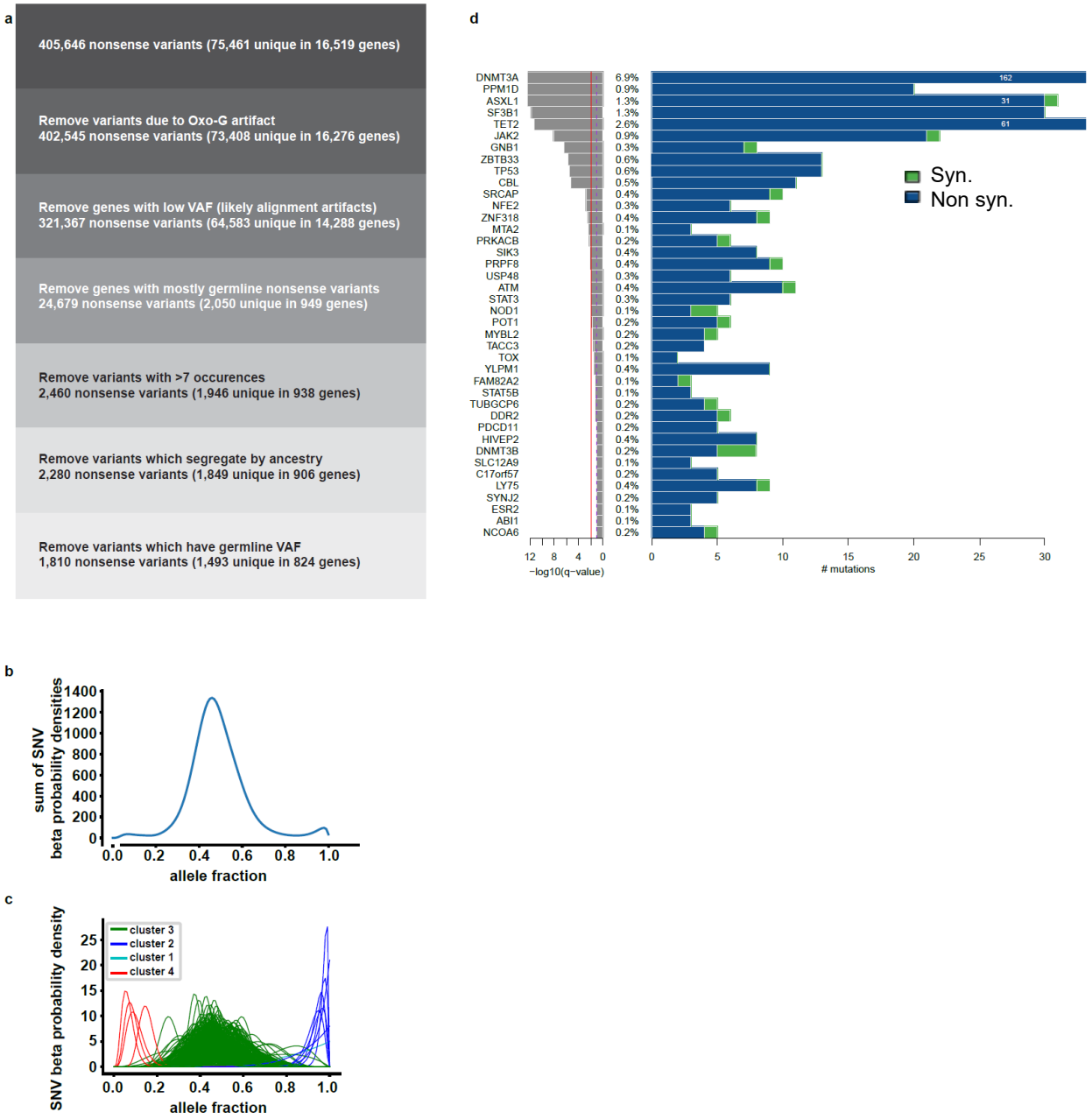
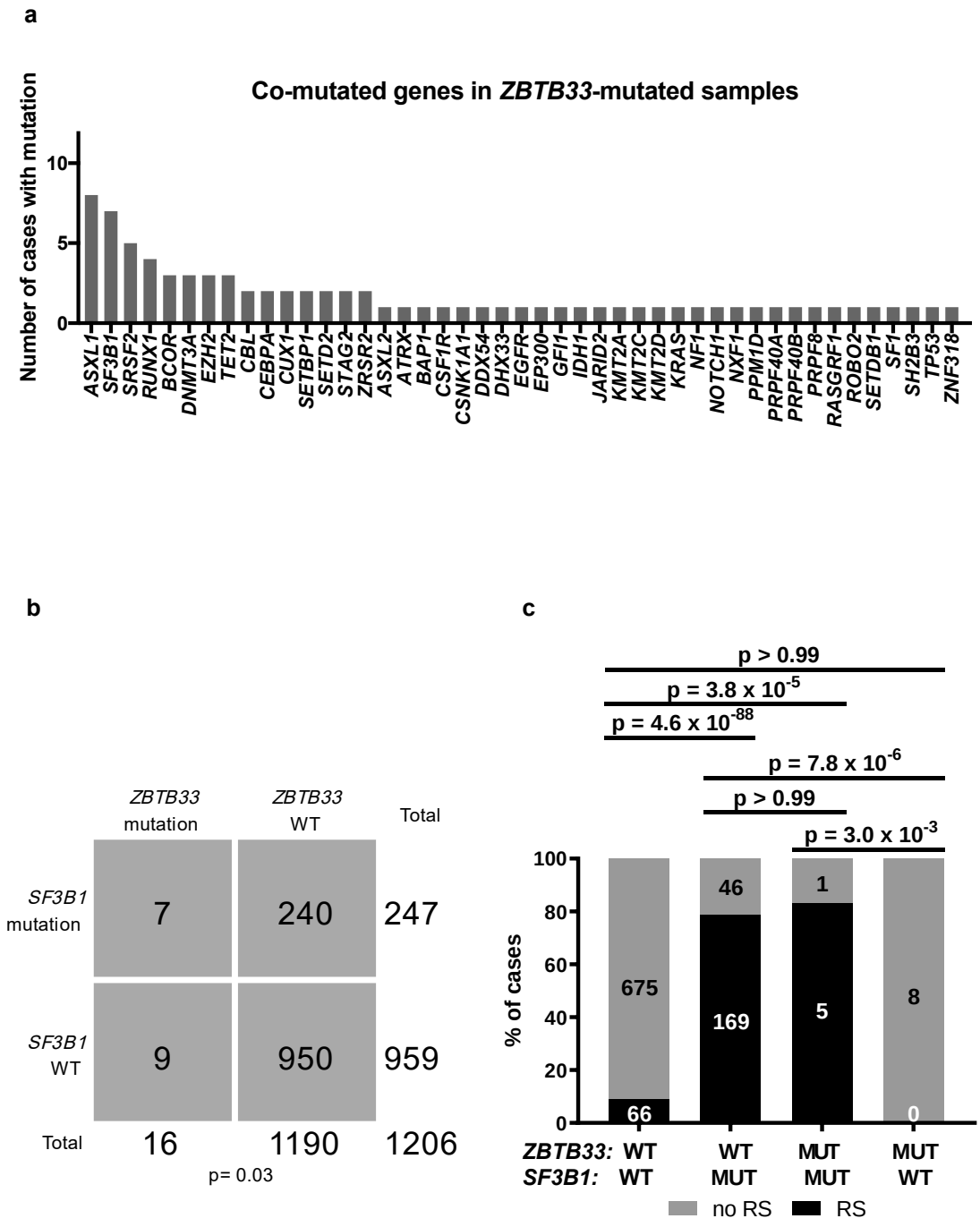
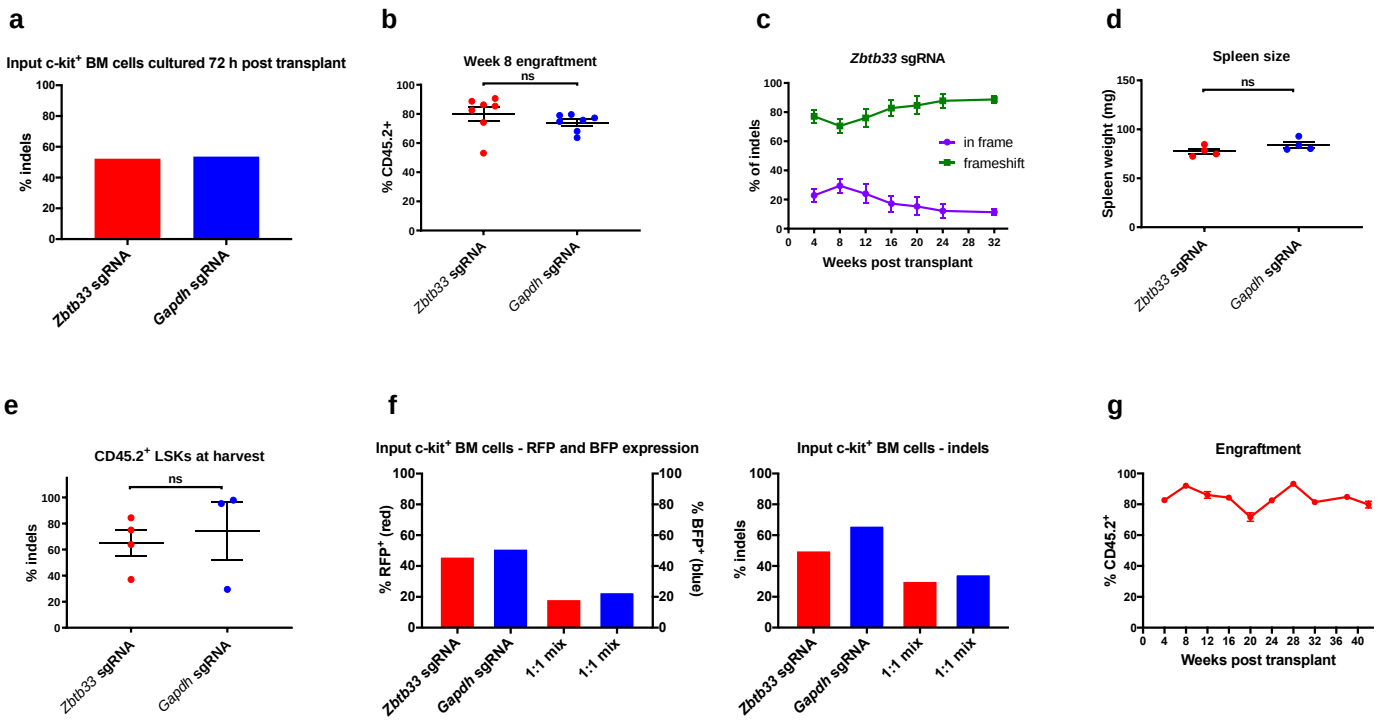


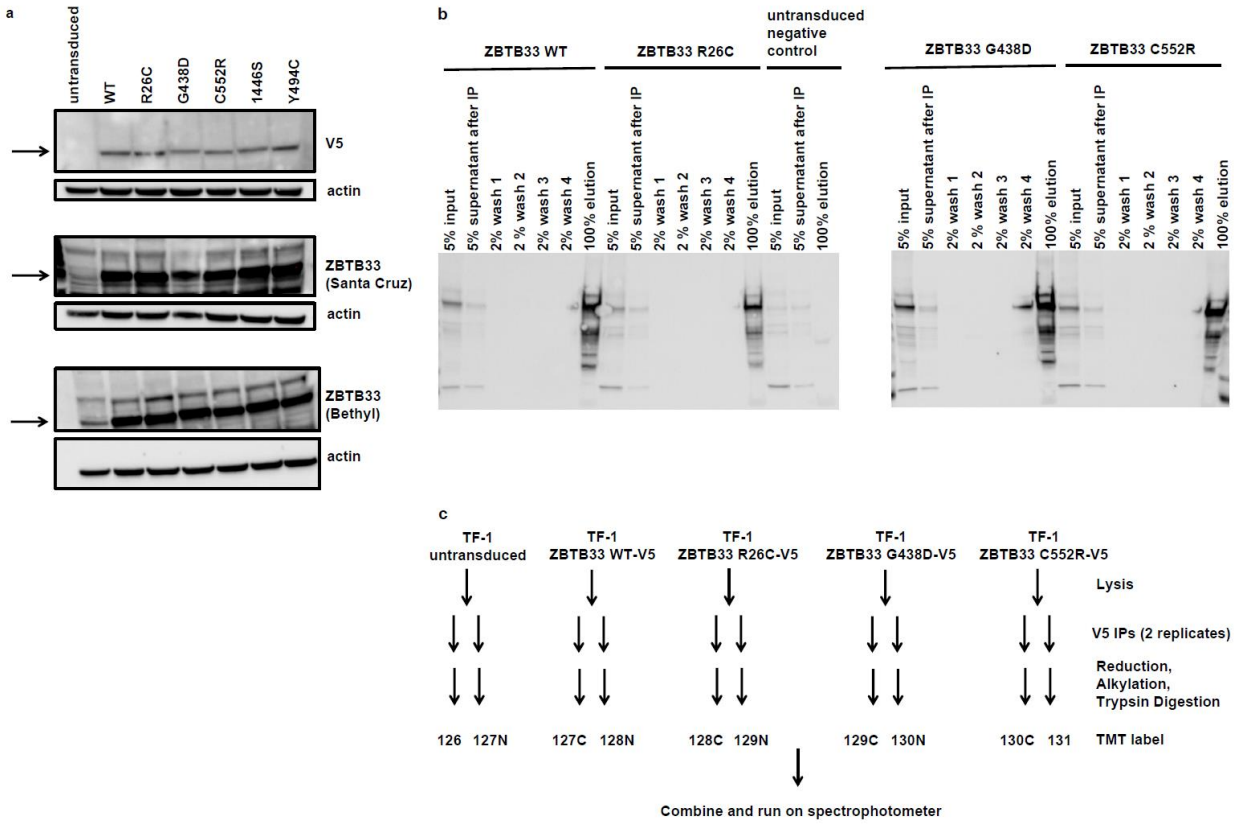
Supplementary Fig. S1: Identification of novel candidate CHIP genes.
a, Schematic depicting strategy for nonsense somatic mutation screen. **b**, Plot showing the sum of beta probability density functions of single nucleotide variants (SNVs) for a representative sample. The majority of SNVs are germline heterozygous variants with VAFs of about 0.5. A smaller number of SNVs are somatic mutations with lower VAFs. **c**, Representative depiction of Dirichlet clustering of the individual SNV beta distributions for each variant in a given sample. Clustering distinguishes germline heterozygous and homozygous mutations (shown in green, cyan, and blue) from somatic clonal mutations (red). For our analyses, we considered only the “clonal somatic heterozygous cluster,” defined as the somatic cluster with the largest purity and at least 4 mutations. **d**, Plot showing significantly mutated ($q<0.1$) genes in the clonal somatic heterozygous clusters of 2266 samples.



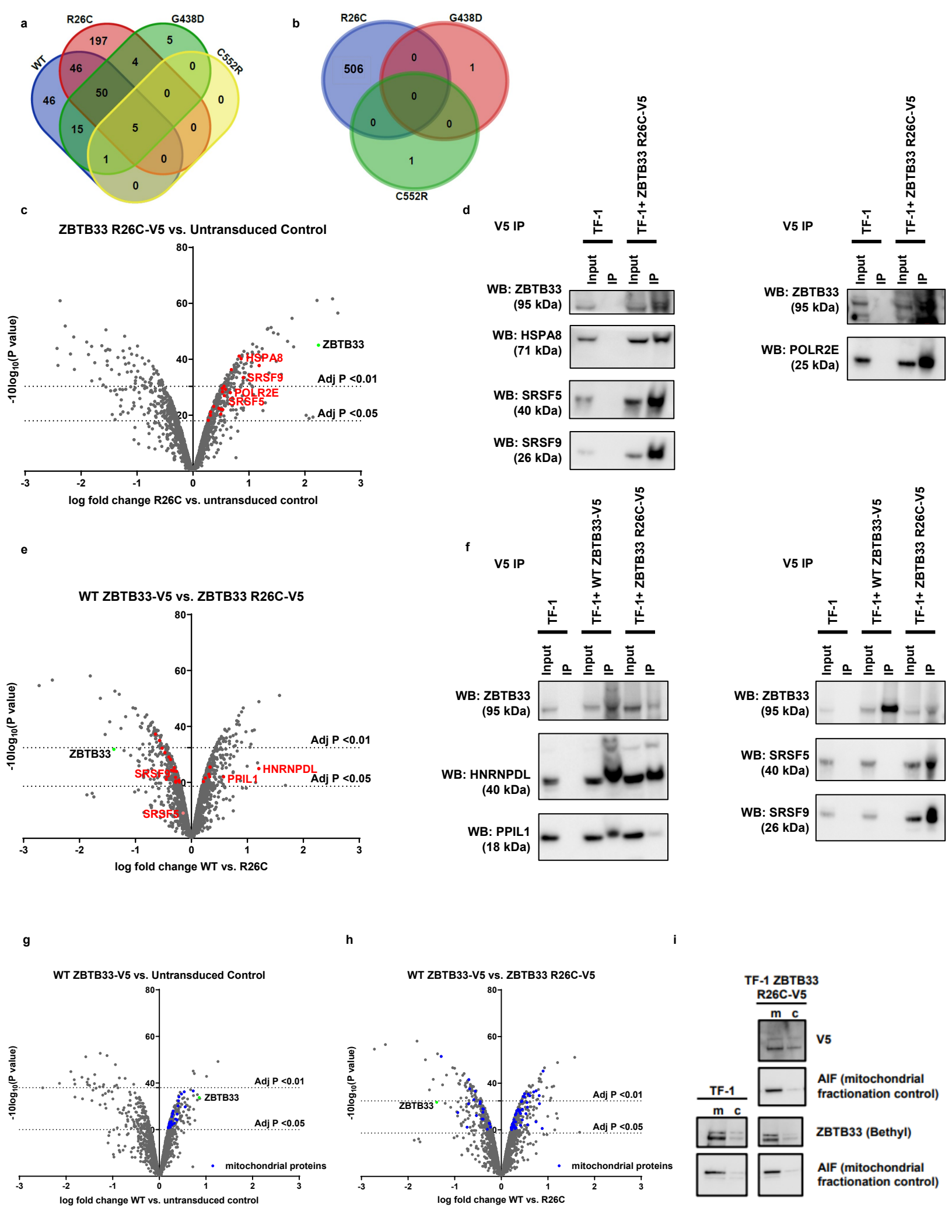
Supplementary Fig. S2: Identification of mutations in candidate genes in MDS patients. a, Bar chart depicting the full co-mutation spectrum of the 16 *ZBTB33* mutated cases. **b**, 2x2 contingency table showing the number of cases with and without *ZBTB33* and/or *SF3B1* mutations. *P* value = 0.03, calculated using Fisher's Exact Test. **c**, Graph depicting the percentages and number of cases of different genotypes with and without ring sideroblasts (RS). *P* values were computed using Fisher's Exact Tests.



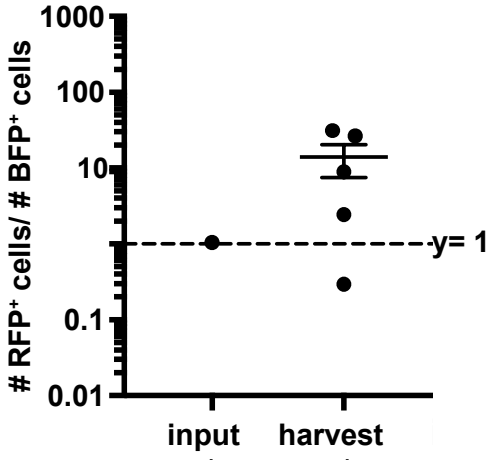
Supplementary Figure S3: Expansion of *Zbtb33* edited mouse HSPCs.
a, Bar graphs plotting the percentage of sequencing reads with indels near the CRISPR cut sites for transduced c-kit$^+$ transplant donor cells. **b**, Graph plotting the percentage of peripheral blood cells expressing CD45.2 in individual recipient mice 8 weeks post transplant. *P* value = 0.28, computed with two-tailed unpaired *t* test. **c**, Graphs plotting the relative amounts of in frame versus frameshift indels over time. **d**, Graph plotting the spleen weights of recipient mice at the conclusion of the experiment, 44 weeks post transplant. *P* value = 0.15, computed with two-tailed unpaired *t* test. **e**, Graphs plotting the percentage of sequencing reads with indels near the CRISPR cut sites for CD45.2$^+$ LSKs, FACS sorted from recipient mice BM. *P* value = 0.70, computed with two-tailed unpaired *t* test. **f**, Comparison of editing in competitive transplant donor cells 72 hours post transplant, as measured by RFP and BFP expression by FACS and % indels by sequencing. Shown are the cells transduced with each sgRNA and the 1:1 mix of cells that was transplanted into recipient mice. **g**, Graph plotting the percentage of CD45.2$^+$ donor cells in recipient peripheral blood over time. All error bars in each panel represent s.e.m.



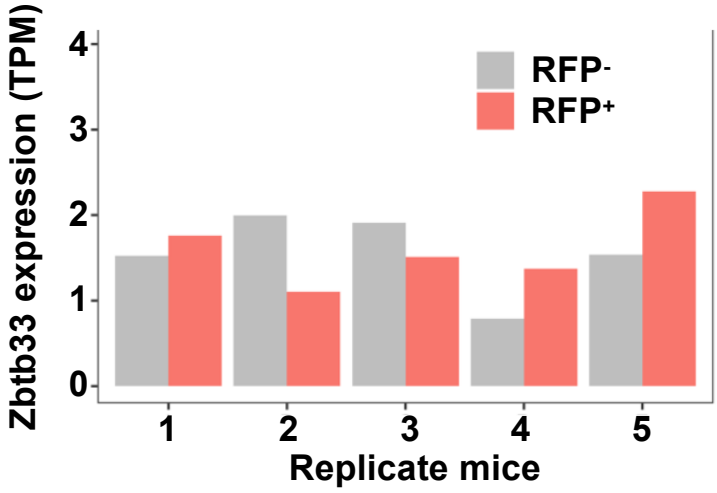
Supplementary Fig. S4: Experimental design and optimization for V5 IP/MS experiment.
a, Western blots performed to overexpression of V5-tagged wildtype or mutant ZBTB33 protein in TF-1 cells. **b**, Western blots performed to confirm optimal IP of V5-tagged WT and mutant ZBTB33. IPs were performed with anti-V5 antibody, and western blots were performed with anti-ZBTB33 antibody, allowing for assessment of relative amount of total (including endogenous) ZBTB33 remaining in supernatant compared to input. **c**, Schematic depicting experimental design. V5 IPs were performed in duplicate, followed by reduction, alkylation, overnight trypsin digest, and TMT labeling. TMT Labeled peptides were combined before being run on the spectrophotometer.


a
R26C
G438D
WT
C552R
197
5
46
4
0
46
50
0
0
15
5
0
1
0
0
b
R26C
G438D
506
0
1
0
0
0
1
C552R
c
ZBTB33 R26C-V5 vs. Untransduced Control
-10log10(P value)
log fold change R26C vs. untransduced control
ZBTB33
HSPA8
SRSF9
POLR2E
SRSF5
Adj P <0.01
Adj P <0.05
d
V5 IP
TF-1
TF-1+ ZBTB33 R26C-V5
Input
IP
WB: ZBTB33 (95 kDa)
WB: HSPA8 (71 kDa)
WB: SRSF5 (40 kDa)
WB: SRSF9 (26 kDa)
V5 IP
TF-1
TF-1+ ZBTB33 R26C-V5
Input
IP
WB: ZBTB33 (95 kDa)
WB: POLR2E (25 kDa)
e
WT ZBTB33-V5 vs. ZBTB33 R26C-V5
-10log10(P value)
log fold change WT vs. R26C
ZBTB33
SRSF9
SRSF5
HNRNPDL
PPIL1
Adj P <0.01
Adj P <0.05
f
V5 IP
TF-1
TF-1+ WT ZBTB33-V5
TF-1+ ZBTB33 R26C-V5
Input
IP
WB: ZBTB33 (95 kDa)
WB: HNRNPDL (40 kDa)
WB: PPIL1 (18 kDa)
V5 IP
TF-1
TF-1+ WT ZBTB33-V5
TF-1+ ZBTB33 R26C-V5
Input
IP
WB: ZBTB33 (95 kDa)
WB: SRSF5 (40 kDa)
WB: SRSF9 (26 kDa)
g
WT ZBTB33-V5 vs. Untransduced Control
-10log10(P value)
log fold change WT vs. untransduced control
ZBTB33
Adj P <0.01
Adj P <0.05
mitochondrial proteins
h
WT ZBTB33-V5 vs. ZBTB33 R26C-V5
-10log10(P value)
log fold change WT vs. R26C
ZBTB33
Adj P <0.01
Adj P <0.05
mitochondrial proteins
i
TF-1 ZBTB33 R26C-V5
m
c
V5
AIF (mitochondrial fractionation control)
TF-1
m
c
ZBTB33 (Bethyl)
AIF (mitochondrial fractionation control)

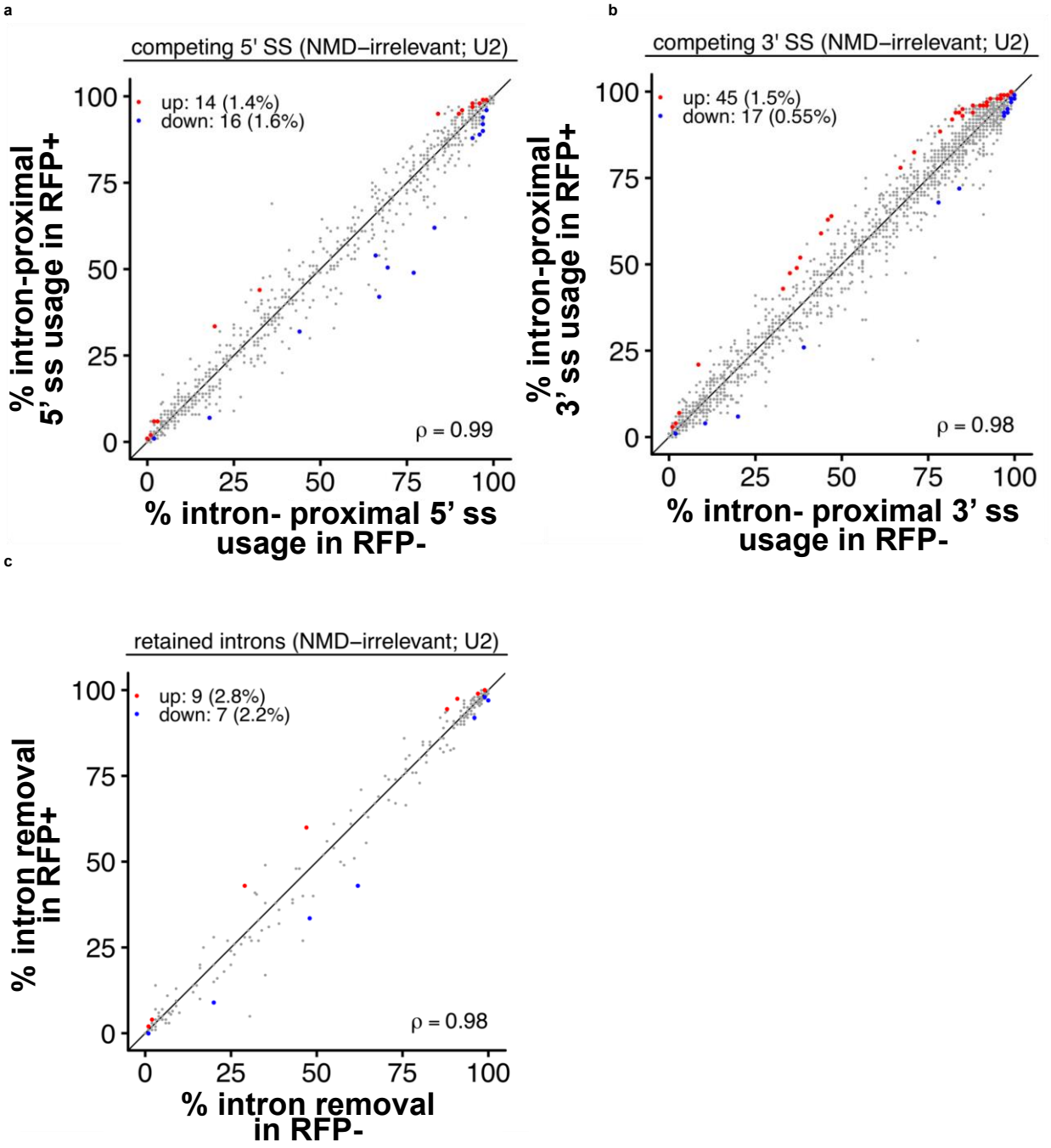
Supplementary Fig. S5: Identification of protein interaction partners of WT and mutant ZBTB33 by IP/MS. **a**,**b**, Venn diagrams plotting the number of protein interacting partners that were (**a**) significantly enriched (Adj. $p<0.05$) in each IP compared to the untransduced control IP or (**b**) differentially enriched ($p<0.05$) in each mutant ZBTB33-V5 IP compared to the WT ZBTB33-V5 IP. **c**, Volcano plot highlighting selected splicing associated proteins that were most highly enriched in ZBTB33 R26C V5 IPs compared to control IPs. Splicing associated proteins are colored red, and proteins validated by IP/WB are labeled. **d**, IP/western blots performed to validate interactions highlighted in **c**. **e**, Volcano plot highlighting selected splicing associated proteins that were most differentially enriched between ZBTB33 WT and ZBTB33 R26C IPs. Splicing associated proteins are colored red, and proteins validated by IP/WB are labeled. **f**, IP/western blots performed to validate interactions highlighted in **e**. **g**-**h**, Volcano plots visualizing significant protein interacting partners enriched in the WT ZBTB33-V5 IP compared to control (**g**) and differentially enriched in the WT-V5 and R26C-V5 IPs (**h**). Proteins which localize to the mitochondria are colored blue. **i**, Western blots for V5 or ZBTB33 performed following mitochondrial fractionation of TF-1 or TF-1 ZBTB33 R26C-V5 cells. m: mitochondrial fraction. c: cytoplasmic fraction.



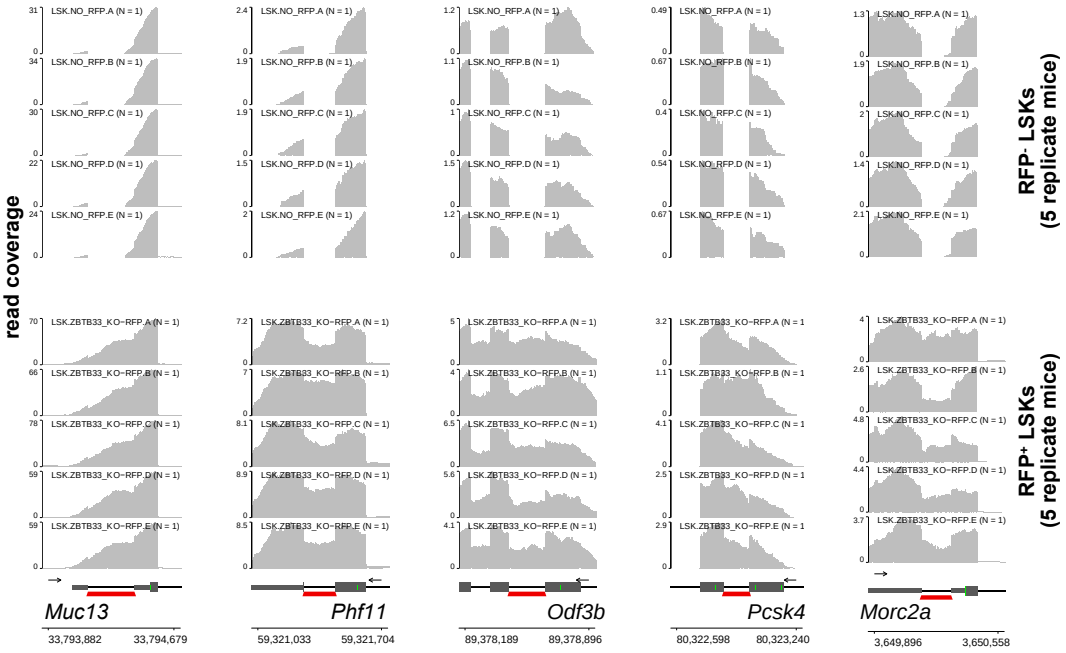
Supplementary Fig. S6: Confirmation of expansion of RFP$^+$ LSKs.
Graph plotting the ratio of RFP$^+$:BFP$^+$ cells in recipient LSKs harvested 38 weeks post transplant compared to donor transplant input LSKs. Bars represent mean and s.e.m. Harvested RFP$^+$ and RFP$^-$ LSKs from this experiment were used for RNA-seq.



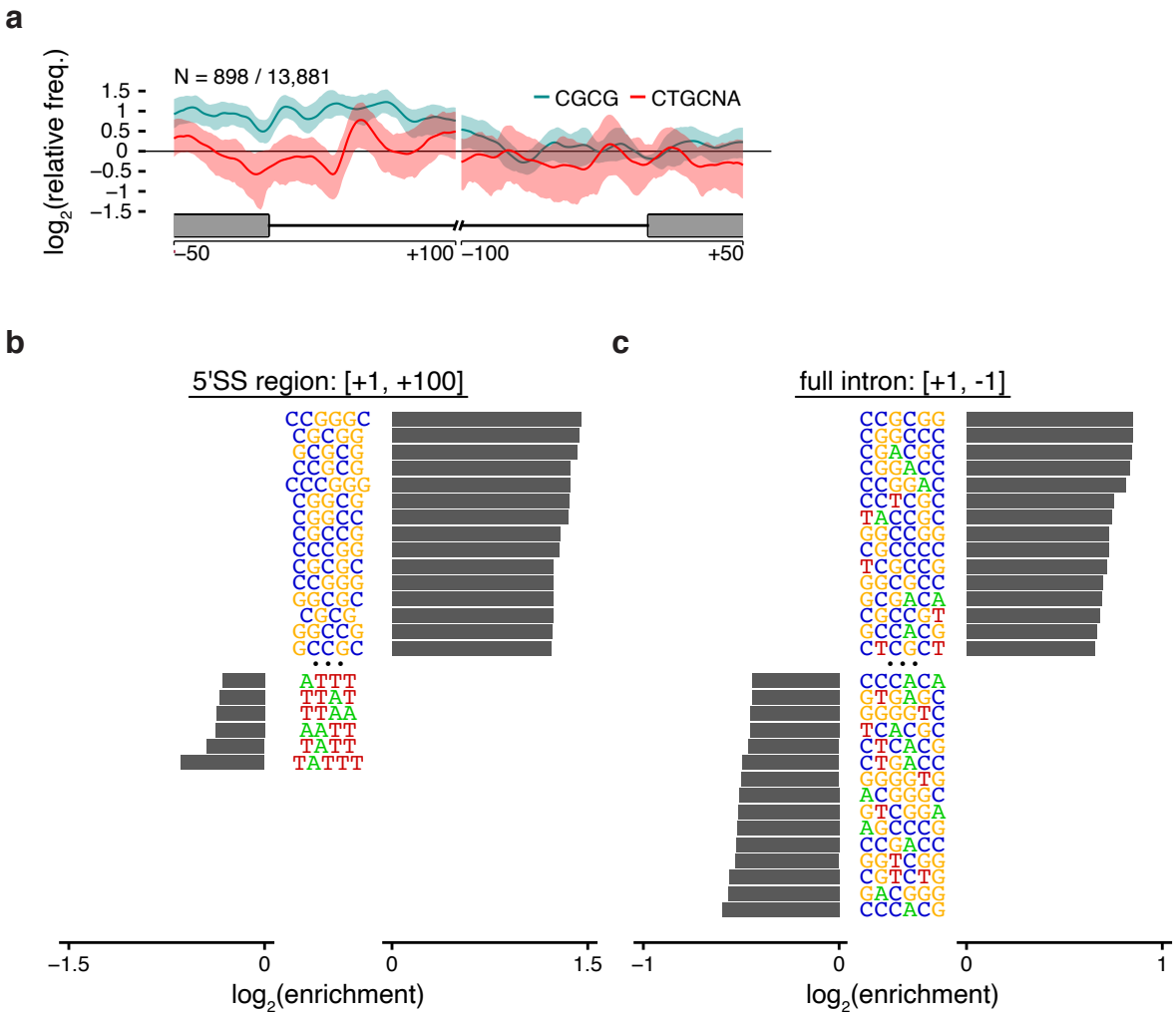
Supplementary Fig. S7: Normalized *Zbtb33* expression in RFP$^+$ and RFP$^-$ LSKs from 5 individual mice estimated from RNA-seq.



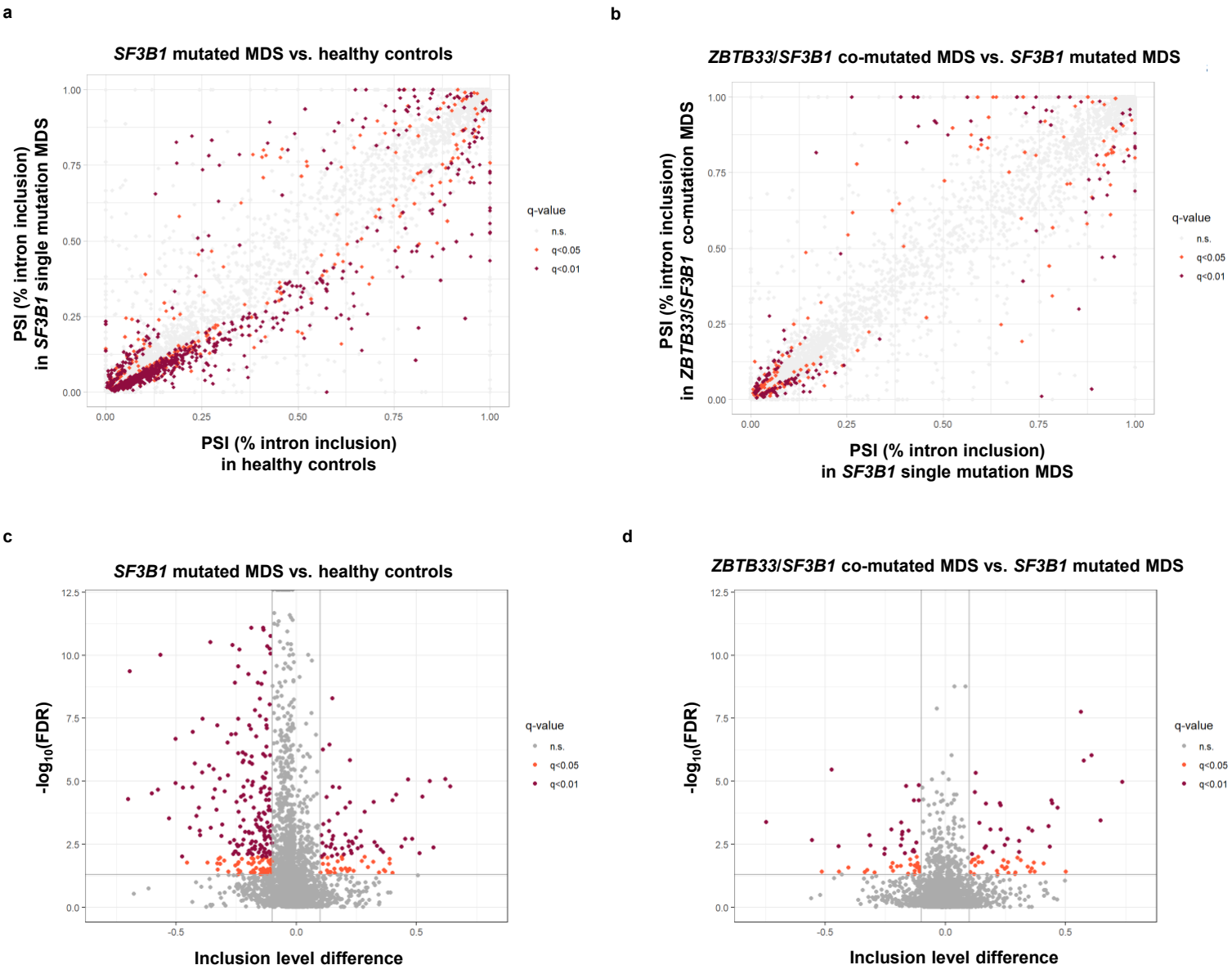
Supplementary Fig. S8: Alternative splicing in *Zbtb33* edited versus control LSKs.
Scatter plots comparing (**a**) 5’ splice site usage, (**b**) 3’ splice site usage, and (**c**) intron removal of frequently retained introns in *Zbtb33* edited RFP+ LSKs versus RFP- control LSKs. Red and blue dots represent individual events with higher and lower usage, respectively, of the intron-proximal alternative splice site or the intron removal isoform that met the specified thresholds (see Methods) in *Zbtb33* edited versus control cells.



Supplementary Fig. S9: Identification of intron retention events in *Zbtb33* edited LSKs.
Example RNA-seq coverage plots corresponding to 5 selected intron retention events with the highest fold change between RFP^{+} and RFP^{-} cells. Read counts are scaled to reads per million.



Supplementary Fig. S10: Motif enrichment at intron retention events in *Zbtb33* edited LSKs.
a, Plots depicting enrichment of previously described ZBTB33 motifs across the regions flanking the 5' and 3' splice sites of constitutive U2 introns. **b**,**c**, ab initio motif enrichment for motifs of length 4, 5, or 6 nucleotides in the 5' splice site adjacent region (**b**) or the full intron (**c**).



Supplementary Fig. S11: Intron retention in *ZBTB33*/*SF3B1* co-mutated MDS, *SF3B1* mutated MDS, and healthy control CD34⁺ cells.
a,**b**, Scatter plots comparing PSI values for intron retention events between *SF3B1* mutated MDS samples and healthy controls (**a**) and between *ZBTB33/SF3B1* co-mutated MDS samples and *SF3B1* mutated MDS samples (**b**). **c**,**d**, Plots comparing inclusion level differences for intron retention events between *SF3B1* mutated MDS samples and healthy controls (**c**) and between *ZBTB33*/*SF3B1* co-mutated MDS samples and *SF3B1* mutated MDS samples (**d**).