$ B6.SJL unfractionated bone marrow cells and transplanted into lethally irradiated (10.5 Gy) B6.SJL recipients (n = 5 mice per group). Granulocytes: Gr1^{hi} CD11b$^+$. M-progen: Myeloid progenitors (Gr1^{lo} CD11b$^+$), B cells B220$^+$, T cells CD3$^+$. Data presented as mean ± SEM. * $p < 0.05$.


A
ABCB7
SF3B1 WT
SF3B1 K700E
PTC
Alt 3' ss


B
Log2 Fold Change
4
2
0
-2
p = 0.0003
NMD-sensitive
NMD-insensitive

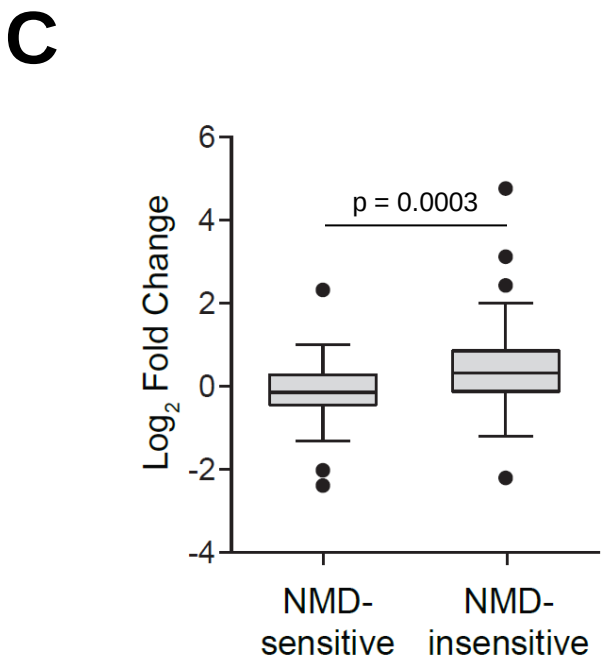

C
Log2 Fold Change
6
4
2
0
-2
-4
p = 0.0003
NMD-sensitive
NMD-insensitive

Figure S4, related to Figure 4. Over 30% of all aberrantly spliced genes in SF3B1-mutant cells are putative targets of nonsense mediated decay (NMD). (A) Sequence fragment density from RNA sequencing performed on two *SF3B1* wild-type and six *SF3B1*-mutant MDS patient samples. An aberrant 3' splice site between exon 8 and exon 9 of *ABCB7* caused the addition of a premature termination codon (PTC) within a seven amino acid addition to the protein sequence, prior to the start of the next canonical exon as described previously for *SF3B1*-mutant isogenic Nalm-6 cells. (B) Comparison of whole gene expression of predicted NMD sensitive (n = 22) and NMD insensitive (n = 36) transcripts given as the $\log_2$ fold change between *Sf3b1*$^{+/K700E}$ and *Sf3b1*$^{+/+}$ myeloid progenitors (LK cells, n = 3 mice per group). NMD predictions were derived from aberrant junctions identified in *Sf3b1*$^{+/K700E}$ myeloid progenitor samples (see Table S1). Box plots and whiskers are represented as per the Tukey method. (C) Comparison of whole gene expression of predicted NMD sensitive (n = 36) and NMD insensitive (n = 72) transcripts given as the $\log_2$ fold change between *SF3B1*–mutant and *SF3B1* wild-type MDS patient samples. NMD predictions were derived from aberrant junctions identified in *SF3B1*-mutant patient samples (see Table S2). Data presented as mean ± SEM.

Table S1, related to Figure 4. List of Aberrant Splice Junctions and Predicted NMD Candidates Identified in Sf3b1K700E Myeloid Progenitor Cells. Provided as an Excel File.

Table S2, related to Figure 4. List of Aberrant Splice Junctions and Predicted NMD Candidates Identified in SF3B1-Mutant MDS Patient Sample Bone Marrow. Provided as an Excel File.

Table S3, related to Figure 4. Significantly Upregulated and Downregulated Pathways Identified by GSEA

MDS UP	LK UP
Ribosome Peptide Chain Elongation	**RPS14 Down V1 Up**
3' UTR[1] Mediated Translational Regulation	
Ribosome	
Translation	
Nonsense Mediated Decay Enhanced by the Exon Junction Complex	
Structural Constituent of Ribosome	
Metabolism of mRNA	
Metabolism of Proteins	
Translated RNA	
Peptide Chain Elongation	
RNA Binding	
RPS14[2] Down V1 Up	
MDS DOWN	**LK DOWN**
Heme Matabolism	Ribosome
Cell Cycle	RNA Binding
Cell Cycle Mitotic	3' UTR Mediated Translational Regulation
DNA Replication	
Normal Quiescent vs Normal Dividing Down	Metabolism of RNA
CML Dividing vs Normal Quiescent Up	Metabolism of mRNA
Hematopoiesis Mature Cell	Translation
Proliferation	Transcription
CDC27[3]	**Cell Cycle Mitotic**
RRM1[4]	RNA Splicing
Cell Cycle G0	mRNA Processing
Response to DNA Damage Stimulus	**RPS14 Down V1 Down**
Regulation of Cell Cycle	**G2-M Checkpoint**
RPS14 Down V1 Down	**Heme Metabolism**
G2-M Checkpoint	

[1]Untranslated Region
[2]Ribosomal Protein S14
[3]Cell Division Cycle 27
[4]Ribonucleotide Reductase M1

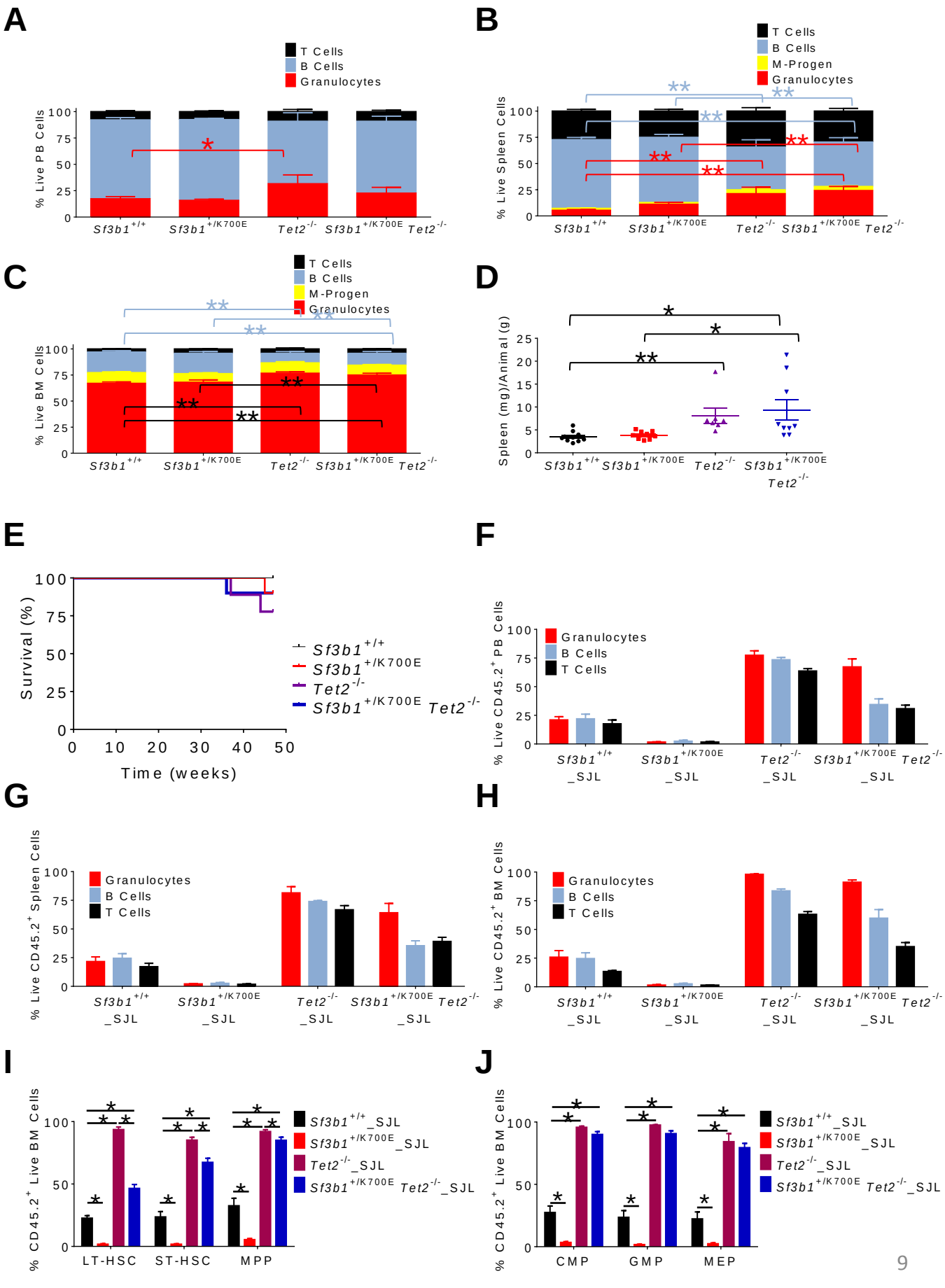

A
T Cells
B Cells
Granulocytes
% Live PB Cells
Sf3b1+/+
Sf3b1+/K700E
Tet2-/-
Sf3b1+/K700E Tet2-/-
B
T Cells
B Cells
M-Progen
Granulocytes
% Live Spleen Cells
C
T Cells
B Cells
M-Progen
Granulocytes
% Live BM Cells
D
Spleen (mg)/Animal (g)
E
Survival (%)
Time (weeks)
Sf3b1+/+
Sf3b1+/K700E
Tet2-/-
Sf3b1+/K700E Tet2-/-
F
% Live CD45.2+ PB Cells
Granulocytes
B Cells
T Cells
Sf3b1+/+ _SJL
Sf3b1+/K700E _SJL
Tet2-/- _SJL
Sf3b1+/K700E Tet2-/- _SJL
G
% Live CD45.2+ Spleen Cells
H
% Live CD45.2+ BM Cells
I
% CD45.2+ Live BM Cells
LT-HSC
ST-HSC
MPP
J
% CD45.2+ Live BM Cells
CMP
GMP
MEP

Figure S5, related to Figure 5. *Sf3b1*$^{+/K700E}$ *Tet2*$^{-/-}$ double mutant mice have additional features of MDS and a competitive advantage in repopulation experiments. (A - C) Percentage of mature white blood cells in the peripheral blood (A), spleen (B) and bone marrow (C) of *Sf3b1*$^{+/+}$, *Sf3b1*$^{+/K700E}$, *Tet2*$^{-/-}$ and *Sf3b1*$^{+/K700E}$ *Tet2*$^{-/-}$ primary mice, 45 weeks post-pIpC (n = 7 - 10 mice per group). Granulocytes: Gr1^{hi} CD11b^{+}. M-progen: Myeloid progenitors (Gr1^{lo} CD11b^{+}), B cells B220^{+}, T cells CD3^{+}. (D) Normalized spleen weights [spleen weight (mg) / animal weight (g)] from *Sf3b1*$^{+/+}$, *Sf3b1*$^{+/K700E}$, *Tet2*$^{-/-}$ and *Sf3b1*$^{+/K700E}$ *Tet2*$^{-/-}$ primary mice, 45 weeks post-pIpC (n = 7 - 10 mice per group). (E) Kaplan-Meier survival curve for the group of *Sf3b1*$^{+/+}$, *Sf3b1*$^{+/K700E}$, *Tet2*$^{-/-}$ and *Sf3b1*$^{+/K700E}$ *Tet2*$^{-/-}$ primary mice (n = 9 – 10 mice per group). (F - H) Percentage of mature white blood cells in the peripheral blood (F), spleen (G) and bone marrow (H) of recipient mice, 20 weeks after competitive repopulation assay. One million unfractionated bone marrow cells from CD45.2^{+} *Sf3b1*$^{+/+}$, *Sf3b1*$^{+/K700E}$, *Tet2*$^{-/-}$ or *Sf3b1*$^{+/K700E}$ *Tet2*$^{-/-}$ mice were mixed 1:1 with 1 x 10^{6} CD45.1^{+} B6.SJL unfractionated bone marrow cells and transplanted into lethally irradiated (10.5 Gy) B6.SJL recipients (n = 5 mice per group). Granulocytes: Gr1hi CD11b^{+}, B cells B220^{+}, T cells CD3^{+}. (I) Analysis of the percentage of LT-HSCs, ST-HSCs, and MPPs in the bone marrow of *Sf3b1*$^{+/+}$, *Sf3b1*$^{+/K700E}$, *Tet2*$^{-/-}$, and *Sf3b1*$^{+/K700E}$ *Tet2*$^{-/-}$ mice, 24 weeks after competitive repopulation assay (n = 5 mice per group). (J) Analysis of the percentage of CMPs, GMPs, and MEPs in the bone marrow of *Sf3b1*$^{+/+}$, *Sf3b1*$^{+/K700E}$, *Tet2*$^{-/-}$, and *Sf3b1*$^{+/K700E}$ *Tet2*$^{-/-}$ mice, 24 weeks after competitive repopulation assay (n = 5 mice per group). Data presented as mean ± SEM. * $p < 0.05$; ** $p < 0.01$.

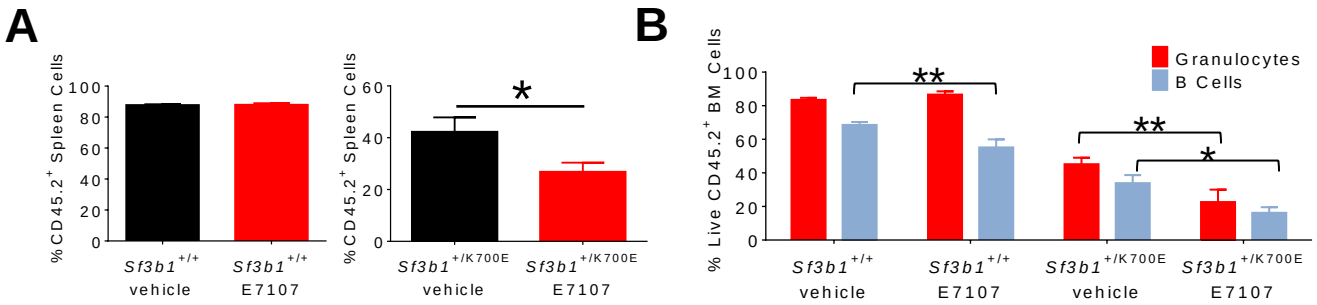


Figure S6, related to Figure 6. Sf3b1^{K700E} expressing cells are sensitive to the spliceosome modulator, E7107, in vivo. (A) Percentage of CD45.2$^+$ donor chimerism in the spleen of *Sf3b1*$^{+/+}$ or *Sf3b1*$^{+/K700E}$ recipients treated with E7107 or vehicle (n = 5 mice per group). (B) Mature myeloid and lymphoid cell chimerism in the bone marrow of *Sf3b1*$^{+/+}$ and *Sf3b1*$^{+/700E}$ competitive transplant recipients treated with E7107 or vehicle (n = 5 mice per group). Granulocytes: Gr1$^+$. B cells B220$^+$. Data presented as mean ± SEM. * $p < 0.05$; ** $p < 0.01$.

Supplemental Experimental Procedures

Generation of a *Sf3b1*^{K700E} Conditional Knock-In Mouse

Mouse embryonic stem cells with a targeting construct containing inverted copies of exons 15 (with the K700E mutation) and 16 in a C57BL/6/129 genetic background were generated by Biocytogen (Worcester, MA). The neomycin/lacZ cassette was flipped out *in vitro* by transfection of a plasmid expressing the flippase recombinase (FRT). Successful FRT recombination was validated by PCR (forward primer: 5'-TCGCACTTGAGC TATTGGGGAGT-3'; reverse primer: 5'-AGGCATGGTAGCTCACACCTGA-3'). Following confirmation of germline transmission, mice were crossed with the Mx1-Cre mouse strain (Jackson Laboratory, stock number 002527). To cause inversion of exons 15 and 16, *Sf3b1*$^{fl/+}$ mice were given four rounds of poly(I:C) (InvivoGen, San Diego, CA) by intraperitoneal injection. *Sf3b1*$^{+/K700E}$ mice were backcrossed against C57BL/6J inbred mice for 9 generations. Littermate controls were used for all experiments, unless stated otherwise in the figure legends.

Antibodies and Reagents

Bone marrow cells were isolated by flushing and crushing pelvic and hind leg bones with PBS (Life Technologies™) + 2% fetal bovine serum (FBS) + penicillin/streptomycin (Life Technologies™). Whole bone marrow was lysed on ice with red blood cell lysis solution (Invitrogen/ Life Technologies™), and washed in PBS (Life Technologies™) plus 2% FBS. Single-cell suspensions of spleen were prepared by pressing tissue through a 70 μM cell strainer followed by red blood cell lysis. Cells were labeled with monoclonal antibodies in 2% FBS/PBS for 30 min on ice and analyzed using an LSR II flow cytometer (BD Biosciences). Antibodies and reagents are listed in the table below:

Antibodies/Reagents Used for Flow Cytometry						
	Fluorochrome	Company	Fluorochrome	Company	Fluorochrome	Company
live/dead	Dapi	Roche Diagnostics				
Gr1	efluor 450	eBioscience	APCCy7	eBioscience	APC	eBioscience
B220	efluor 450	eBioscience	APCCy7	eBioscience	PECy7	BD Pharmingen
CD3	efluor 450	eBioscience	APCCy7	eBioscience		
CD11b	efluor 450	eBioscience	APCCy7	eBioscience		
Ter119	efluor 450	eBioscience	APCCy7	eBioscience	APC	eBioscience
CD150	PE/Cy5	Biolegend				
CD48	APCCy7	Biolegend				
CD34	FITC	eBioscience				
CD16/32	PE	eBioscience				
Sca1	PECy7	eBioscience				
c-kit	APC	eBioscience				
CD45.1	PE	eBioscience				
CD45.2	FITC	eBioscience				
Annexin V	APC	BD Pharmingen				
CD71	PE	eBioscience				
CD44	FITC	BD Biosciences				
CD45	APCCy7	BD Pharmingen				
Ki67	FITC	BD Pharmingen				
Hoechst		Life Technologies				

Erythropoietin Assay
100 mL serum was used in Mouse Erythropoietin Quantikine ELISA Kit (R&D Systems, Minneapolis, MN) according to the manufacturer's instructions.

Phenylhydrazine Stress Test
PHZ (Sigma-Aldrich) was diluted in PBS and 30 mg/kg was injected subcutaneously on 2 consecutive days (day 0 and day +1). Blood was collected 4 days prior to injection and at days 3, 6, and 9-post injection for hemoglobin and reticulocyte count measurements.

Colony-Forming Assays
Whole BM and spleen cells were harvested, subjected to RBC lysis, and resuspended in IMDM with 10% FBS and 5% penicillin-streptomycin. Cells were plated in triplicate in M3434 methylcellulose media (Stem Cell Technologies) at 20,000 cells/dish. Colonies were scored after 7 days.

Histopthology
Murine organs were fixed in formalin with 3.7% paraformaldehyde, and histology slides were prepared and stained by the Brigham and Women's core Rodent Histology Facility. Digital images were acquired using a Nikon Eclipse E400 microscope equipped with a digital camera and analyzed using Spot Advanced software (model 2.1.1, Diagnostic Instruments).

Repopulation and Transplantation Assays
For competitive repopulation experiments, competitor bone marrow was obtained from age-matched B6.SJL mice. Transplant recipients were 6- to 8-week-old B6.SJL mice. For lethal irradiation, mice were given 1 doses of 10.5 cGy on the day prior to transplant. Donor bone marrow was injected into the retro-orbital plexus.

Primary Murine Cell Isolation and Culture
c-kit^{+} cells were isolated from whole bone marrow using the autoMACS pro Separator (Miltenyi Biotec). C-kit^{+} bone marrow cells cultured in StemSpan™ Serum-Free Expansion Medium (Stem Cell Technologies) supplemented with murine thrombopoietin (m-tpo) (50 ng/mL; Peprotech), and murine stem cell factor (m-scf) (50 ng/mL; Peprotech).

Quantitative Real-Time PCR (qRT-PCR)
Unfractionated bone marrow cells were isolated as described above from *Sf3b1*$^{+/+}$, *Sf3b1*$^{+/K700E}$, and Flox Cre negative (*Sf3b1*$^{+/K700E}$ *Mx1*-Cre negative) mice. The bone marrow cells were lysed with TRIzol® Reagent (Invitrogen / Life Technologies), according to the manufacturer's specifications for RNA extraction. Purified RNA was dissolved in sterile distilled water. cDNA was synthesized using the QuantiTect Reverse Transcription Kit (Qiagen), according to the manufacturer's protocol. qRT-PCR reactions were performed using TaqMan Expression Assays for Sf3b1 and Gapdh genes (Applied Biosystems) and TaqMan Universal PCR Mastermix (Applied Biosystems) in an ABI Prism 7900 Real-Time PCR System (Applied Biosystems). Each condition was run in triplicate. Expression of Sf3b1 was normalized to Gapdh.

Quantification of RNA sequencing data
RNA sequencing analysis was performed as described previously (Darman et al., 2015). Briefly, for all RNA-seq data, reads were aligned to hg19 by STAR (Dobin et al., 2013), gene isoform counts were quantified by Sailfish (Li and Dewey, 2011) against

GENCODE annotation version v19. Gene and isoform quantification is given in unit of TPM (transcripts per million). Raw junction counts generated by STAR were used for all further downstream analyses; however splice junctions which did not meet a minimum average threshold of 10 counts in either wild-type SF3B1 or SF3B1-mutant cohorts were removed in order to reduce false positives. In order to identify intron retention events, estimates of read counts spanning each intron-exon junction were required. For every splice junction in each BAM file, we counted reads which fully overlapped artificially defined 6 bp regions, 3 bp intronic and 3 bp exonic, across each of the 3' and 5' intron-exon junctions.

Identification of Differential Splice Junctions

Analysis was performed as described previously (Darman et al., 2015). Briefly, instead of raw splice junction counts obtained during the quantification process, percentages of splice junctions relative to shared splice sites were calculated and used to assess differential splicing activities between SF3B1-mutant and wild-type SF3B1. In brief, the junction usage percentage, or percent spliced in (PSI) is a measurement of the usage of one splice junction relative to all other splice junctions that share the same splice site. Therefore for each shared splice site, the raw counts of each splice junction are divided by the total counts of all splice junctions that utilize the same shared splice site in order to derive a ratio which is then multiplied by 100 to convert to a percentage. Differential PSI was assessed between SF3B1-mutant and wild-type SF3B1 using moderated t-test defined in the Bioconductor's limma package (Wettenhall and Smyth, 2004). The statistical p values were corrected using the Benjamini-Hochberg procedure.

Canonical splice junctions were identified as those junctions which shared a splice site with an aberrant junction and which were also differentially upregulated in the wild-type SF3B1 setting, i.e., exhibit "switch-like" behavior. To be conservative, we reported only those aberrant junctions which had corresponding canonical splice junctions which could also pass a q-value threshold of 0.20. Intron retention events were required to be upregulated (have significantly differential PSI between SF3B1-mutant and wild-type SF3B1) for both 3' and 5' intron-exon junctions as defined above. The canonical junction was therefore defined to be the junction which spans the intronic sequence, and this was also required to be upregulated in wild-type SF3B1 as was required for all other canonical splicing events.

Motif and Cross-Species Analyses

Analysis was performed as described previously (Darman et al., 2015). Briefly, consensus sequences for 3' splice sites were defined using 35 intronic bases upstream and 3 exonic bases downstream of both aberrant AGs preferentially used bySF3B1-mutant and canonical AGs preferentially used by wild-type SF3B1. All alternative 3' ss motifs were derived from cryptic AGs found upstream of their canonical 3' ss. Reference 3' ss were derived from hg19 (human) and mm10 (mouse) annotations given in RefSeq (Tatusova et al., 2014), KnownGene (Hsu et al., 2006), AceView (Thierry-Mieg and Thierry-Mieg, 2006), and Ensembl (Cunningham et al., 2015) databases. Pictograms of consensus sequences were generated using WebLogo (Crooks et al., 2004). In cases where the number of sequences exceeded 2000 (as in all reference 3' ss), 2000 sequences were randomly sampled in order to generate the image.

NMD Prediction

Analysis was performed as described previously (Darman et al., 2015). Briefly, the existence of PTCs found > 55 nt upstream from the last exon-exon junction in a

transcript has been shown to be most predictive of degradation by NMD (Rivas et al., 2015). In order to predict the location of PTCs in transcripts, we attempted to predict the complete sequence of all RefSeq transcripts which could contain the aberrant junction, e.g., an alternative 3' ss shares its 5' ss with a known RefSeq transcript, causing the elongation of the downstream exon. If a given transcript contained more than one aberrant junction, changes effected by each aberrant junction were applied and evaluated independently. Additionally, we detect aberrant transcripts which extend well beyond the canonical transcript termination (lack of PTC), which have been shown to be subject to nonstop mediated decay, a separate pathway which nonetheless results in transcript degradation (Wu and Brewer, 2012). Protein sequence prediction was carried out for each transcript independently using standard codon translation of the coding sequence. Each prediction is given as either a) True, a PTC is predicted > 55 nt from the final exon-exon junction in all RefSeq transcripts which could contain the aberrant junction, b) False, no PTCs < 55 nt upstream from the last exon-exon junction would be generated in any RefSeq transcript, c) Ambiguous, some transcripts which could contain the aberrant transcript at that gene locus are predicted to be NMD targets while others are not, and d) No CDS, which is assigned if the aberrant junction does not affect genes with any coding transcripts (lncRNA, etc) or the aberrant junction could not be assigned to any known RefSeq transcript.

Gene Set Enrichment Analysis (GSEA)

To identify pathways that were enriched in the sorted Sf3b1$^{+/K700E}$ myeloid progenitor (LK, n = 3) and SF3B1-mutant MDS patient bone marrow mononuclear cells (n = 6), Gene Set Enrichment Analysis (GSEA; http://www.broadinstitute.org/gsea/msigdb/annotate.jsp) was performed on genes which were consistently upregulated or downregulated in Sf3b1$^{+/K700E}$ myeloid progenitor (LK) or SF3B1-mutant unfractionated bone marrow samples when compared to wild-type controls (n = 3 Sf3b1$^{+/+}$ and n = 4 wild-type SF3B1 MDS patient samples).

Supplemental References

Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. Genome research *14*, 1188-1190.

Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S.*, et al.* (2015). Ensembl 2015. Nucleic acids research *43*, D662-669.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Hsu, F., Kent, W. J., Clawson, H., Kuhn, R. M., Diekhans, M., and Haussler, D. (2006). The UCSC Known Genes. Bioinformatics *22*, 1036-1046.

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics *12*, 323.

Rivas, M. A., Pirinen, M., Conrad, D. F., Lek, M., Tsang, E. K., Karczewski, K. J., Maller, J. B., Kukurba, K. R., DeLuca, D. S., Fromer, M.*, et al.* (2015). Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. Science *348*, 666-669.

Tatusova, T., Ciufo, S., Fedorov, B., O'Neill, K., and Tolstoy, I. (2014). RefSeq microbial genomes database: new representation and annotation strategy. Nucleic acids research *42*, D553-559.

Thierry-Mieg, D., and Thierry-Mieg, J. (2006). AceView: a comprehensive cDNA-supported gene and transcripts annotation. Genome biology *7 Suppl 1*, S12 11-14.

Wettenhall, J. M., and Smyth, G. K. (2004). limmaGUI: a graphical user interface for linear modeling of microarray data. Bioinformatics *20*, 3705-3706.

Wu, X., and Brewer, G. (2012). The regulation of mRNA stability in mammalian cells: 2.0. Gene *500*, 10-21.