Supplemental Data

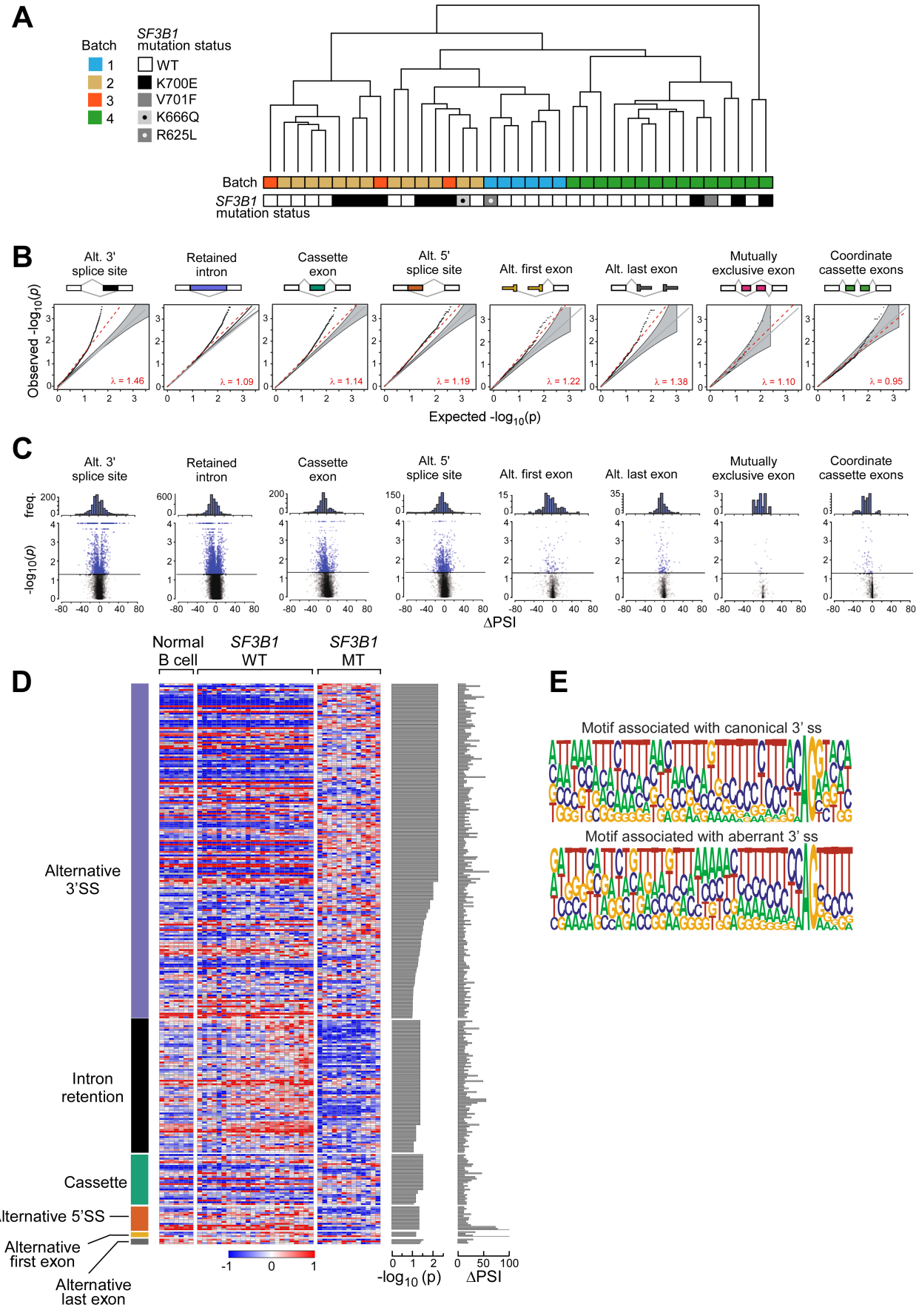

Figure S1, related to Figure 1. Batch effect and alternative splicing characterization from 37 samples used in the study.

(A) Unsupervised hierarchical clustering of 37 CLL samples for analyses of bulk RNA sequencing of bulk poly-A selected libraries, with batches and mutation sites indicated.
(B) Q-Q plots comparing observed empirical p values with expected p values given a uniform distribution of differential splicing events between *SF3B1* wild-type and mutant CLL samples. Red line - the least-squares linear fit to the lower 95 percentile of points and the λ is the slope. Grey shaded areas indicate 95% confidence intervals for the expected distribution.
(C) Frequency of PSI changes associated with *SF3B1* mutation (top) from RNA-Seq data derived from the 37 CLL samples, with volcano plots of ΔPSI versus $-\log_{10}(p)$ of all splicing changes (bottom).
(D) Heat map of 304 significantly alternatively spliced events associated with *SF3B1* mutation using JuncBASE.
(E) Motif frequency plots for canonical, aberrant, and first non-GAG AGs for 3' splice sites. The motifs are given 35 nt upstream of the 3' AG and 3 nt downstream.

Table S1, **related to Figures 1.** Supplied as an Excel file. Summary of CLL samples used for RNA-Seq and validation of splice variant expression.

Table S2, **related to Figure 1.** Supplied as an Excel file. Altered splice variants significantly associated with *SF3B1* mutation in CLL samples.

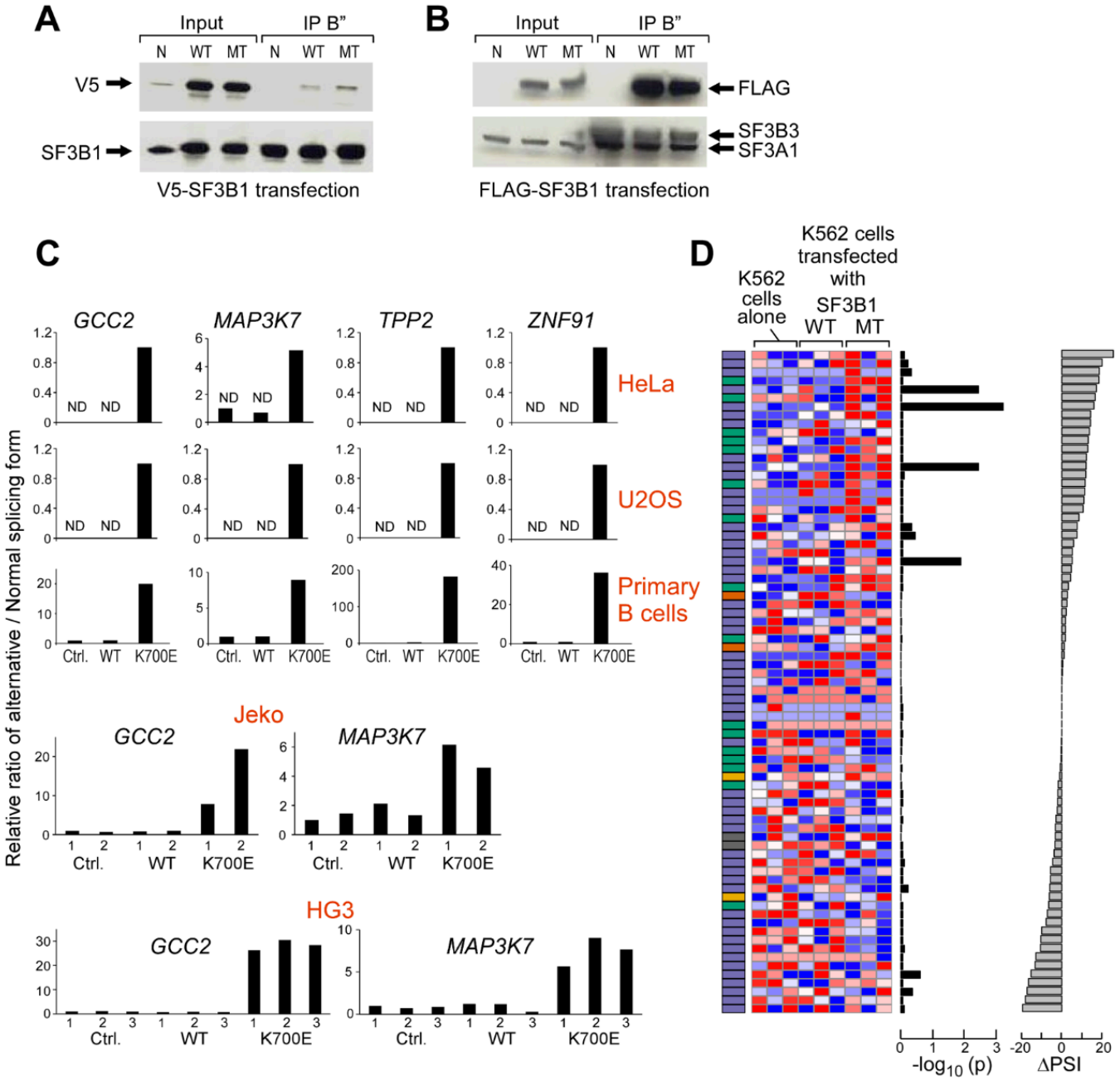

Figure S2, related to Figure 2. Expression of SF3B1-K700E using N-terminal-tagged SF3B1 construct in different cell lines is associated with altered splicing.

(A, B) Binding of transiently expressed SF3B1 protein with other protein in the U2 snRNA complex was examined in HeLa cells. Nuclear extract from the HeLa cells transfected with either N-terminal or C-terminal tag SF3B1 was immunoprecipitated with an antibody against B" which pulls down the U2 snRNA complex (provided by Dr. Robin Reed, Harvard Medical School) (Mattaj et al., 1986), and the immunocomplexes were subjected to SDS-PAGE. Immunoblots were probed with antibodies against V5 and SF3B1 (A), FLAG, SF3B3 and SF3A1 (B), respectively.
(C) Expression of splicing in *GCC2*, *MAP3K7*, *TPP2*, and *ZNF91* was examined using quantitative Taqman RT-PCR assays in HeLa, U2OS, JeKo, HG3 and primary B cells that were either transfected or nucleofected with either empty vector (control), wild-type or mutant SF3B1 constructs.

(D) Heat map of 77 significant alternatively spliced events associated with SF3B1 mutation in transfected K562 cells analyzed by JuncBASE. The category of splice variant is annotated by the color panels to the left of each heat map, while p value for each individual event is provided on the right-sided tracks.

Table S3, related to Figure 2. Supplied as an Excel file. Altered splice variants significantly associated with *SF3B1* mutation in K562 cells.

Table S4, related to Figure 3. Supplied as an Excel file. Primer sequences for detection of splice variants used in the single CLL cell.

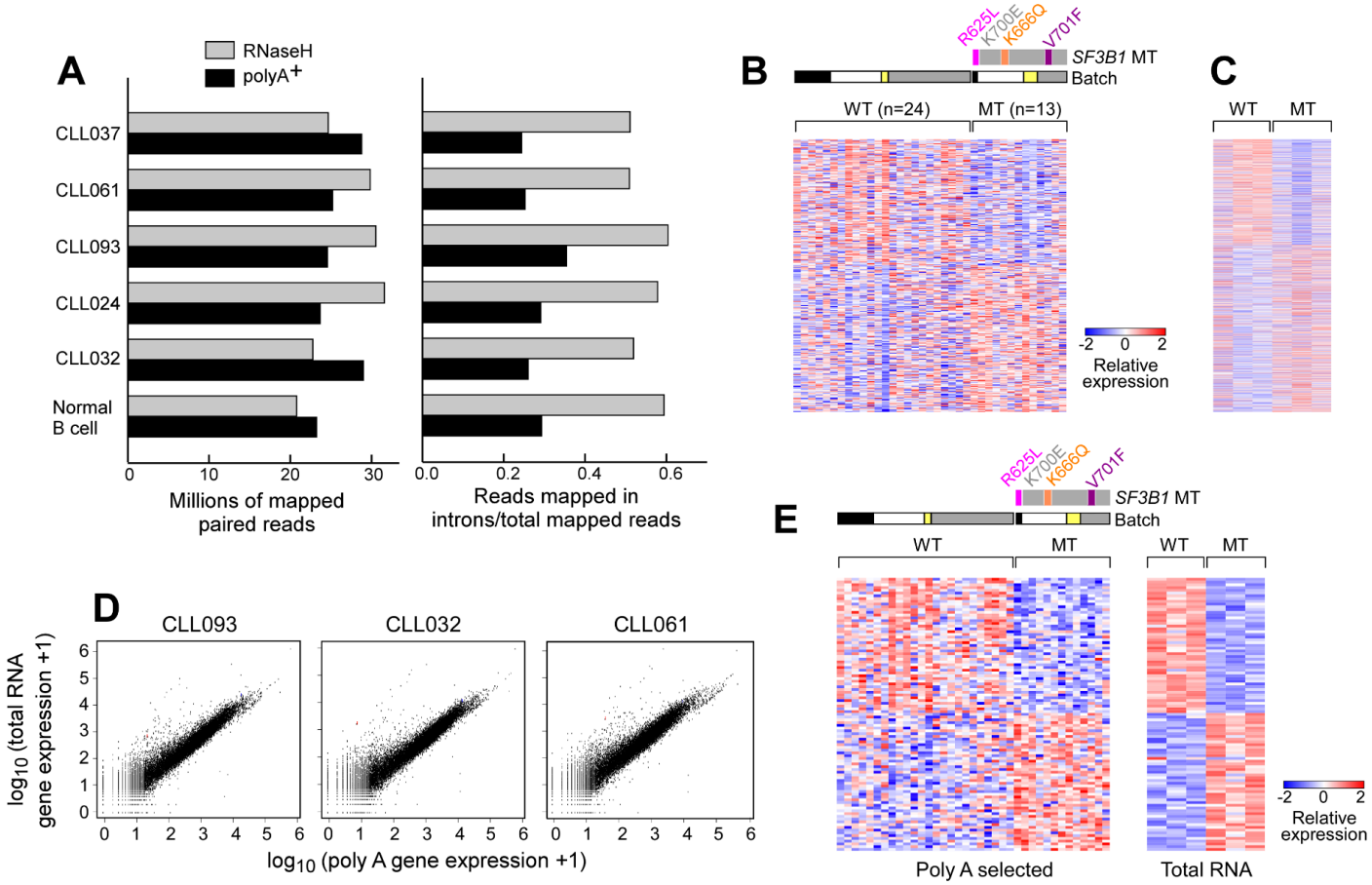

Figure S3, related to Figure 4. Comparisons between poly-A selected and total RNA libraries.

(A) Comparison of intronic rate between total RNA and poly-A RNA-Seq libraries from the same sample. The intronic rate (right) is the number of reads mapped within annotated introns over the total number of mapped reads (left) in the sample.
(B) Heatmap of differentially expressed genes identified from RNA-seq of total RNA libraries in samples from poly-A selected libraries.
(C) Heatmap of differentially expressed genes identified from RNA-seq analysis of poly-A selected libraries in samples from total RNA libraries.
(D) Correlation between total and poly-A RNA libraries in 3 CLL samples.
(E) Visualization of expression of those genes identified as at the intersection between differentially expressed genes identified from the poly-A and total RNA libraries.

Table S5, related to Figure 4. Supplied as an Excel file. Alternative splicing events significantly associated with SF3B1 mutation and with branchpoint support identified from non-poly A selected RNA-Seq.

Table S6, related to Figure 4. Supplied as an Excel file. Genes significantly associated with *SF3B1* mutation identified from poly-A selected and total RNA libraries.

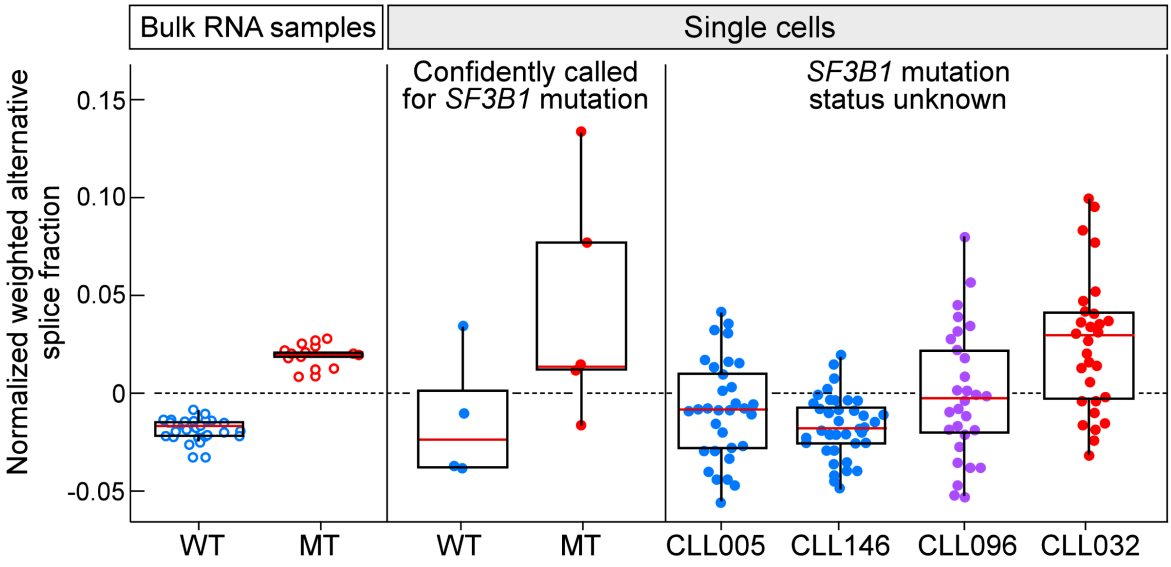

Figure S4, related to Figure 5. Inference of single cell *SF3B1* mutation status by splice variants.

The ability of expression of splice variants to enable confident call for *SF3B1* mutation status in the single cell transcriptome analysis was explored. A trend in which a collection of *SF3B1* mutation associated splice variants (identified from bulk RNA-seq analysis) generally could, on average, distinguish between single cells derived from bulk samples with clonal *SF3B1* mutant (CLL032) and wild-type (CLL005, CLL146) status. The normalized weighted average alternative splice fraction score for each bulk and single cell sample is plotted, such that a score greater than 0 may be interpreted as higher alternative splicing and a score less than 0 as lower alternative splicing. Each hollow dot represents a bulk RNA-seq sample and each filled dot represents a single cell. Blue dots represent samples or cells known to be *SF3B1* wild-type, red dots represent samples or cells known to be *SF3B1* mutant, and purple dots represent unknown *SF3B1* mutation status. Box plots visualize the distribution of scores with the red bar corresponding to the median. Median is represented as a line inside the box. Lines at the bottom and top of the box represent, respectively, the 25th and the 75th quartile.

Table S7, related to Figure 5. Supplied as an Excel file. Gene set enrichment analysis of alternatively spliced and differentially expressed genes in CLL samples with *SF3B1* mutation.

Table S8, related to Figure 5. Supplied as an Excel file. Gene expression associated with *SF3B1* mutation in the single cells from CLL samples.

Table S9, related to Figure 5. Supplied as an Excel file. Primer sequences for detection of gene expression associated with *SF3B1* mutation in the single CLL cell.

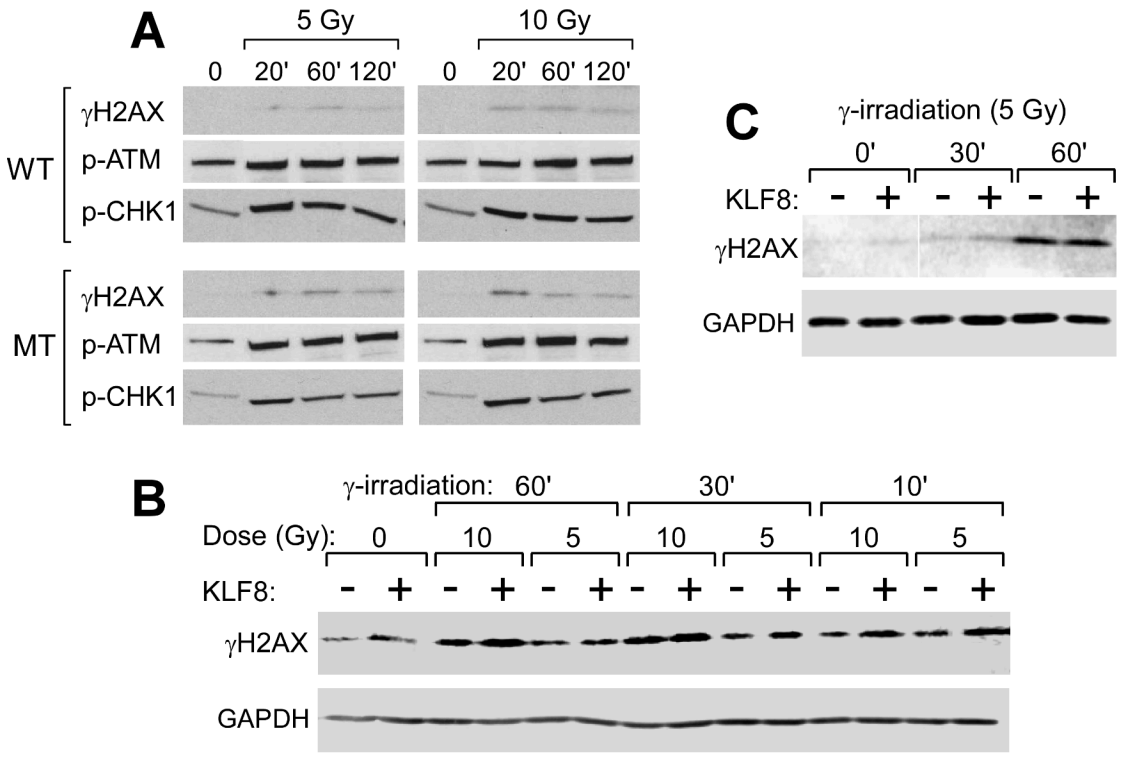

Figure S5, related to Figure 6. Effects of *SF3B1* mutation DNA damage pathway.

(A) Protein expression of phosphorylated H2AX, ATM and CHK1 in HeLa cells overexpressing either wild-type or mutant SF3B1 at different times after variable dosage of γ-radiation.
(B, C) Protein expression of phosphorylated H2AX was assessed in HEK293 (B) or Nalm-6 SF3B1^{K700K} (C) that overexpress control or KLF8 constructs upon γ-radiation using immunoblot.

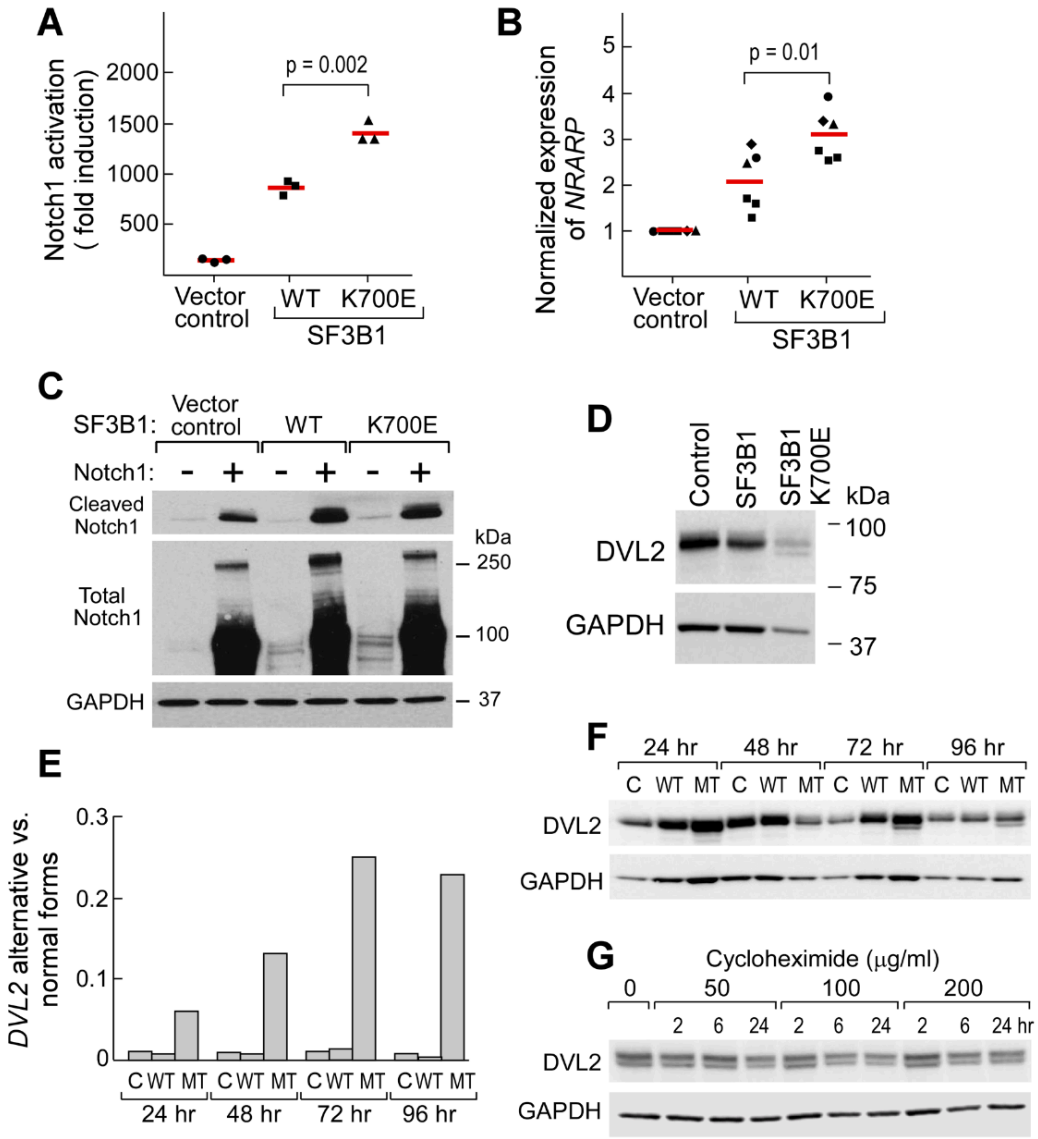

Figure S6, related to Figures 7. *SF3B1* mutation impacts Notch signaling and *DVL2* splicing.

(A-C) Notch activity was assessed in K562 cells that were nucleofected with control, wild-type or mutant SF3B1 in the presence or absence of a Notch 1-expressing construct. Notch activity was measured either with luciferase activity (A) or target gene expression by real-time PCR (B). Expression of activated and total Notch1 protein was also assessed from these cells (C).
(D) Expression of DVL2 was also assessed from K562 cells overexpressing control, wild-type or mutant SF3B1 plasmids.
(E, F) DVL2 alternative form was examined at the RNA and protein level using quantitative Taqman assay (E) and immunoblotting (F), respectively. HEK293 cells were transfected with control, wild-type and mutant SF3B1. RNA and protein lysate were generated in a time-course starting from 24 to 96 hours after transfection.
(G) Expression of DVL2 was examined from SF3B1^{K700E} cells that were treated with different dosages of cycloheximide for 2 to 24 hours.

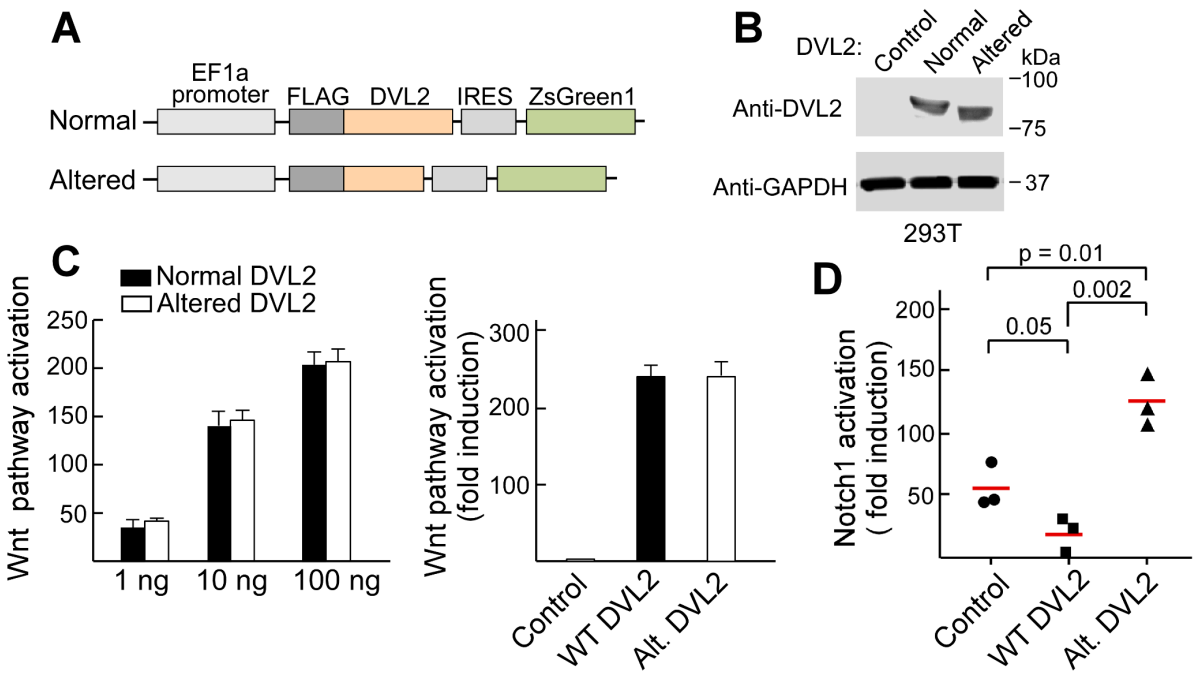

Figure S7, related to Figure 8. *SF3B1* mutation impacts Notch signaling through splicing of *DVL2*.

(A) Schematic map of constructs of normal and alternative forms of DVL2 used in the study.
(B) Detection of normal and alternative forms of DVL2 in HEK293T cells that overexpress the control, normal and altered DVL2 constructs using monoclonal antibody that recognizes N-terminal DVL2.
(C) Wnt pathway activation was examined in HEK293T (left) and K562 cells (right) that were transfected or nucleofected with wild-type and alternative form of DVL2 along with a Wnt pathway reporter. Error bars denote mean±SD, n=3.
(D) Notch pathway activation was assessed in K562 cells that were nucleofected with control vector, or plasmids expressing wild-type or alternative DVL2 along with a Notch reporter in the presence or absence of Notch1-expressing plasmid. Luciferase activity was examined and fold-induction with Notch1 was plotted. Red line indicates mean.

Supplemental Experimental Procedures

Sample processing

Peripheral blood mononuclear cells (PBMC) from normal donors and patients were isolated by Ficoll/Hypaque density gradient centrifugation. $CD19^+$ B cells from normal volunteers were isolated by immunomagnetic selection (Miltenyi Biotec, Auburn CA). Mononuclear cells were used fresh or cryopreserved with FBS 10% DMSO and stored in vapor-phase liquid nitrogen until the time of analysis.

Generation of expression constructs and stable cell lines

Full-length *SF3B1* was cloned using a partial fragment of SF3B1 cDNA (gift from Dr. Robin Reed, Harvard Medical School), and ligating this to a codon-optimized synthetic fragment encoding the uncloned region of 414 nucleotides (Blue Heron Biotechnology, Bothell, WA). A FLAG tag was added at the N-terminal of codon-optimized full-length SF3B1, and the FLAG-tagged SF3B1 cDNA was cloned into the expression vector pLVX-EF1α-IRES-ZsGreen (Clontech, Mountain View, CA) using the endonuclease cut sites of SpeI and BamHI. The K700E mutation was introduced into the codon-optimized full-length SF3B1 cDNA through site-directed mutagenesis according to manufacturer's instructions (QuickChange Site-Directed Mutagenesis Kit, Agilent Technologies, Lexington, MA). The constructs were transformed into Stbl2 competent bacterium (Life Technologies, Thermo Fisher Scientific Inc, Grand Island, NY) and cultured at 30ºC for plasmid DNA extraction. Wild-type DVL2 (clone from OriGene Technologies, Rockville, MD) and its altered splice isoform (synthesized by Life Technologies, Grand Island, NY) were likewise subcloned into the expression vector pLVX-EF1α-IRES-ZsGreen. The full-length KLF8 expression construct, in the pCMV6-Entry vector, was purchased from OriGene (RC211273, Rockville, MD).

Stable OCI-Ly1 cells that express either wild-type, altered or wild-type+altered DVL2 were generated through spin infection with lentivirus. To produce the virus, 293T cells were plated on 6-well plates at 3×10^5 cells per well in complete DMEM medium with 10% serum and allowed to adhere for 16 hrs. Transfections were performed as per manufacture's instructions with 3 µg of wild-type or altered DVL2 cloned in pLVX-EF1α-IRES-ZsGreen, 1.5 µg psPAX2, and 0.4ug VSV-G packaging plasmids along with 12 µl of Lipofectamine ™ 2000 in serum-free Opti-MEM media (Life Technologies, Thermo Fisher Scientific Inc, Grand Island, NY). 24 hours following transfection, fresh complete DMEM with 10 % FBS was added and the cells were incubated for an additional 24 hours for virus production. Virus supernatant was collected and filtered through a 0.45 micron Nalgene syringe filter SFCA (Whatman, Clifton, NJ), and concentrated by a size exclusion column (Vivaspin 20 column; (Satorius, Goettingen, Germany) at 2200rpm for 15 minutes at 4°C. 0.5 million OCI-Ly1 cells were infected with 100µl concentrated virus by spin infection (2200rpm, 90 minutes) at 37°C with a final concentration of polybrene at 8ug/ml. GFP-positive cells were sorted 5 days after the infection and the stable cell lines were maintained with more than 90% GFP positivity.

Immunofluorescence staining, immunoblotting, and immunoprecipitation

For immunofluorescence staining of transfected K562 cells, cells were cytospun at 700 rpm for 5 minutes. Cells were subsequently fixed in 4% paraformaldehyde in PBS and permeabilized with 100% methanol. Cells were stained with primary and secondary antisera as previously described (Bloch et al., 2005), using the following antibodies: murine anti-FLAG antibody (F1804, Sigma, St. Louis, MO); rabbit anti-Sf3b1 antibody (ab66774, Abcam, Cambridge, MA); and species-specific secondary antibodies, conjugated to FITC or rhodamine (Jackson ImmunoResearch Laboratories, West Grove, PA).

For immunoblotting, cells were harvested, washed once with phosphate-buffered saline (PBS), and lysed for 30 minutes at 4°C with RIPA buffer (Boston Bioproducts, Boston, MA) supplemented with a cocktail of protease inhibitors (1 tablet in 10 ml RIPA buffer, Roche, San Francisco, CA). Cell debris was removed by

centrifugation at 10,000×g for 5 minutes, and cell extracts were fractionated by SDS-PAGE and transferred to Immun-Blot PVDF membranes (Bio-Rad laboratories, Hercules, CA). Membranes were stained with primary antibodies and appropriate secondary antibodies, e.g., horseradish peroxidase-coupled goat anti-rabbit and goat anti-mouse antibodies (Rockland Immunochemicals Inc, Limerick, PA). The bands were visualized by using the ECL chemiluminescence Western blotting detection system (GE Healthcare, Boston, MA).

Antibodies used in the study include: murine anti-FLAG antibody (Sigma, St. Louis, MO), rabbit anti-β actin (ab8227) (Abcam, Cambridge, MA), rabbit anti-DVL2 (3224), anti-Notch1 Val1744 (activated form) (4147), anti-Notch1 (total) (3608), anti-ATM Ser1981 (5883), anti-CHK1 Ser317 (2344), and anti-GAPDH (2118) antibodies (Cell Signaling (Cell Signaling, Danvers, MA). For some experiments, cycloheximide (Cell Signaling, Danvers, MA) was added to cultures prior to the generation of cell lysates.

For immunoprecipitation experiments, nuclear extracts were prepared from HeLa cells 48 hours after transfection, as described previously (Das et al., 2006; Mattaj et al., 1986). Immunoprecipitation was performed with either U2 snRNP protein- or the murine anti-FLAG antibody, that were coupled to protein A (polyclonal) or G (monoclonal) and then covalently crosslinked using dimethylpimelimidate (Sigma Aldrich, S). For each sample, 40 µl of nuclear extract was incubated for 20 minutes at 30 ˚C in 0.5 mM ATP, 3.2 mM $MgCl_2$, 20 mM creatine phosphate (di-Tris salt), 2 mM PMSF, 2 ml protease inhibitor EDTA-free (Roche, Indianapolis, IN). After incubation, reaction mixtures were spun at 4 ˚C for 5 minutes at 14,000rpm. Supernatants were then added to the antibody-crosslinked beads and the immunoprecipitations were rotated for 2 hours at 4 ˚C and washed 5 times with 1 ml of wash buffer (PBS with 0.1% Triton-X). Immunoprecipitated protein complex was boiled in the SDS lysis buffer and resolved on NuPAGE 4-12% pre-cast gels (Life Technologies, Grand Island, NY), and the gel was transferred to nitrocellulose membranes. Western bloting was performed using rabbit anti-Sf3b1 antibody.

Detection of activity of SF3B1 on CLL cellular pathways

Wild-type and mutant SF3B1 were introduced into K562 (ATCC), HG3 (gift from Dr. Anders Rosen, Linkoping University, Sweden) and Jeko cells (ATCC) by nucleofection (Amaxa nucleofector, Lonza, Walkersville, MD; V buffer and program T016 for K562 and HG3; V buffer and program X-001 for Jeko). The constructs were introduced into HEK293T and Hela cells through transfection using Lipofectamine 2000 (Life Technologies, Thermo Fisher Scientific Inc, Grand Island, NY) per manufacturer's recommendations. Alternatively, we tested pre-B Nalm-6 isogenic cell lines, in which $SF3B1^{K700K}$, $SF3B1^{K700E}$ or $SF3B1^{H622Q}$ were introduced by AAV-mediated homology. To control for the expression of mutant vs wild-type alleles in suspension cell lines, we sorted GFP-expressing cells before performing functional analyses. Pre-sorting of cells based on GFP-expression was not performed in transfected HEK293T and HeLa cells, since the transfection efficiencies in these settings were consistently high (>70%).

Pathway activation was screened using the following methods: (1) Plasmid-based luciferase reporter assays (SuperTOP lucierase reporter and Wnt1 expressing plasmid for Wnt signaling (Wang et al., 2014); TP-1 luciferase reporter Notch 1 expressing plasmid (Aster et al., 2000); (2) Gene expression assays (RNA splicing and Notch target gene); (3) Western blot (DNA damage, Notch signaling); (4) Flow cytometry based assays (cell cycle, apoptosis).

DNA damage response was assessed in HeLa, HEK293 and Nalm-6 cells. First, HeLa and HEK293 cells were plated in 6-well plates 24 hours before transfection with SF3B1 (3µg) or KLF8 (3µg, RC211273, OriGene, Rockville, MD) constructs using Lipofectamine 2000 (Invitrogen) according to manufacturer's protocol. Next, 48 hours after overexpression, cells were irradiated with different dosage of γ-radiation.

Nalm-6 SF3B1^{K700K} cells were nucleofected either with control or KLF8 constructs (10μg) (Amaxa, C-005, V buffer, Lonza, Walkersville, MD) along with GFP expressing plasmids; GFP positive cells were sorted 48 hours after nucleofection. For studies of DNA damage response, cells were treated with γ-radiation and cells lysates were generated using the methods described above.

For measuring Notch activity in K562 cells, 3 million cells were nucleofected with SF3B1 (10μg), TP-1 luciferase reporter (2μg) and pRL-TK (80ng) plasmids, with or without 5μg Notch1 expression plasmid (Amaxa, U-017, V buffer, Lonza, Walkersville, MD). In the case of DVL2, 2μg was used in the same setting. Flow-sorted GFP cells expressing the constructs were isolated and lysed prior to interrogation in downstream assays (ie. for testing for luciferase activity using the Dual-Luciferase Reporter Assay (Promega, Madison, MI), for RNA gene expression, or for immunoblotting). Notch activities in the Nalm-6 isogenic cell lines were similarly measured.

Notch signaling in the stable OCI-Ly1 cell lines was activated through co-culturing with OP9-DL1 cells (expressing Notch ligand Delta 1). OP9 and OP9-DL1 stromal cells were cultured in Minimum Essential Medium Alpha (Invitrogen), supplemented with 20% fetal bovine serum, 100 U/mL penicillin, 100 μg/mL streptomycin, and 1uM L-glutamine (all from Invitrogen), and plated 24 to 48 hours before use in 6-well plates to achieve a confluent monolayer of cells. 2.5 million OCI-Ly1 cells were co-cultured with either OP9 or OP9-DL1 cells for 48 hours before assessing for Notch pathway activation. Notch target gene expression and NOTCH1 protein expression were assessed from the co-cultured OCI-Ly1 cells using qRT-PCR and immunoblot, respectively.

Detection of RNA splice variants in CLL samples and cell lines

Total RNA was extracted from normal B, CLL-B, Nalm-6, transfected HEK293T, U2OS cells, or nucleofected K562 cells (TRIZOL; Invitrogen, Carlsbad CA). 2 μg of total RNA per sample were treated with DNase I (2 units/sample; New England BioLabs, Ipswich MA) at 37°C for 20 minutes to remove contaminating genomic DNA, followed by heat-inactivation of DNase I at 75°C for 15 minutes, and then used as template to synthesize cDNA by reverse transcription (SuperScript® III First-Strand kit; Invitrogen, Carlsbad CA). We designed quantitative Taqman assays to detect the normal and alternative spliced transcripts in which probes were localized within the splicing junctions. Gene expression assays for the Notch pathway target *NRARP* (Life technologies, Grand Island, NY) were assayed. All assays were run in duplicate using the 7500 Fast System (Applied Biosystems, Carlsbad CA), and all values were normalized to *GAPDH* gene expression. Relative splicing activity was measured by calculating the ratio of alternative to normal spliced forms of each target gene.

Bulk and single cell RNA-Seq library generation and data processing

Total RNA was extracted from primary CLL cells, normal B cells or transfected cell lines (RNeasy Mini Kit, Qiagen, Valencia, CA), and assessed for quality and quantity (RNA 6000 Nano LabChip kit on a 2100 Bioanalyzer, both from Agilent). Poly-A selected RNA-Seq libraries were prepared according to standard Illumina protocols. cDNA libraries, except as noted below, were checked for quality and quantified using the DNA-1000 kit (Agilent, Santa Clara, CA) on a 2100 Bioanalyzer. Each library was sequenced with the Illumina Sequencing Kit v4 on one lane of a Gene Analyzer IIX sequencer to obtain 76-base paired-end reads. For six samples, poly-A selected and total (RNase H rRNA)-depleted Illumina libraries were prepared from two aliquots of 1 microgram of total RNA with the following exceptions. RNA was fragmented using 10x RNA fragmentation buffer (New England Biolabs, Ipswich, MA) for 4 min at 85 degrees C. SUPERase-In (Life Technologies, Grand Island, NY) was added during the Turbo DNase step to minimize RNA degradation. After rRNA depletion, we obtained 39 to 81 nanograms of RNA. After poly-A selection, we obtained 5 to 10 ng of RNA. We used 7 and 9 PCR cycles to generate the RNase H and poly-A libraries, respectively. Libraries were pooled and sequenced on a HiSeq2000 (Illumina) to obtain 101-base paired-end reads.

Single-cell CLL RNA libraries were generated from viable $CD19^{+}CD5^{+}$ tumor cells that were pre-sorted by flow cytometry. Subsequently, the bulk cell concentration was adjusted to 250 cell/ml and applied to the C1 Single-Cell Auto Prep System for single cell capture with a 5-10 micron chip (Fluidigm, South San Francisco, CA). The capture rate was measured at > 80%. Following capture, whole transcriptome amplification (WTA) was immediately performed using the C1 Single-Cell Auto Prep System with the SMARTer Kit (Clontech, Mountain View, CA) on up to 96 individual cells. The C1 WTA products were then converted to Illumina sequencing libraries using Nextera XT (Illumina, San Diego, CA). RNA-Seq was performed on a MiSeq instrument (Illumina, San Diego, CA). Reads were aligned to the human reference genome (hg19) using TopHat v1.4 with Gencode v12 gene annotations.

Quality control was performed using Picard (http://broadinstitute.github.io/picard/index.html). Lowly expressed genes (detected in fewer than 4 cells) were removed prior to SCDE model fitting. Poor cells or empty wells (estimated library size < $1x10^6$, mode mapping quality = 0, passed filter aligned bases < $1.5x10^8$) were also removed from prior to model fitting, resulting in 255 out of 384 single cells.

Analysis of alternative splicing of bulk and single cell RNA-Seq data

To identify splicing changes significantly associated with *SF3B1* mutation, we used JuncBASE. JuncBASE identifies and quantifies alternative splicing events and performs differential splicing analysis. To control for potential batch effects, we implemented a permutation-based method to find events that were differentially spliced between *SF3B1* wild-type and mutant CLL samples. This approach performs a Wilcoxon rank sum test between PSI values of *SF3B1* mutated versus unmutated samples and calculates an empirical p value based on the permuted phenotype labels within the same sequencing batch. As PSI values are derived from abundance estimates of inclusion and skipping of each alternative splicing event, permutations were also performed on these abundance estimates given a beta-binomial distribution. We selected splicing events with a Benjamini-Hochberg false discovery rate of 5% and a difference in PSI > 10% between the median values of wild-type and mutant *SF3B1* samples as our most confident set of splicing changes. For differential splicing analysis in K562 cells and in total (RNaseH) libraries, a t-test between the observed PSI values of *SF3B1* WT versus mutant samples was used.

We focused on alternative 3' splicing events observed in both bulk and single cells, using the inclusion junction as the unique identifier to map between single cell and bulk alternative splice variants. We computed the PSI for each splice variant in each of the bulk and single cell samples. We used the bulk *SF3B1* WT samples to gauge whether the inclusion or exclusion isoform is the 'normal' isoform: if the average PSI among bulk *SF3B1* wild-type samples was higher than the average PSI among bulk *SF3B1* mutant samples, the inclusion isoform would be deemed normal and the exclusion isoform deemed alternative. In this manner, we calculated the alternative splice fraction for each splice event observed bulk and single cell sample. We further computed a weighted average alternative splice fraction for each bulk and single cell sample, weighting each alternative splice fraction by the ΔPSI from bulk *SF3B1* mutant versus wild-type samples. We centered this weighted average alternative splice fraction score using the average of bulk *SF3B1* mutant and wild-type samples such that a score greater than 0 may be interpreted as higher alternative splicing and a score less than 0 as lower alternative splicing.

Variant calling from single cell RNA-Seq data was performed using the Genome Analysis Toolkit (v3.0.0) (McKenna et al., 2010) allowing for 2 mismatches, local realignment around indels using the Mills and 100G gold standard, filtering on dbsnp release (build 138), with VQSR and otherwise default settings. Manual inspection using the Integrative Genomics Viewer (IGV) was also used to confirm the absence and presence of the expected nucleotide changes in *SF3B1* for each single cell.

Density plots of the relative positions of alternative splice sites and motif frequency for canonical as well as alternative splice sites were performed. For simplicity, events used for detecting the distance between alternative 3' splices sites and sequence motifs were performed using only junction reads and only from current 22 DFCI samples. Similar results were found when using the full cohort of 37 samples (22 DFCI samples + 15 ICGC samples (data not shown).

To check for expression of the altered *DVL2* isoform across normal tissues, 693 RNA-Seq BAM files were downloaded from the GTEx consortium from a variety of tissues including blood, brain, breast, lung, colon. We performed JuncBASE analysis to quantify PSI values of *DVL2* alternative splicing event across these 693 samples. From this data, there was no expression of the altered *DVL2* isoform above a 10% relative inclusion (PSI). We note that CLL RNA-Seq reads were aligned using the same aligner method (TopHat v1.4) with the same parameters as reads aligned in GTEx.

Analysis of differential gene expression of bulk and single cell RNA-Seq data

Differential gene-expression analysis on bulk samples, both poly-A and RNAse-H derived, was performed using the DESeq2 R package with *SF3B1* mutation status as the model matrix and condition batch labels as the covariate with standard settings. We applied a lenient threshold of 0.2 on the p value adjusted for multiple testing using the Benjamini-Hochberg method to identify 1963 out of 28456 statistically significant differentially expressed genes between 24 *SF3B1* wild-type and 13 mutant samples.

Differential gene-expression analysis on single cell samples was performed using the SCDE R package with standard settings in order to better account for drop-outs and other technical artifacts more common to single cell RNA-Seq data. A p value threshold (non-corrected) of 0.05 was applied to identify 326 out of 63677 differentially expressed genes between 52 *SF3B1* WT and 12 mutant cells identified using the variant calling method described above.

For gene-set-enrichment analysis, we focused on the C2 curated gene sets and C6 oncogenic signatures gene set from MSigDB along with 14 hand-curated gene sets pertaining to DNA damage and cell cycle, apoptosis, and Notch signaling (**Table S7**). To test whether particular gene sets and pathways were enriched in our set of differentially expressed genes, we used a custom implementation of the GSEA algorithm (Subramanian et al., 2005) on the significantly differential expressed genes ranked by fold change. Similarly, to test for pathway association of genes exhibiting alternative splicing, the significantly affected splice variants ranked by absolute ΔPSI. We applied a 10% FDR threshold, empirically estimated from gene rank randomization to identify significantly enriched gene sets. To enhance interpretability of results, we focused on gene sets associated with positive fold change and positive enrichment scores (significantly upregulated) and gene sets associated with negative fold change and negative enrichment scores (significantly downregulated).

Single cell targeted mutation, gene expression, and alternative splicing analysis

Cryopreserved peripheral blood normal $CD19^+$ and CLL-B cells samples were thawed and stained with anti-CD19 FITC and anti-CD5 PE antibodies (Beckman Coulter, Indianapolis, IN) and 7-AAD (Invitrogen, Grand Island, NY). Live single $CD19^+CD5^+$ tumor cells were directly flow sorted into wells of 96-well plates with 5µl of a lysis buffer consisting of 10 mM Tris-HCl, pH 8.0; 0.1 mM EDTA; 0.5% NP-40 (Thermo Scientific, Waltham, MA) and 0.1U/µl SUPERase-In (Life Technologies). Immediately following cell sorting, each plate of cells was gently vortexed, spun down, and flash frozen on dry ice. Integrated detection of somatic mutations, splice variants and gene expression in the same single CLL and B cells was performed on a Fluidigm Biomark instrument (Fluidigm Corporation, South San Francisco CA) using an adaptation of a previously reported protocol (Livak et al., 2013). Targeted detection assays were designed using a nested design, with outer primers for preamplification and inner primers for qPCR detection. For splice variant detection, separate splice-specific primer pairs were designed for the normal splice form and

the variant splice form. For mutation detection, separate assays were designed to cDNA sequence for the normal allele and the mutation allele. The primer sequences for all of the RNA expression, splice variant, and mutation assays are listed in **Table S4 and S9**. For each patient sample, a 10× Preamplification Primer Mix (500 nM each primer) was prepared containing all the outer primers for all the RNA expression, splice variant, and mutations assays to be used for that patient. As previously described, single cell cDNA was synthesized by adding 1 μL Reverse Transcription Master Mix (Fluidigm 100-6299), preamplified using a mix of 2 μL 5× PreAmp Master Mix (Fluidigm 100-5744), 1 μL 10× Preamplification Primer Mix, and 1 μL H_2O, and then processed for qPCR detection. qPCR detection was performed using 96.96 Dynamic Array™ IFCs and the Biomark™ HD System from Fluidigm, per manufacturer's recommendations.

Single cells were deemed 'mutant' if fractional mutant allele expression was above the expected background level; or 'wild-type' if fractional mutant allele expression was below the expected background. We did not differentiate between homozygous and heterozygous mutants. Expected background level of mutant allele detection was determined by linear regression modeling against wild-type allele detection in negative controls (i.e. normal B cells, known to not express the mutated allele). We restricted analyses to cells for which we could confidently call 'wild-type' or 'mutant' status. For analyses of alternative splicing and gene expression in single cells, cells were first binned into classes by mutation status (mutant vs wild-type). For RNA expression, the distribution of expression levels was directly compared between the two classes. Cells for which less than 10 of 96 assayed genes were detected were deemed as failed cells and omitted. For alternative splicing, the distribution of fractional expression of the alternatively spliced allele, defined as alternatively spliced allele over total, was compared between the two classes. A two-sided Wilcoxon rank-sum test is used to assess statistical significance. A non-multiple testing corrected *p value* threshold of 0.05 was used to determine significance.

Branch point identification in RNAse H RNA-Seq data

To identify branch points in RNase-H RNA-Seq data, we applied Mercer's method which identified lariat reads. Lariat reads are reads in which one portion of the read mapped in an intron directly downstream of the 5' splice site and a second portion of the read mapped in an intron near a 3' splice site. Additionally, these portions of the read align in an inverted order. Subsequently, we applied a series of filters to the putative lariat reads. We filtered out lariat reads that did not support lariats from an intron of a single gene. For all lariat reads, we only considered lariats for which half of the alignment started in an intron within 2bp of an annotated 5' splice site. Some identified lariat reads had exact alignments in which all bases of the read aligned and the position of the branchpoint could be determined exactly. Additionally, some lariats reads contained overlapping sequence in the middle, and some contained a gap in the middle. For these types of lariat reads, we filtered out gapped lariat reads with below 95 bases aligned and we filtered out overlapping lariat reads with above 105 bases aligned. Finally, we considered lariat reads with branchpoint positions in introns with alternative 3' splice site events with $p<0.05$ that only had evidence for two splice site choices. We considered alternative 3' splice site events where either a proximal cryptic splice site was preferentially spliced in or a distal cryptic splice site was preferentially spliced in. Additionally, we considered lariat reads in a set of control introns with alternative 3' splice site events that had $q>0.9$. For all events, we considered only lariat reads supporting branchpoints at intronic positions within 100bp from the 3' splice site.

Detection of telomerase activity

Telomerase activity in cell extracts was measured using the TRAPeze Telomerase Detection Kit (S7700, Millipore, Sigma, Billerica, MA). Each assay mixture included 5 μl of 5× TRAPeze Reaction Mix, 2ul γ–^{32}P–ATP, 15.6 μl PCR Grade Water, 0.4 μl of 50× TITANIUM Taq DNA Polymerase (Clontech), and 2 μl cell extract from 10,000 Nalm-6 cells with or without *SF3B1* mutations. Samples were subjected to the following cycling parameters using an Applied Biosystems 7300 Real-Time PCR system: 30 min at 30 °C (extension of telomerase substrate); 2 min at 95 °C then 45 cycles of 15 s at 94 °C, 1 min at 59 °C and 30 s

at 45 °C (PCR amplification of extended telomerase substrate). Internal PCR reaction control was included to indicate the amplification efficiency from each reaction. PCR product was subjected to 12.5% PAGE. After gel was dried, autoradiography was developed.

Taqman assay primers

All primers were purchased from Life technologies (Grand Island, NY).

Assay Name	Assay ID	Forward Primer Seq (5'-> 3')	Probe Sequence (5'-> 3')
ZNF91 altered	AIKALGQ	CCCTGGAATATGAAGCAACATGAGA GGCCAAAAGTCTTGAGGAAAATGAG	ATGAACCCACAGTTTTTT
ZNF91 normal	AILJJMY	CCCTGGAATATGAAGCAACATGAGA GGCCAAAAGTCTTGAGGAAAATGAG	AACCCACAGGTATATGTC
MAPK7 altered	AIPAD5M	TCTTGTGATGGAATATGCTGAAGGG GCAGCAGTATAATATGGCAATGGTT	AAAGAAAGGCACAAACATT
MAPK7 normal	AIPAD5M	AGTGTGTCTTGTGATGGAATATGCT GCAGCAGTATAATATGGCAATGGTT	CCATGCAGCACATTAT
GCC2 altered	AIVI40Q	GAGCTGGAGGCAAGCCA TGATGTTATTTCAGCCAGCTGTATCTG	CAAGTAGAAGTCTATAAAATTTAC
GCC2 normal	AIT96UI	GCCAGCAGCAAGTAGAAGTCTA GTGGATTTTGTGCTTCTCTGATGTT	CAGCCAGCTGTATTTTA
TPP2 altered	AIWR26Y	CTCCCTGCATAACTTTGAAGAACTG CCTAAAGGTTTTGTTTTTGCACTCACT	CACTGCGTTTCCTCATTAC
TPP2 normal	AIX01C6	CTCCCTGCATAACTTTGAAGAACTG GGCAAAACATCTCTTGATCCTAAAGGT	CTGCGCCCAGTGAGTG
DVL altered	AIT97YM	ATTGTGCTGACTGTGGCCAA GGCCTTCACAGCCATCAGGC	TCCCCCGAACCTATCCAGGTTC
DVL normal	AIS09SE	ATTGTGCTGACTGTGGCCAA GTAATGGTGCTCATGGAGGA	TCCCCCGAATGAGCCCAT
TERT	Hs00972650_m1		
TERC	Hs03297287_s1		
NRARP	Hs00171172_m1		
HES1	Hs00172878_m1		

Supplemental References

Aster, J. C., Xu, L., Karnell, F. G., Patriub, V., Pui, J. C., and Pear, W. S. (2000). Essential roles for ankyrin repeat and transactivation domains in induction of T-cell leukemia by notch1. Mol Cell Biol *20*, 7505-7515.

Bloch, D. B., Yu, J. H., Yang, W. H., Graeme-Cook, F., Lindor, K. D., Viswanathan, A., Bloch, K. D., and Nakajima, A. (2005). The cytoplasmic dot staining pattern is detected in a subgroup of patients with primary biliary cirrhosis. J Rheumatol *32*, 477-483.

Das, R., Dufu, K., Romney, B., Feldt, M., Elenko, M., and Reed, R. (2006). Functional coupling of RNAP II transcription to spliceosome assembly. Genes Dev *20*, 1100-1109.

Livak, K. J., Wills, Q. F., Tipping, A. J., Datta, K., Mittal, R., Goldson, A. J., Sexton, D. W., and Holmes, C. C. (2013). Methods for qPCR gene expression profiling applied to 1440 lymphoblastoid single cells. Methods *59*, 71-79.

Mattaj, I. W., Habets, W. J., and van Venrooij, W. J. (1986). Monospecific antibodies reveal details of U2 snRNP structure and interaction between U1 and U2 snRNPs. EMBO J *5*, 997-1002.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res *20*, 1297-1303.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.

Wang, L., Shalek, A. K., Lawrence, M., Ding, R., Gaublomme, J. T., Pochet, N., Stojanov, P., Sougnez, C., Shukla, S. A., Stevenson, K. E.*, et al.* (2014). Somatic mutation as a mechanism of Wnt/beta-catenin pathway activation in CLL. Blood *124*, 1089-1098.