SUPPLEMENTAL INFORMATION

Coordinated mis-splicing of TMEM14C and ABCB7 causes ring sideroblast formation in SF3B1-mutant myelodysplastic syndrome

Courtnee A. Clough[1,2], Joseph Pangallo[1,7,8#], Martina Sarchi[2,4#], Janine O. Ilagan[7,8#], Khrystyna North[7,8,9], Rochelle Bergantinos[2], Massiel C. Stolla[2], Jasmine Naru[5,6], Patrick Nugent[1,7,8], Eunhee Kim[10,12], Derek L. Stirewalt[5,6], Arvind R. Subramaniam[7,8], Omar Abdel-Wahab[10,11], Janis L. Abkowitz[2,3,9], Robert K. Bradley[7,8,9,†], and Sergei Doulatov[2,3,9,†]

[1]Molecular and Cellular Biology Program, University of Washington, Seattle, WA 98195
[2]Division of Hematology, University of Washington, Seattle, WA 98195
[3]Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle WA, 98195
[4]Department of Molecular Medicine, University of Pavia, 27100 Pavia PV, Italy
[5]Seattle Cancer Care Alliance, Seattle, WA 98109
[6]Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109
[7]Computational Biology Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109
[8]Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109
[9]Department of Genome Sciences, University of Washington, Seattle, WA 98195
[10]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[11]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[12]Department of Biological Sciences, College of Information-Bio Convergence Engineering, Ulsan National Institute of Science and Technology, South Korea

Includes:

- **Supplemental Methods**
- **Figures S1-4**
- **Supplemental File legends**

Supplemental Methods

Combined protocol for RS generation: 5F-HPCs derived from WT, SF3B1-only, and SF3B1;EZH2 iPSCs were cultured in SFEM medium (StemCell Technologies) with 50 ng/ml SCF, 50 ng/ml FLT3, 50 ng/ml TPO, 50 ng/ml IL-6, 10 ng/ml IL-3 (all Peprotech) and penicillin/streptomycin. Doxycycline was added at 2 µg/ml (Sigma). Cultures were maintained at <1.5 x 10^6 cells/ml, and media was exchanged every 3-4 days. 5F-HPC lines were maintained in culture between 60-100 days, and could be cryopreserved in 2X freezing solution with 40% FBS and 10% DMSO. Erythroid differentiation was initiated by plating 5F-HPCs into erythroid stage 1 media without doxycycline or exogenous iron. Stage 1 (6 days): IMDM + 1% BSA (Gibco), 20% FBS (Sigma), 1 mM L-glutamine, penicillin/streptomycin, 500 µg/ml holo-transferrin (Sigma), 10 µg/ml human insulin (CellSciences), 6U Epo (Procrit), 100 ng/ml SCF, and 5 ng/ml IL-3. Cells were seeded at a density of 2-3 x 10^5 cells/ml in 24-well plates. Stage 2 (up to 12 days): Media were changed to: IMDM + 1% BSA, 20% FBS, 1 mM L-glutamine, penicillin/streptomycin, 500 µg/ml holo-transferrin, 10 µg/ml human insulin, 6U Epo, and 50 ng/ml SCF. Cells were seeded at a density of 3 x 10^5 cells/ml in 24-well plates as cell number permitted. During stage 2, cells were maintained at a density of <2 x 10^6/ml in a maximum volume of 2.5 ml of media in a 24-well plate. For larger numbers of cells, plates and dish sizes were scaled up accordingly to maintain the 3 x 10^5 cells/ml density. Erythroid differentiation was carried out until day 18 (6 days stage 1 + 12 days stage 2), and differentiation efficiency routinely evaluated by flow cytometry and cell morphology (see Methods) prior to proceeding with RS staining. RS staining was performed starting on day 14, with the highest efficiency of RS formation observed on days 17 and 18, of erythroid culture. For RS staining, 5x10^4 – 1x10^5 cells were collected onto microscope slides with a cytocentrifuge (Thermo Scientific) and air-dried. Cells were fixed in -20C methanol for 12 minutes and air-dried prior to iron staining. Slides were stained with Prussian Blue kit (Abcam, ab150674). Solutions were prepared according to manufacturer's protocol. Slides were incubated in water for 30 seconds prior to a 12-minute incubation in iron stain followed by a 30 second water wash. Cells were counter-stained in nuclear fast red for 5 minutes and rinsed for 4 x 30 second washes in water. Slides were preserved with Vectamount (Vector laboratories) and coverslipped. Images were collected with an EVOS M5000 microscope (ThermoFisher Scientific). Cells were imaged at 60x objective with a brightfield phase ring with RGB Trans combined excitation. Brightness, saturation, and contrast were adjusted to ensure iron staining was visible for ring sideroblast analysis. Images were collected by tiling across slides and RS were manually counted. A ring sideroblast was

defined as an erythroid cell with iron-laden mitochondria covering >1/3 of the nuclear perimeter. >300 cells were counted for RS analysis.

RNA-seq splicing analysis: Isoform expression levels for all *SF3B1*-mutant and *SF3B1*-WT lines were estimated as previously described[1]. In brief, a transcriptome annotation for the hg19/GRCh37 human genome assembly was created by merging gene annotations from Ensembl v71.1[2], the UCSC knownGene track[3], and MISO v2.0 isoform annotations[4]. RSEM v1.2.4[5] was used to map all RNA-seq reads to this transcriptome annotation. Remaining unaligned reads were then mapped a database of all possible junctions between annotated 5' and 3' splice sites of the transcriptome annotation, as well as to the genome sequence, with TopHat v2.0.8b[6], and the resulting aligned reads were merged with the RSEM output. The expression levels of isoforms annotated in MISO v2.0's annotation were estimated with MISO v2.0. Events that were differentially spliced in *SF3B1*-mutant versus *SF3B1*-WT samples were identified as previously described. In brief, we defined the metric deltaPSI (ΔPSI) as the isoform ratio (absolute percentage of mRNA) in *SF3B1*-mutant samples - isoform ratio in *SF3B1*-WT cells, and computed a significance statistic using Wagenmakers's Bayesian framework[7] for single-sample comparisons or the Mann-Whitney *U* test for group comparisons. An event was classified as differentially spliced if it exhibited an absolute change of ΔPSI ≥ 10% and Bayes factor ≥ 5 or $p \leq 0.05$.

Luciferase assay: $5x10^4$ HEK293T cells were plated into a 24-well plate 16-24 hours before DNA transfection. 1µg of DNA was transfected using PEI-MAX™ (Polysciences) at a 3:1 PEI:DNA ratio. After 48-72 hours, each transfected well was split into 3-wells of a 96-well plate and luciferase was measured using the Nano-Glo® Dual-Luciferase® Reporter Assay (Promega) per manufacturer's instructions. Luminescence was quantified using an Infinite M200 plate reader (TECAN), triplicate values were averaged prior to analysis.

Western Blot analysis: Protein lysates were prepared by resuspending 100,000 washed cells/10µl of RIPA buffer (Thermo Fisher). Lysates were resolved by 4-20% SDS-PAGE (Bio-rad) and immunoblotted with antibodies for TMEM14C (Custom, YenZym Antibodies LLC), PPOX (Santa Cruz Biotechnology, sc-271768), MAP3K7 (CST 5206), ABCB7 (LS-B13035), and GAPDH (Abcam, ab9485). Membranes were washed in TBST and then incubated with anti-rabbit or anti-mouse HRP-conjugated secondary antibodies for visualization on a BioRad ChemiDoc. Images were exported to ImageJ for analysis.

Proteomics: Total protein was extracted from $CD34^+$ 5F-HPCs using urea lysis buffer that contained 6M Urea, 25 mM Tris (pH 8.0), 1 mM EDTA, 1 mM EGTA, protease and phosphatase inhibitors (Sigma-Aldrich). Total protein content of protein lysates from each sample was

determined using BCA assay (ThermoFisher Scientific). Protein lysates (30µg each) were processed for digestion and TMT labeled as per manufacturer's instructions (ThermoFisher Scientific). Tryptic peptides were pooled and desalted by solid-phase extraction using Oasis HLB sorbent (Waters, MA, USA). The desalted tryptic digests were fractionated by high-pH reverse phase (RP) liquid chromatography and concatenated into 24 samples. Fractionated samples were analyzed by LC-MS/MS on an Easy-nLC 1200 (Thermo Scientific) coupled to an Orbitrap Eclipse mass spectrometer (Thermo Scientific) operated in positive ion mode. Raw MS/MS spectra were searched using proteome discoverer 2.1 against the reviewed Human Universal Protein Resource (UniProt) sequence database. The search was performed with tryptic enzyme constraint set for up to two missed cleavages, oxidized methionine set as a variable modification, and carbamidomethylated cysteine set as a static modification. The precursor ion tolerance was set to 10 ppm and the fragment ion tolerance was set to 0.6 Da. Peptide identifications were filtered to a peptide FDR of 1%. Data normalization was performed by scaling each TMT channel to the channel median. Differential expression analysis was performed with R package. For these analyses we choose to keep features that had values in all TMT channels leaving us with 6782 quantified proteins. t-statistics were used on log2 scale to assess differences in expression/abundance.

Lentiviral vectors: Open reading frames (ORFs) for TMEM14C, PPOX, ABCB7, and MAP3K7 were purchased from the Mission® TRC3 Human LentiORF Collection and subcloned into the pSMAL overexpression vector (Addgene, Plasmid #161785). A luciferase ORF was synthesized and subcloned into pSMAL as a control. The TRC cloning vector pLKO.1 was used to generate the ABCG2 shRNA vectors. shRNA sequences were selected from the GPP web portal (Broad Institute). A shRNA targeting luciferase was used as a control. Sequences are as follows:

ABCG2-1:

5' CCGGCCTTCTTCGTTATGATGTTTACTCGAGTAAACATCATAACGAAGAAGGTTTTTG

LUC: 5' CCGGCTTACGCTGAGTACTTCGACTCGAGTCGAAGTACTCAGCGTAAGTTTTTG

Lentiviral preparation/5F-HPC infection: Virus was prepared using third generation packaging plasmids pMDLg/RRE and pRSV/Rev. pMD.G was used for VSV-G pseudotyping. Plasmids were transfected into HEK293T cells with calcium phosphate (Takara). Lentivirus was concentrated via ultracentrifugation at 23,000 rpm for 3 hrs and resuspended in SFEM and stored at -80°C. All viruses were titered by serial dilution on 293T cells. CD34-5F cells were transduced in a 96-well plate in standard CD34-5F media (outlined above). 2-3 x 10^5 cells were infected at an MOI of 5 in each well and spun at 2300 rpm, 30 minutes prior to incubation at

37°C. Cells were washed 16-24 hours later and cells were resuspended in fresh CD34-5F media.

Quantitative RT-PCR: RNA was extracted from 100-300,000 cells using TRIzol (ThermoFisher) and resuspended in water for cDNA synthesis using iScript cDNA synthesis kit (BioRad) according to manufacturer instructions. *ABCB7* primers were as previously described[8]. *ABCG2* primers were as previously described[9]. Primers for *PPOX*, *TMEM14C*, and *MAP3K7* were selected from https://pga.mgh.harvard.edu/primerbank/. Sequences are as follows:

PPOX: F – 5' TCTGCCGTGGAGTGTTTGC
R – 5' ATGGAACGATGGGTTTGCTCA

TMEM14C: F—5' CACTGGCTCAGTAGTGCCTTT
R – 5' AGTGGTAGAACCTCATTCCCAT

MAP3K7: F—5' AAACCACCAAACTTACTGCTGG
R – 5' CGCGTTATCACTTCCCAAAGAA

GAPDH: F – 5' TCCTGCACCACCAACTGCTTA
R – 5' TCTTCTGGGTGGCAGTGATGG

Figure S1

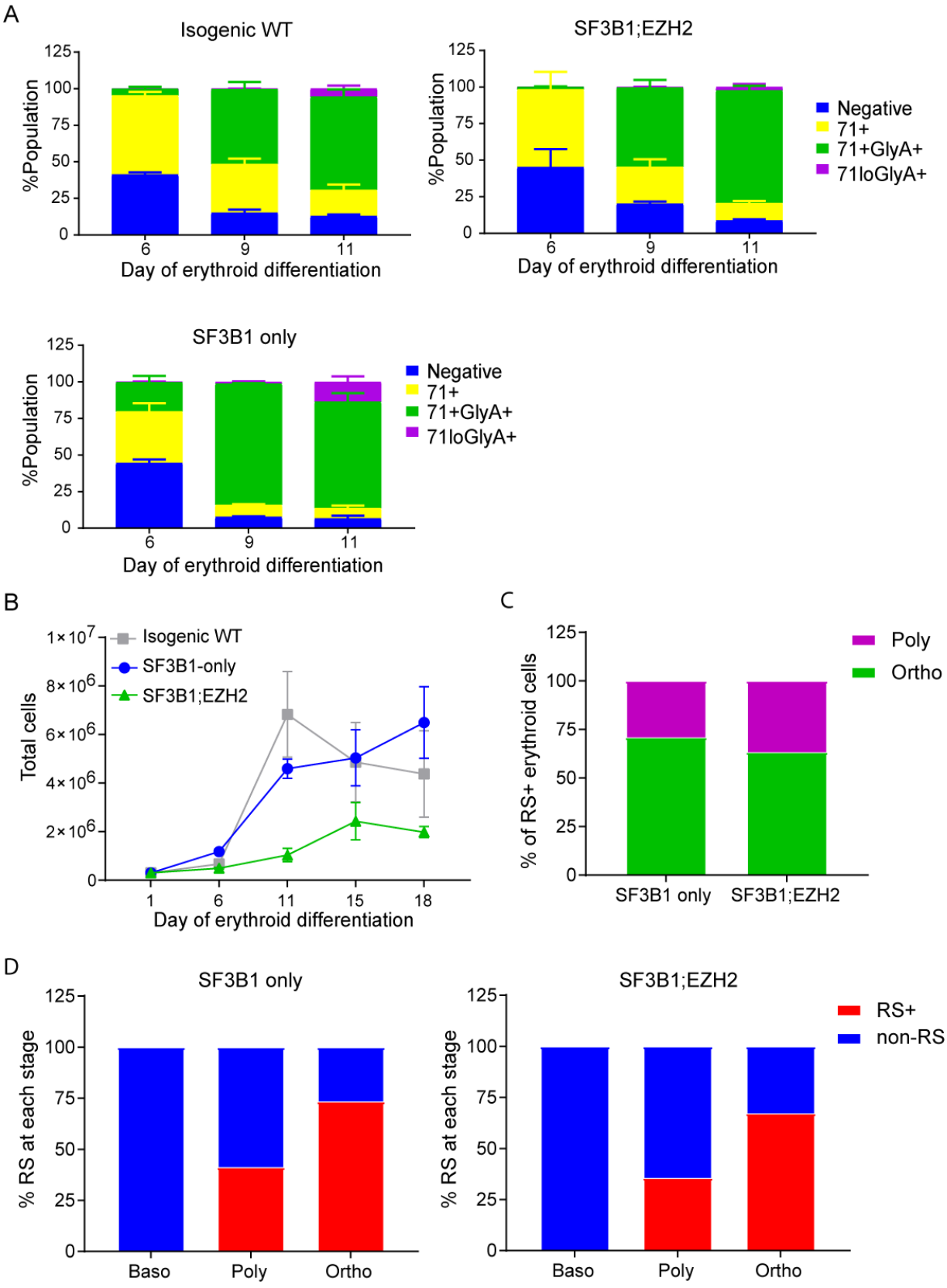

A
Isogenic WT
SF3B1;EZH2
SF3B1 only
%Population
Day of erythroid differentiation
Negative
71+
71+GlyA+
71loGlyA+
B
Isogenic WT
SF3B1-only
SF3B1;EZH2
Total cells
Day of erythroid differentiation
C
% of RS+ erythroid cells
Poly
Ortho
SF3B1 only
SF3B1;EZH2
D
% RS at each stage
RS+
non-RS
Baso
Poly
Ortho

Figure S1. Related to Figure 1.
A. Relative CD71 and GlyA levels as measured by flow cytometry using gating strategy shown in 1B for days 6, 9, and 11 of erythroid differentiation for normal isogenic WT, *SF3B1*$^{G742D/+}$, and *SF3B1*$^{G742D/+}$; *EZH2*$^{R685H/+}$ lines; mean ± SEM of n = 3. **B.** Total erythroid cell number over 18 days of erythroid differentiation of 5F-HPC derived erythroid cells. **C.** Quantification of ring sideroblast erythroid cell morphology. Erythroid morphology was evaluated in RS cells stained with Perls Prussian blue on day 18 of culture. Poly = polychromatic, ortho = orthochromatic erythroblasts; a representative experiment was counted. **D.** Representative quantification of ring sideroblasts (RS+) at each stage of differentiation defined by erythroid cell morphology. Erythroid morphology and RS were evaluated in cells stained with Perls Prussian blue on day 18 of culture. Baso = basophilic, poly = polychromatic, ortho = orthochromatic erythroblasts; a representative experiment was counted.

Figure S2

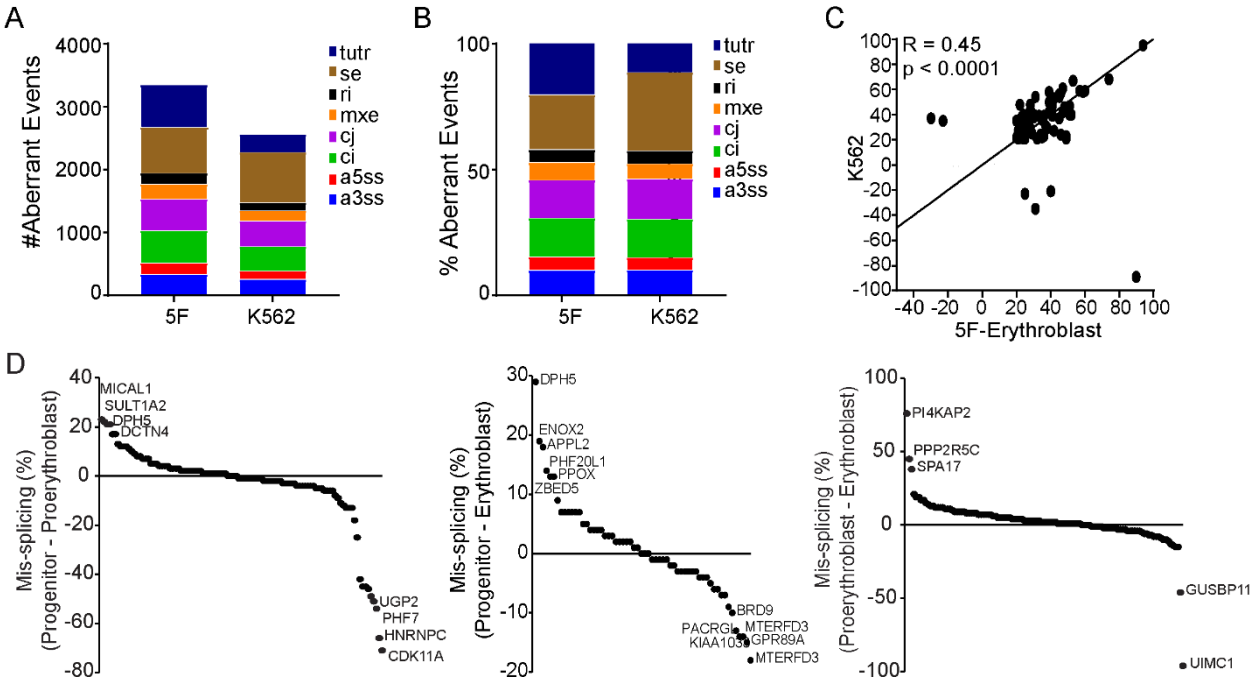


Figure S2. Related to Figure 2.
A,B Total number **(A)** and proportion **(B)** of differentially spliced events in all 5F SF3B1G742D/+ cell populations compared with K562 SF3B1K700E/+ cells (change in isoform ratio ≥ 10%). Splicing events classified as arising from tandem 3' UTRs (tutr), cassette or skipped exons (se), retained introns (ri), mutually exclusive exons (mxe), alternative usage of normally constitutively spliced junctions (cj), alternative retention of normally constitutively spliced introns (ci), alternative 5'ss (a5ss), and alternative 3'ss (a3ss). **C**. Spearman correlation between the quantitative level of differential splicing in SF3B1G742D/+ 5F erythroid cells compared with K562 SF3B1K700E/+ cells. Line of best fit is plotted. **D.** Differential mis-splicing between 5F-HPCs and erythroid cells, and between $CD71^{+}$ proerythroblasts and $CD71^{+}GlyA^{+}$ erythroblasts. Isoforms are ranked by % mis-splicing in HPCs compared with erythroblasts or by %mis-splicing in proerythroblast compared with erythroblasts respectfully. Genes with mis-splicing ≥ 0.2 and total read count >5 were included in the analysis.

Figure S3

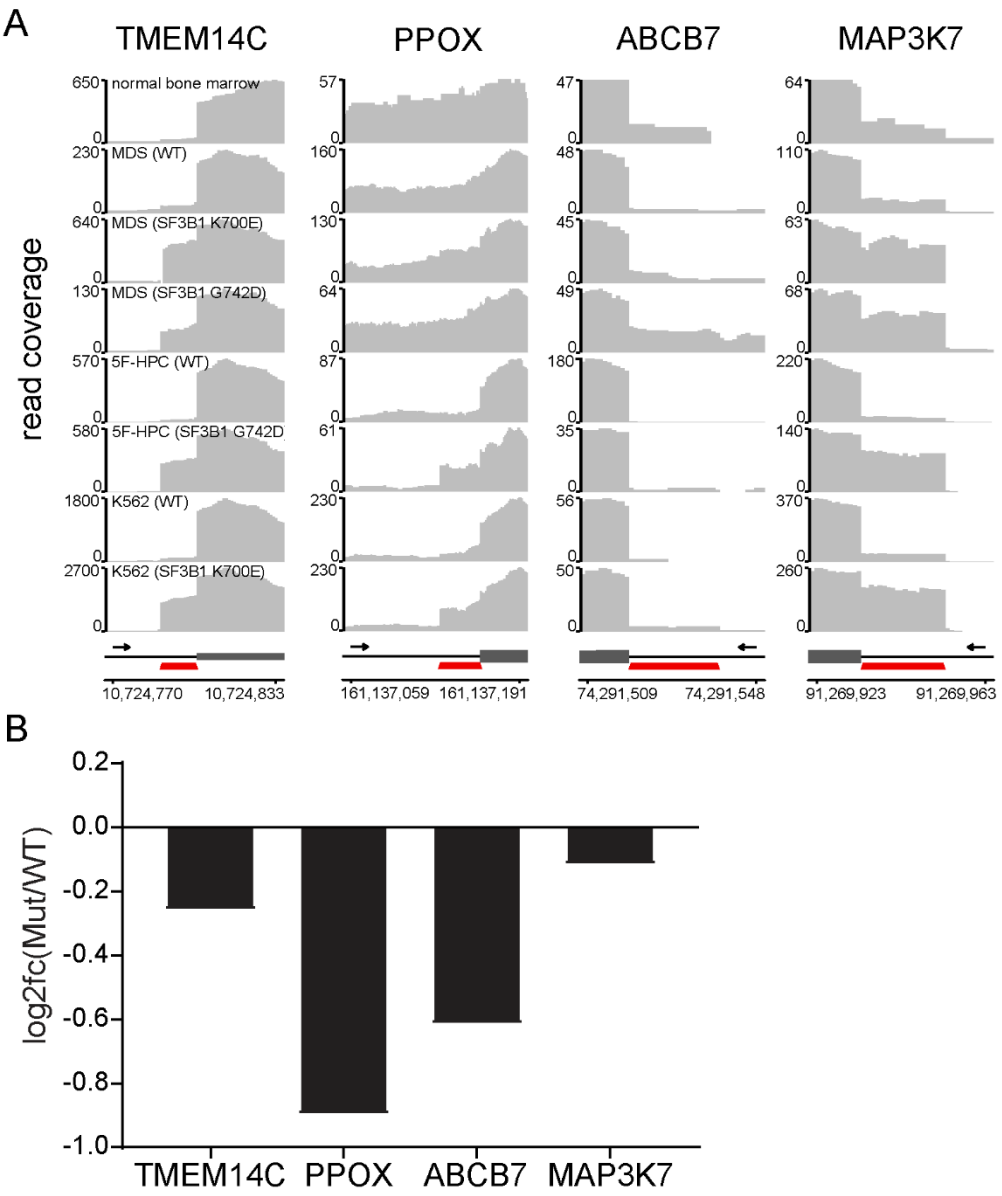


Figure S3. Related to Figure 3.
A. Identification of erythroid-specific mutant *SF3B1* targets based on the level of mis-splicing vs. fold change of transcript expression during normal erythroid differentiation. Selected genes were mis-spliced in both *SF3B1*-mutant MDS patient cells and iPSC 5F-HPC derived erythroid cells. Arrows indicate direction of transcription. Red bar indicates location of aberrant splicing event. Genomic coordinates relative to respective chromosomes are indicated for each mis-splicing event. **B.** Log fold change of protein expression in *SF3B1*-mutant compared to *SF3B1*-wild type 5F-HPCs for the selected candidate mis-spliced genes.

Figure S4

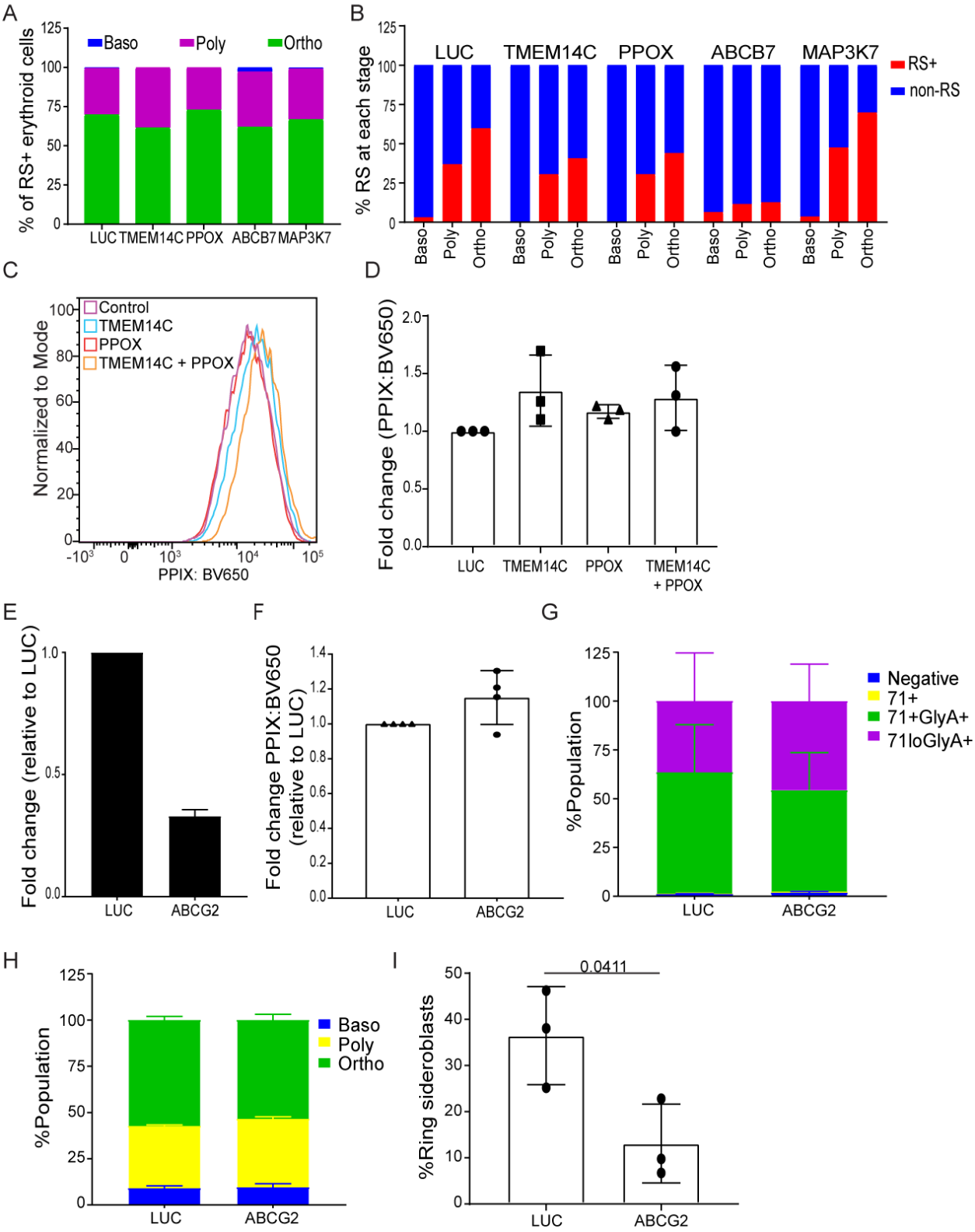
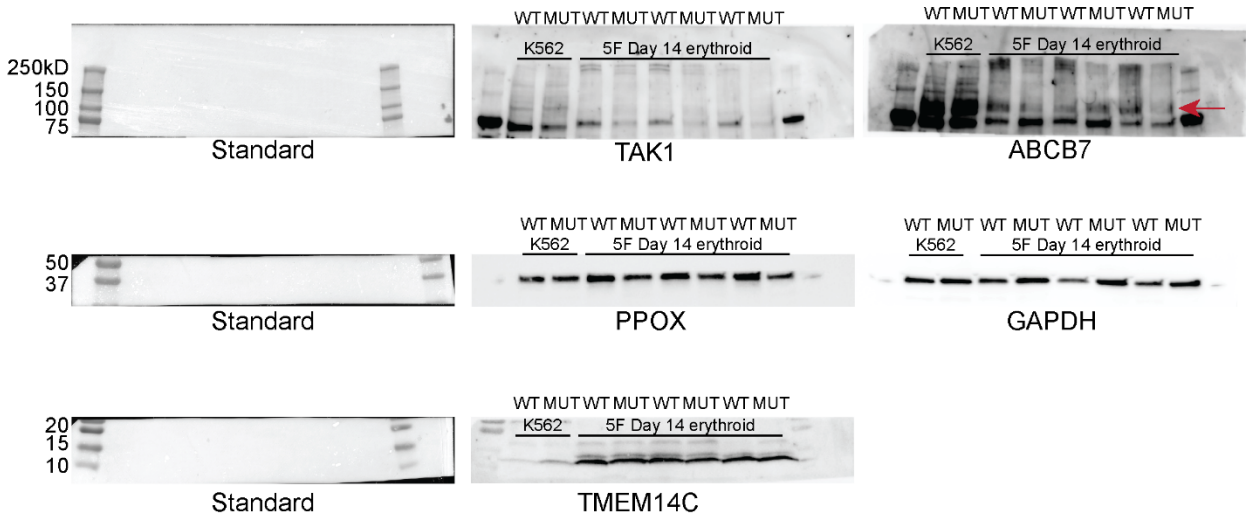

A
% of RS+ erythroid cells
Baso
Poly
Ortho
LUC TMEM14C PPOX ABCB7 MAP3K7
B
% RS at each stage
LUC
TMEM14C
PPOX
ABCB7
MAP3K7
RS+
non-RS
Baso
Poly
Ortho
C
Normalized to Mode
Control
TMEM14C
PPOX
TMEM14C + PPOX
PPIX: BV650
D
Fold change (PPIX:BV650)
LUC
TMEM14C
PPOX
TMEM14C + PPOX
E
Fold change (relative to LUC)
LUC
ABCG2
F
Fold change PPIX:BV650 (relative to LUC)
LUC
ABCG2
G
%Population
Negative
71+
71+GlyA+
71loGlyA+
LUC
ABCG2
H
%Population
Baso
Poly
Ortho
LUC
ABCG2
I
%Ring sideroblasts
0.0411
LUC
ABCG2

Figure S4. Related to Figure 4.
A. Quantification of ring sideroblast erythroid cell morphology. Erythroid morphology was evaluated in RS cells stained with Perls Prussian blue on day 18 of culture. Baso = basophilic, poly = polychromatic, ortho = orthochromatic erythroblasts; a representative experiment was counted. **B.** Representative quantification of ring sideroblasts (RS+) at each stage of differentiation defined by erythroid cell morphology. Erythroid morphology and RS were evaluated in cells stained with Perls Prussian blue on day 18 of culture. Baso = basophilic, poly = polychromatic, ortho = orthochromatic. **C**. Representative flow cytometry plot of PPIX fluorescence measured in the BV650 channel (405 nm excitation, 670/30 nm emission). **D**. Fold-change of PPIX levels in *SF3B1*-mutant cells transduced with TMEM14C and/or PPOX ORFs normalized to *SF3B1*-mutant cells transduced with LUC control; mean ± s.d. n = 3 independent experiments. **E**. Expression of *ABCG2* as measured by quantitative PCR relative in cells transduced with shRNA targeting luciferase control or *ABCG2.* Expression was normalized to LUC shRNA, mean ± s.d., n = 3 independent experiments. **F.** Fold-change of PPIX levels in *SF3B1*-mutant cells transduced with shRNA targeting LUC control or *ABCG2.* Each point represents an independent experiment, mean ± s.d. **G.** Relative CD71 and GlyA levels as measured by flow cytometry at day 18 of erythroid differentiation for *SF3B1*-mutant cells transduced with shRNA targeting LUC or *ABCG2*, mean ± SEM of n = 3. **H.** Quantification of May-Grunwald Giemsa staining of erythroid cell morphology from day 18 of erythroid differentiation; n = 2. **I.** RS counts on day 18 of erythroid differentiation in *SF3B1*-mutant erythroid cells transduced with shRNA constructs targeting luciferase (LUC) control or *ABCG2,* mean ± s.d. of n = 3 independent experiments. P-value indicated, t-test.

Western Blot uncropped images related to Figure 3E



SUPPLEMENTAL FILES

Supplemental File 1. Global RNA-sequencing splicing analysis of 5F-HPCs and K562s. Table includes Ensembl, geneID, descriptions, and genomic coordinates for each gene identified as mis-spliced. For each splicing event, genomic coordinates, splicing type, isoform, spliceosome information and nonsense mediated RNA decay prediction is annotated. The degree to mis-splicing (dpsi) and total splicing (psi) are listed for each stage of 5F-HPC $CD34^+$ progenitors, $CD71^+$ proerythroblasts, $CD71^+GlyA^+$ erythroblasts, and K562 cells. Log fold-change of splicing events and respective Bayes factor for each cell line is also indicated. Total RNA expression as measured by TPM is indicated.

Supplemental File 2. Global proteomics analysis of 5F-HPC progenitor cells. Table includes accession, Pfam ID, Entrez Gene ID, Ensembl Gene ID, Gene symbol ID and description for each gene. Total peptide levels for SF3B1 WT and SF3B1-G742D mutant cells and respective fold change are indicated. Table also includes various mass spec quality control analyses.

REFERENCES

1. Inoue D, Chew G-L, Liu B, et al. Spliceosomal disruption of the non-canonical BAF complex in cancer. *Nature*. 2019-10-17 2019;574(7778):432-436. doi:10.1038/s41586-019-1646-9
2. Flicek P, Ahmed I, Amode MR, et al. Ensembl 2013. *Nucleic Acids Research*. 2012-11-30 2012;41(D1):D48-D55. doi:10.1093/nar/gks1236
3. Goldman M, Craft B, Swatloski T, et al. The UCSC Cancer Genomics Browser: update 2013. *Nucleic Acids Research*. 2013-01-01 2013;41(D1):D949-D954. doi:10.1093/nar/gks1008
4. Katz Y, Wang ET, Airoldi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods*. 2010-12-01 2010;7(12):1009-1015. doi:10.1038/nmeth.1528
5. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011-12-01 2011;12(1):323. doi:10.1186/1471-2105-12-323
6. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009-05-01 2009;25(9):1105-1111. doi:10.1093/bioinformatics/btp120
7. Wagenmakers E-J, Lodewyckx T, Kuriyal H, Grasman R. Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*. 2010-05-01 2010;60(3):158-189. doi:10.1016/j.cogpsych.2009.12.001
8. Nikpour M, Scharenberg C, Liu A, et al. The transporter ABCB7 is a mediator of the phenotype of acquired refractory anemia with ring sideroblasts. *Leukemia*. 4 2013;27:889-896. doi:10.1038/leu.2012.298
9. Tang Y, Hou J, Li G, et al. ABCG2 regulates the pattern of self-renewing divisions in cisplatin-resistant non-small cell lung cancer cell lines. *Oncology Reports*. 2014-11-01 2014;32(5):2168-2174. doi:10.3892/or.2014.3470