Cancer Cell

Transcriptomic Characterization of *SF3B1* Mutation Reveals Its Pleiotropic Effects in Chronic Lymphocytic Leukemia

Highlights

- *SF3B1* mutation causes alternative splicing in cell lines and primary CLL cells

- *SF3B1* mutation-associated splice variants are enriched for 3′ splice sites

- *SF3B1* mutation induces RNA changes affecting multiple CLL-associated pathways

- *SF3B1* mutation modulates Notch signaling through an RNA splice variant of *DVL2*

Authors

Lili Wang, Angela N. Brooks, Jean Fan, ..., Matthew M. Meyerson, Peter V. Kharchenko, Catherine J. Wu

Correspondence

cwu@partners.org

In Brief

Wang et al. perform transcriptomic characterization of bulk or single primary human chronic lymphocytic leukemia cells harboring *SF3B1* mutations, identifying several dysregulated cancer-related pathways resulting from altered expression of *KLF8*, *TERC*, or *TERT* or altered splicing of *DVL2* mRNA.

CrossMark

CellPress

Transcriptomic Characterization of *SF3B1* Mutation Reveals Its Pleiotropic Effects in Chronic Lymphocytic Leukemia

Lili Wang,[1,11,15] Angela N. Brooks,[1,2,3,15] Jean Fan,[4,15] Youzhong Wan,[1,13,15] Rutendo Gambe,[1] Shuqiang Li,[7] Sarah Hergert,[1] Shanye Yin,[12] Samuel S. Freeman,[2] Joshua Z. Levin,[2] Lin Fan,[2] Michael Seiler,[14] Silvia Buonamici,[14] Peter G. Smith,[14] Kevin F. Chau,[11] Carrie L. Cibulskis,[2] Wandi Zhang,[1] Laura Z. Rassenti,[6] Emanuela M. Ghia,[6] Thomas J. Kipps,[6] Stacey Fernandes,[1] Donald B. Bloch,[10] Dylan Kotliar,[11] Dan A. Landau,[1,11] Sachet A. Shukla,[1] Jon C. Aster,[8] Robin Reed,[12] David S. DeLuca,[2] Jennifer R. Brown,[1,9] Donna Neuberg,[5] Gad Getz,[2] Kenneth J. Livak,[7] Matthew M. Meyerson,[1] Peter V. Kharchenko,[4] and Catherine J. Wu[1,2,9,11,16,*]

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, Dana 540, 44 Binney Street, Boston, MA 02115, USA
[2]Broad Institute of MIT and Harvard, Cambridge, MA 02141, USA
[3]University of California, Santa Cruz, CA 95064, USA
[4]Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA
[5]Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02115, USA
[6]Moores Cancer Center, University of California, San Diego, La Jolla, CA 92093, USA
[7]Fluidigm Corporation, South San Francisco, CA 94080, USA
[8]Department of Pathology
[9]Department of Medicine
Brigham and Women's Hospital, Boston, MA 02115, USA
[10]Center for Immunology and Inflammatory Disease, Massachusetts General Hospital, Boston, MA 02115, USA
[11]Harvard Medical School, Boston, MA 02115, USA
[12]Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA
[13]National Engineering Laboratory of AIDS Vaccine, School of Life Science, Jilin University, Changchun, Jilin, PRC
[14]H3 Biomedicine, Cambridge, MA 02141, USA
[15]Co-first author
[16]Lead Contact
*Correspondence: cwu@partners.org
http://dx.doi.org/10.1016/j.ccell.2016.10.005

SUMMARY

Mutations in *SF3B1*, which encodes a spliceosome component, are associated with poor outcome in chronic lymphocytic leukemia (CLL), but how these contribute to CLL progression remains poorly understood. We undertook a transcriptomic characterization of primary human CLL cells to identify transcripts and pathways affected by *SF3B1* mutation. Splicing alterations, identified in the analysis of bulk cells, were confirmed in single *SF3B1*-mutated CLL cells and also found in cell lines ectopically expressing mutant SF3B1. *SF3B1* mutation was found to dysregulate multiple cellular functions including DNA damage response, telomere maintenance, and Notch signaling (mediated through *KLF8* upregulation, increased *TERC* and *TERT* expression, or altered splicing of *DVL2* transcript, respectively). *SF3B1* mutation leads to diverse changes in CLL-related pathways.

Significance

SF3B1 encodes an RNA splicing factor and is among the most frequently mutated genes in CLL; however, mechanistic insights into its role in the oncogenic process are lacking. We report a comprehensive transcriptomic analysis of CLL bulk and single cells, coupled with functional examination, to investigate the impact of mutated *SF3B1* on CLL-associated pathways. Our analyses suggest that *SF3B1* mutation induces subtle but broad changes in gene expression and splicing across multiple pathways including DNA damage, Notch signaling, and telomere maintenance. We found that *SF3B1* mutation modulates DNA damage response through *KLF8* upregulation and increases Notch signaling through altered splicing of *DVL2*. Our studies point to diverse mechanisms by which RNA splicing factors can affect cancer phenotype.

INTRODUCTION

Large-scale cancer-sequencing efforts have enabled the discovery of paths to carcinogenesis. In chronic lymphocytic leukemia (CLL), these initiatives unexpectedly led to the identification of highly recurrent mutations in *SF3B1*, which encodes a key component of the spliceosome, at restricted sites (50% at K700E), implicating a role of altered RNA splicing in CLL (Quesada et al., 2012; Rossi et al., 2011; Wang et al., 2011). Mutated *SF3B1* has been associated with adverse clinical outcome in CLL and has a higher incidence in refractory CLL (Jeromin et al., 2014; Landau et al., 2015). However, mechanistic insights into downstream paths affected by mutated *SF3B1* and its role in the oncogenic phenotype are still unclear.

SF3B1 is an essential component of the U2 small nuclear ribonucleoprotein particle (snRNP), which interacts with branchpoint sequences close to 3′ splice sites during pre-mRNA splicing (Chen and Manley, 2009). The critical function of SF3B1 in pre-mRNA splicing leads to the hypothesis that SF3B1 mutations contribute to CLL through the generation of alternatively spliced transcripts. A variety of previous studies have identified splicing alterations associated with mutated SF3B1 in CLL (Alsafadi et al., 2016; Darman et al., 2015; DeBoever et al., 2015; Ferreira et al., 2014; Kesarwani et al., 2016), but the breadth of its functional impact on CLL biology has remained elusive. The study of SF3B1 function has been complicated by difficulties in the genetic manipulation of human B cells and the complex biology associated with altering an essential component of the splicing machinery.

In the current study, we set out to examine the functional impacts of *SF3B1* mutations by overcoming these challenges.

RESULTS

Mis-Splicing in CLL Samples with *SF3B1* Mutations Is Enriched for Alternative 3′ Splice Sites

Given the key role of SF3B1 in pre-mRNA splicing, we hypothesized that features of altered splicing associated with this recurrently mutated gene could provide mechanistic insights into the functional impact of this putative CLL driver. We therefore performed RNA sequencing (RNA-seq) from poly(A)-selected RNA of 22 CLL samples and combined these results with a published set of 15 CLL RNA-seq data (Ferreira et al., 2014) to yield a total of 13 and 24 cases with and without the *SF3B1* mutation, respectively. Thirteen of 37 cases (4 of 10 *SF3B1*-mutated cases with known *IGHV* status) had unmutated *IGHV*. Of the nine samples with cancer cell fraction data available, five had clonal and four had near-clonal (cancer cell fractions of 60%–90%) *SF3B1* mutations (Table S1).

To identify and classify altered splicing events associated with *SF3B1* mutation, we applied the tool JuncBASE (Brooks et al., 2011). We also used JuncBASE to detect unannotated alternative splicing and calculate a "percent spliced in" (PSI) value for each individual splicing event to quantify the inclusion of an alternative exon relative to the total abundance of all isoforms. Unsupervised hierarchical clustering of the samples based on the top 25% most variable splicing events among the 37 CLL cases revealed clustering of CLL cases with *SF3B1* mutations, separate from *SF3B1* unmutated samples; however, batch effects were observed (Figure S1A).

To account for these batch effects, we implemented a permutation-based approach in the JuncBASE package to identify robustly altered splicing events associated with *SF3B1* mutated samples (Experimental Procedures). We found pervasive changes in 3′ splice site selection as observed by a large skew toward lower p values in a Q-Q plot (Figure 1A). To a lesser extent, *SF3B1* mutations also were associated with changes in other types of alternative splicing (e.g., alternative 5′ splice sites, cassette exons) (Figure S1B). Although significant splicing changes (p < 0.05) were consistent among *SF3B1*-mutated CLLs, the majority of changes were subtle, such that 65% showed a <10% difference in the median PSI value (ΔPSI) of *SF3B1* wild-type (WT) and mutated samples (Figure S1C, Table S2). When randomly sampling 13 versus 24 cases, 92% of ΔPSI values were <10%, supporting a difference in PSI of > 10% as an appropriate cutoff to identify alterations with stronger effects (Figure 1B).

Because altered splicing may generate splice variants that could in turn affect cellular circuitry, we focused on splicing changes most strongly associated with *SF3B1* mutation (false discovery rate < 10% and ΔPSI > 10%), resulting in 304 splicing changes (Table S2). As expected, there was a significant difference in the types of alternative splicing events strongly associated with *SF3B1* mutation compared with highly variable splicing events among samples without *SF3B1* mutation (p < 0.0001, Figure 1C). CLL samples without *SF3B1* mutation had PSI values for these splicing events similar to normal B cells from seven healthy donors, further indicating that these splicing alterations are specific to *SF3B1* mutation and not a general feature of CLL (Figures 1D and S1D). We observed a bias in the distance between the canonical to alternative 3′ splice sites associated with *SF3B1* mutation. In addition, branchpoints used in *SF3B1* WT conditions (Mercer et al., 2015) were found to map at or <10 nt from the aberrant 3′ splice site (Figure 1E), and we observed A's enriched upstream of the aberrant 3′ splice site (Figure S1E), suggesting altered branchpoint usage in the presence of *SF3B1* mutation, as recently described (Alsafadi et al., 2016; Darman et al., 2015; DeBoever et al., 2015). Of four randomly selected candidate *SF3B1* mutation-associated splice variants (*GCC2*, *MAP3K7*, *TPP2*, *ZNF91*), all were validated as present by quantitative real-time RT-PCR in ten independent CLL samples, but not in 11 WT *SF3B1* CLL samples (Table S1, Figure 1F).

SF3B1 Mutation Causes Alternative Splicing

To confirm the effects of *SF3B1* mutation on RNA splicing, we cloned full-length WT and K700E-mutated SF3B1 using codon-optimized cDNA sequences (GenBank: KX881377). Overexpression of these constructs in the hematopoietic cell line K562 revealed SF3B1 expression within the cell nucleus, demonstrating that mutation does not affect the nuclear localization of SF3B1 (Figure 2A). N-terminal tagged, but not C-terminal tagged, mutated SF3B1 appropriately bound to other protein components of the U2 snRNP (Figures 2B, S2A, and S2B). Hence, mutated *SF3B1* still interacts with its binding partners, suggesting its ability to participate in pre-mRNA splicing.

Following the introduction of these full-length expression constructs into cell lines including K562, HeLa, U2OS, JeKo-1,
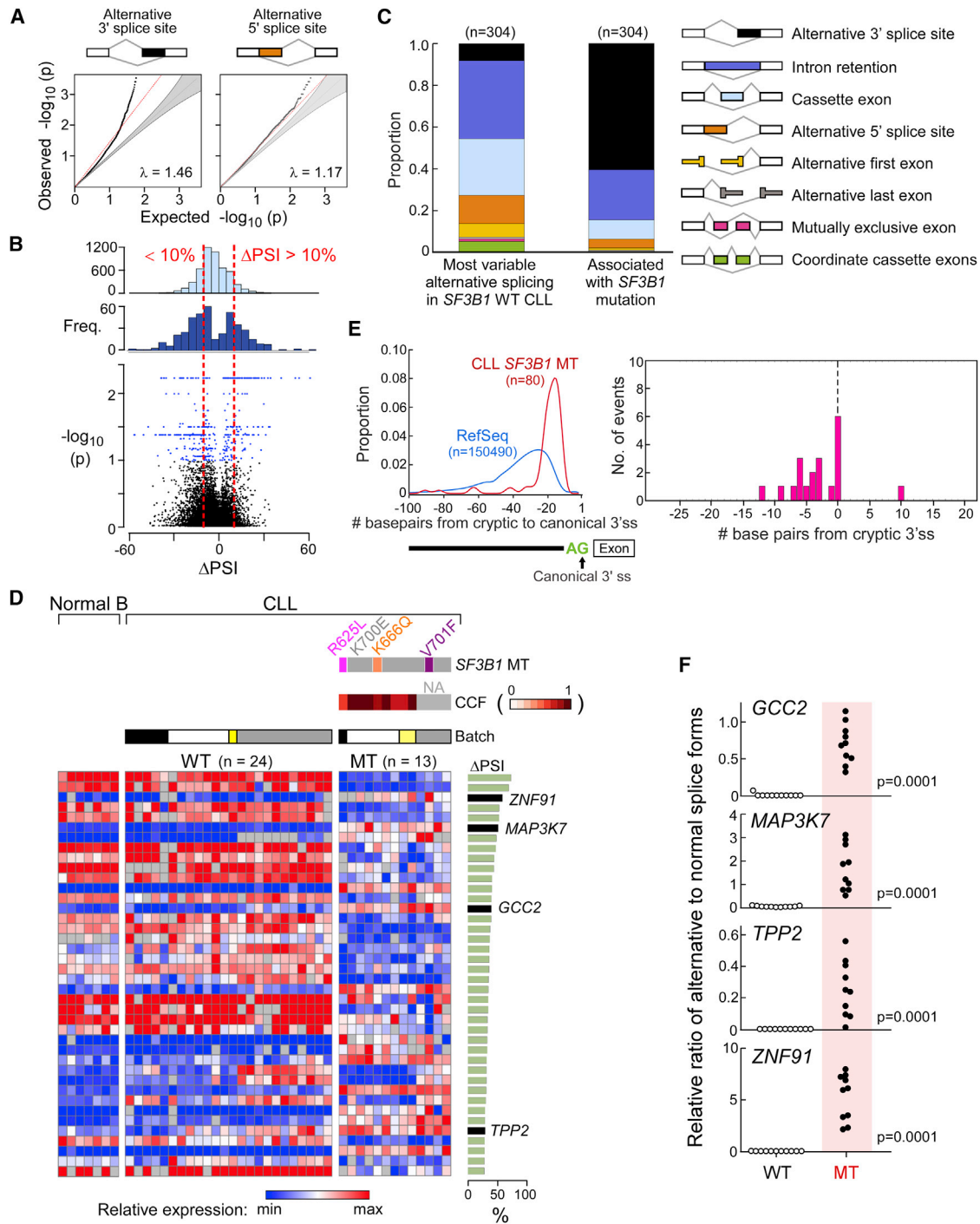
Figure 1. *SF3B1* Mutation Is Associated with Alternative Splicing at 3′ Splice Sites in CLL

(A) Q-Q plots comparing observed empirical values with expected p values between *SF3B1* wild-type (WT) and mutated (MT) CLL identified through the analysis of bulk poly(A) selected RNA-seq from 37 CLLs. Red line, the least-squares linear fit to the lower 95 percentile of points with slope λ. Gray-shaded areas, 95% confidence intervals for the expected distribution.

(B) Frequency of ΔPSI from random comparisons (top) or significant splice changes (middle, p < 0.05) from the RNA-seq data above and volcano plot of ΔPSI versus log₁₀(p) of all splicing changes (bottom). Red dotted lines, thresholds of ΔPSI of 10%. Blue dots, significant splicing events.

(C) Categories of alternative splicing within the 304 splice events significantly associated with mutant *SF3B1* in CLL versus the 304 most variable alternatively spliced events in WT CLL from bulk poly(A) selected RNA-seq.

(D) Heatmap of the top 40 alternatively spliced events with the highest ΔPSI between CLL samples with mutant (n = 13) and WT (n = 24) *SF3B1*. Expression of splice variants from RNA-seq analysis of CD19⁺ selected B cells from seven healthy adult volunteers indicated, along with RNA-seq batch labels, *SF3B1* mutation type, and clonality status. Right: ΔPSI for each splice event.

(legend continued on next page)
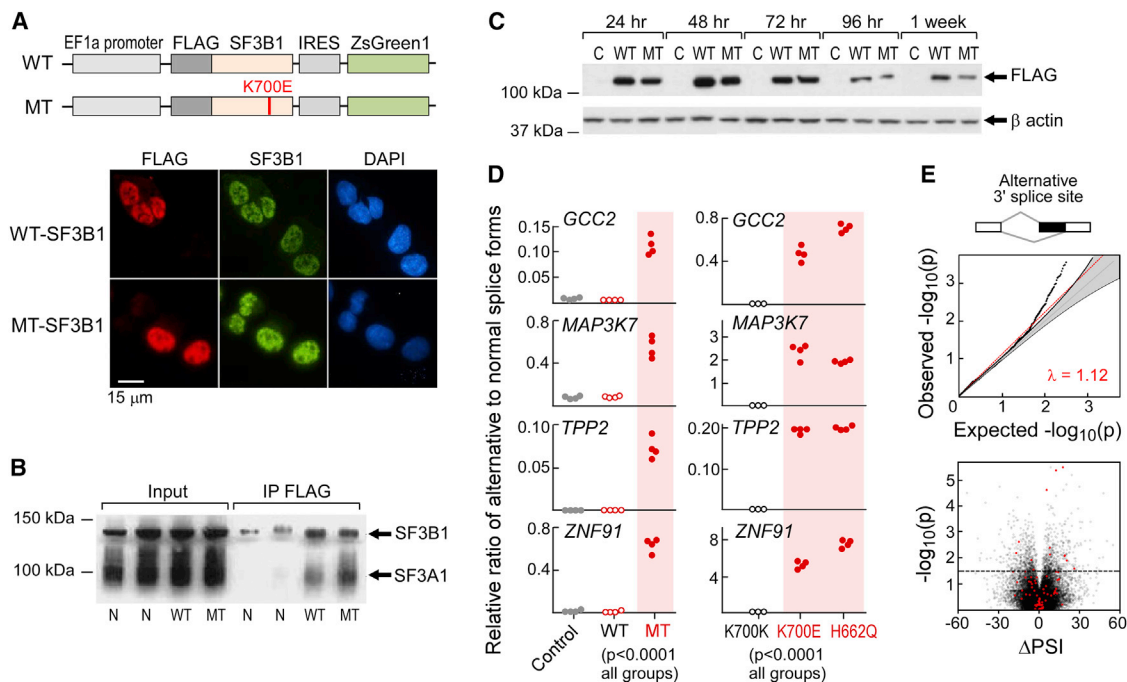
Figure 2. Expression of Mutant *SF3B1* Causes Alternative Splicing

(A) Schematics of WT and K700E MT SF3B1 expression constructs (top) and immunoflourescence staining of FLAG-tagged SF3B1 in K562 cells that were nucleofected with SF3B1 constructs (bottom).

(B) Cell lysate from HeLa cells overexpressing either WT or MT SF3B1 was immunprecipated with anti-FLAG antibody and probed with anti-SF3B1 antibody.

(C) FLAG-tagged mutant and WT SF3B1 protein were transiently expressed in HEK293T cells and detected by immunoblotting.

(D) Expression of alternative splicing associated with SF3B1 mutations in transfected K562 cells (left, n = 4 for each group) and in isogenic Nalm-6 cells (right) was assessed with qRT-PCR assays.

(E) Analysis of bulk poly(A) selected RNA-seq of K562 cells expressing SF3B1-K700E. Top panel- Q-Q plot of alternative 3′ splice sites between empirical p value of observed and expected spliced events. Lower panel, volcano plot of ΔPSI in relation to significance between WT and mutant SF3B1 overexpression in K562 cells for each splicing events. Red dots, significantly differentially spliced events also identified in the primary CLL analysis.

See also Figure S2.

and HG3, and also primary B cells, we observed expression of mutated SF3B1 protein over 24–72 hr (Figure 2C), which was associated with upregulation of splice variants identified from the aforementioned bulk RNA-seq analysis (Figure S2C, Table S2). For example, K562 cells transfected with the SF3B1-K700E expression construct revealed a more than 10-fold increase in candidate splice variant expression compared with cells expressing WT SF3B1 (p < 0.0001; Figure 2D). Likewise, expression of altered splice variants was observed in isogenic Nalm-6 cells, gene-edited at the endogenous *SF3B1* locus to express SF3B1[K700E] or SF3B1[H662Q] but not in matched control (SF3B1[K700K]) cells (p < 0.0001; Figure 2D) (Darman et al., 2015). JuncBASE analysis of RNA-seq data from K562 cells expressing mutant SF3B1 showed pervasive altered 3′ splice site use, and enrichment of the same splice variants observed in primary CLL samples (35 events, chi-square test, p < 0.0001) (Figures 2E and S2D, Table S3). Altogether, these results

link *SF3B1* mutation with the generation of splicing changes in CLL, either through direct or indirect interactions.

Single CLL Cells with Mutated *SF3B1* Demonstrate Altered Splicing

We confirmed the association of *SF3B1* mutation with splicing through the analysis of single cells. We adapted a sensitive microfluidics-based approach that uses multiplexed targeted amplification of RNA to simultaneously detect expressed genes, somatic mutations, and alternative splicing within a single cell (Supplemental Experimental Procedures). Assays for targeted detection of specific *SF3B1* WT versus mutant alleles in single cells were generated. In this manner, we confirmed the association of mutation in *SF3B1* with altered splicing through the analysis of hundreds of single cells within each sample. By bulk analysis, the *SF3B1*-G742D mutation was estimated to be present in 15% of the sample population of CLL096, while the

(E) Left: density plot of the positions of cryptic AGs relative to their canonical splice sites in *SF3B1* mutant samples, compared with the distance to the first AG (non-GAG trimer) from all RefSeq canonical 3′ splice sites. Right: relative positions of mapped branchpoints (BP) (n = 16) from Mercer et al. (2015). x axis, distance in nucleotides (nt) of the BP to the cryptic AG (upstream positions are negative distances); y axis, the frequency of BP found at that position.

(F) Validation of RNA-seq analysis through qPCR of selected significantly altered spliced events in independent CLL samples (11 with WT and 10 with MT *SF3B1*).

See also Figure S1.
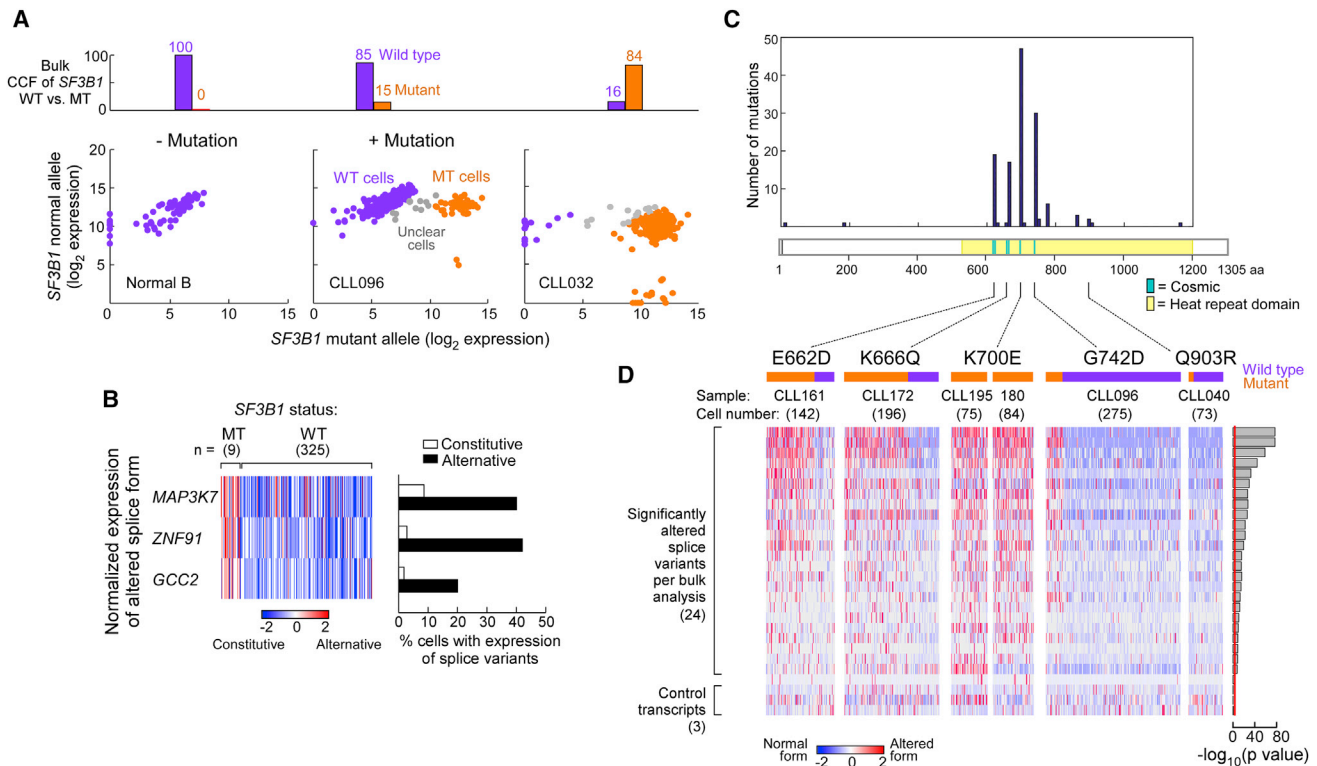
Figure 3. Single CLL Cells with *SF3B1* Mutation Express Alternatively Spliced RNAs

(A) Example of *SF3B1* mutation call in single normal B cells and from CLL cells from samples with either subclonal (CLL096) or clonal (CLL032) *SF3B1* mutation. Log$_2$ transformed expression of mutant versus WT *SF3B1* alleles are plotted, with each dot representing one single cell. Purple, cells identified to be *SF3B1* WT; orange, cells inferred to be *SF3B1* mutant; gray, cells with ambiguous calls.

(B) Expression of alternative versus constitutive transcript relative to total expression of the genes *MAP3K7, ZNF91*, and *GCC2* from single cells of sample CLL096.

(C) Frequency of mutations in *SF3B1* at different mutation sites from a recent study of 538 CLL samples (Landau et al., 2015). Yellow shading, heat repeat region; blue, mutation sites reported previously from the COSMIC database.

(D) Single-cell profiling of splice variant expression across five *SF3B1* mutations from six CLL samples (one cell per column). Expression of the alternative transcript relative to total gene expression in 24 selected genes, identified from the bulk poly(A) RNA-seq analysis, was scored along with three control genes. Orange and purple bars indicate cells with and without mutation, respectively.

SF3B1-K700E mutation was nearly clonal (>84%) in sample CLL032. We established background expression thresholds using 96 single CD19$^+$ B cells from a healthy adult with no *SF3B1* mutation (Figure 3A, Table S4). Consistent with the bulk estimates, 44 of 367 (12%) CLL096 single cells were positive for *SF3B1*-G742D, whereas 215 of 288 (75%) CLL032 single cells were positive for *SF3B1*-K700E. Assays for detection of constitutive and alternative forms of *MAP3K7, ZNF91*, and *GCC2* transcripts (identified as significantly alternatively spliced from the bulk analysis) were also generated for single cells. In the single cells with mutated *SF3B1* from CLL096, we observed significantly higher levels of the altered *SF3B1* mutation-associated splice variants (Figure 3B, Table S4).

To explore if other *SF3B1* mutations result in a spectrum of altered splicing similar to K700E, we examined 845 cells from six patient samples across five different *SF3B1* mutations using 45 splice variant assays (Table S4). For 4 of 5 mutations (E622D, K666Q, K700E, and G742D), we observed highly similar patterns of significant altered splicing (p < 0.05) for 24 of 45 splice variants. The one *SF3B1* mutation (Q903R) positioned farther from

the mutation-enriched region (but still in the HEAT repeat domain) did not exhibit increased alternative splicing for the selected K700E mutation-associated variants (Figure 3C and Table S1). Splicing events not found to be associated with the *SF3B1*-K700E mutation also did not show splicing changes for other HEAT repeat mutations, supporting the idea that the shared splicing events are specific to *SF3B1* mutation rather than a generalized splicing defect in CLL.

Preferential 3′ Splice Site Alterations in *SF3B1* Mutated Samples Are Also Observed in CLL Total RNA

Since SF3B1 acts on pre-mRNA and splicing can occur co-transcriptionally in advance of polyadenylation, the full effects of *SF3B1* mutation on splicing may be masked by sequencing of only fully processed and stable mRNAs. We therefore examined the transcriptome changes of non-poly(A)-selected total RNA by RNA-seq from three WT and three *SF3B1* mutated samples, using an rRNA depletion method (Adiconis et al., 2013; Levin et al., 2010). Three of six samples (one WT and two *SF3B1* mutated) were included in the aforementioned poly(A)-selected
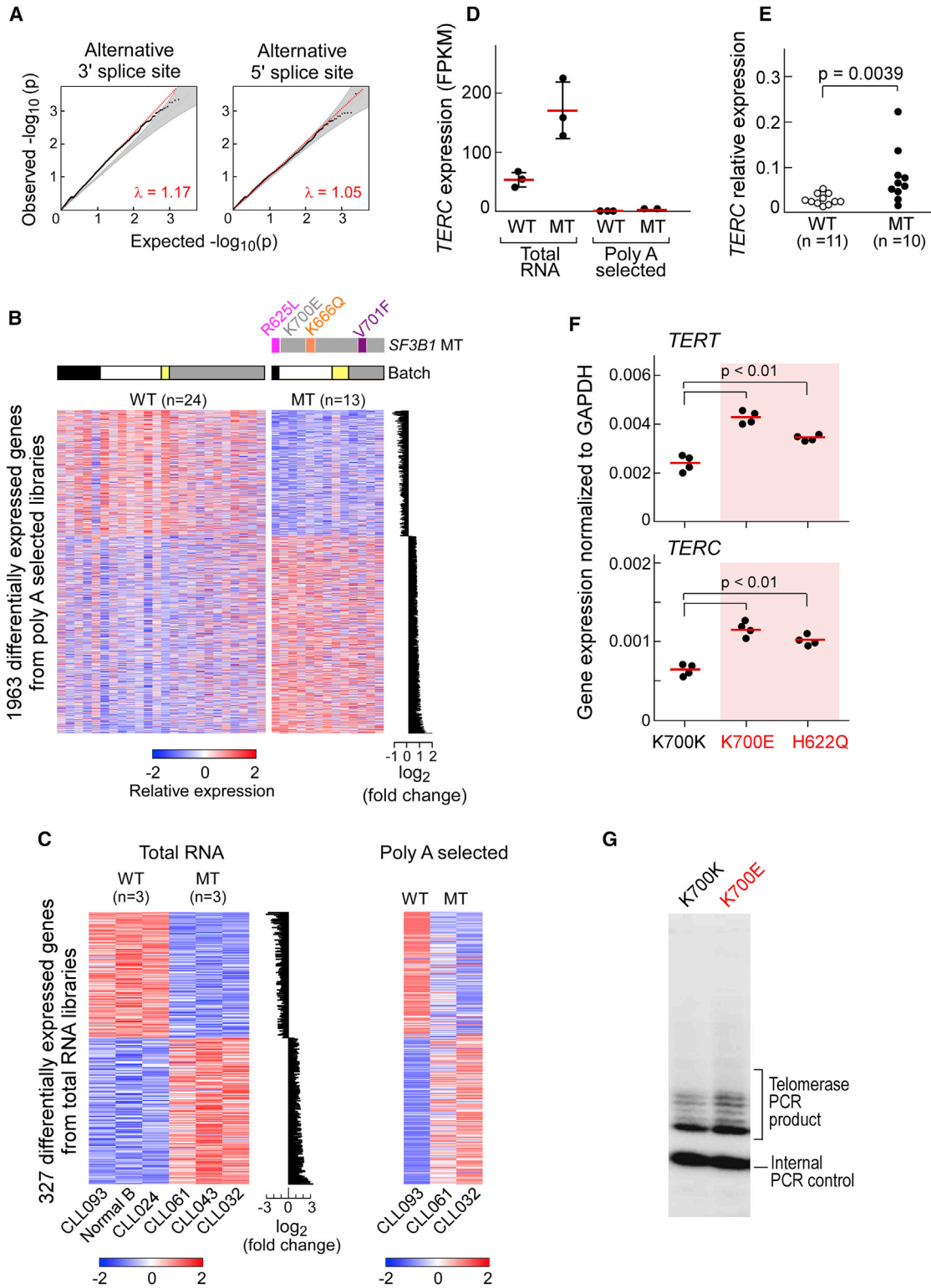
Figure 4. Affected Genes Associated with *SF3B1* Mutation in CLL Samples

(A) Q-Q plots of alternative 5′ and 3′ splice sites between empirical p value of observed and expected spliced events from RNA-seq analysis of CLL total RNA libraries.

(B) Heatmap visualizations of significantly differentially expressed genes (batch corrected, adjusted p value < 0.2) between CLL samples with or without *SF3B1* mutation derived from poly(A) selected RNA. The barplot at the right shows the average log$_2$ fold expression difference between the two conditions.

(legend continued on next page)

RNA-seq analysis (Table S1). From the total RNA libraries, the number of mapped paired reads (median of 27 million) was comparable with those of the poly(A)-selected libraries (median of 25 million) (Figure S3A, left). However, 55% of reads from sequenced total RNA mapped within introns compared with 28% from the same RNA sample processed by poly(A) selection, consistent with the increased sequencing of pre-mRNA from these libraries (Figure S3A, right). We again observed an enrichment of 3′ splice site changes associated with *SF3B1* mutation, further supporting the idea that this effect is unlikely due to a bias in the stability of fully processed mRNAs (Figure 4A, Table S5).

Sequencing data from total RNA provided an opportunity to directly investigate branchpoint usage when *SF3B1* is mutated by identifying reads derived from spliced lariats that spanned the branchpoint (Taggart et al., 2012). As expected, since we did not enrich for spliced lariats (Mercer et al., 2015), we only identified branchpoint support for three events that were differentially spliced in the presence of *SF3B1* mutation and 14 in a set of control splicing events after stringent filtering (Experimental Procedures, Table S5). For the event in *GCC2* where cryptic 3′ splice site use was observed upstream of the annotated splice site in *SF3B1* mutated samples, there were 11 lariat reads among three *SF3B1* mutated samples supporting upstream alternative branchpoint usage and four lariat reads supporting downstream branchpoint use compared with two branchpoint positions identified in an *SF3B1* WT sample (Table S5). Improved genome-wide approaches to enrich for branchpoint spanning reads in *SF3B1* mutated samples will provide more conclusive evidence of differential branchpoint usage.

SF3B1 Mutation Impacts Multiple Cellular Pathways

To investigate the global effects of *SF3B1* mutation on cellular processes in CLL, we examined whether coherent changes in cellular pathways could be observed in samples with *SF3B1* mutation compared with those without the mutation. We identified 1,963 and 327 significantly differentially expressed genes (batch corrected, adjusted p value <0.2) between these two groups through analysis of the poly(A) and total RNA libraries, respectively (Figures 4B and 4C, Table S6). Differentially expressed genes from the total RNA data exhibited similar trends and directionalities in matched poly(A) data (Figure 4C) and vice versa (Figures S3B–S3D). Ninety-nine of 327 genes from the total RNA libraries overlapped with differentially expressed genes identified from the poly(A)-selected libraries (Figure S3E). In addition, 25% of significantly differentially expressed genes were found in common between the primary CLL datasets and with previously reported data from the Nalm-6 lines (Darman et al., 2015).

We explored the spectrum of non-polyadenylated transcripts associated with *SF3B1* mutation through RNA-seq analysis of the total RNA libraries. One noteworthy target with significantly increased expression in mutant *SF3B1* samples was *TERC*, encoding an essential RNA component of telomerase. A *TERC* transcript could be amplified in total RNA from *SF3B1* mutated samples, but was undetected in the matched poly(A) RNA (Figure 4D). We validated this discovery in independent CLL samples (p = 0.0039) using cDNA generated by random hexamers (Figure 4E). Examination of the Nalm6 isogenic cell lines showed both K700E and H622Q mutations to lead to upregulation of *TERC* and *TERT* gene expression (Figure 4F). Since *TERC* overexpression can increase telomerase activity in activated lymphoid cells (Weng et al., 1997), and the spliceosome has been implicated in telomerase RNA processing in yeast (Qi et al., 2015), we investigated if *SF3B1* mutation impacted telomerase activity. Indeed, SF3B1^{K700E} Nalm-6 cells had higher telomerase activity than cell lines with the silent mutation by a sensitive in vitro PCR-based assay (Figure 4G). Altogether, these data suggest that mutant *SF3B1* may affect telomerase activity through dysregulated *TERC* and *TERT* expression.

To more generally identify possible cellular processes altered by *SF3B1* mutation, we performed gene set enrichment analysis (GSEA) against 1,970 gene sets from MSigDb, and 14 additional manually curated gene sets from pathways known to be altered in CLL. From the 1,963 differentially expressed genes from the poly(A) RNA-seq data, we identified 180 significantly upregulated and 37 downregulated gene sets (q value < 0.1), with enrichment across diverse cellular processes related to cancer and CLL (Table S7). Likewise, GSEA of 327 differentially expressed genes from the total RNA data and of 341 *SF3B1* mutation-associated splice variants (all q value < 0.1) revealed involvement in a similar spectrum of diverse cellular processes (Figure 5A, Table S7).

Signals obtained from bulk analyses reflect only the average characteristics of a population. Since *SF3B1* mutation is commonly a subclonal event in CLL (Landau et al., 2013), we assessed the effects of *SF3B1* mutation on the transcriptome of the subpopulation of single cells within CLL cases with mutated *SF3B1* (versus cases with WT *SF3B1*). We performed whole-transcriptome amplification and sequencing of up to 96 individual cells per sample from two primary CLL samples with (CLL032 and CLL096) and two without (CLL003, CLL005) *SF3B1* mutation. We noted a poor ability to directly call *SF3B1* mutation status from these data, as *SF3B1* expression across single cells was highly variable (mean *SF3B1* expression 870.20 counts ± 1677.15 SD) and lacking in 11 cells. Likewise, coverage at the *SF3B1* mutation site was highly variable and

(C) Left: visualization of the 327 significantly differentially expressed genes (batch corrected, adjusted p value < 0.2) identified from total RNA libraries derived from one sample of normal CD19$^+$ B cells, two *SF3B1* wild-type samples, and three *SF3B1* mutant CLL samples. Right: heatmap of the same genes, extracted from RNA-seq data prepared from poly(A) selected libraries of matched CLL samples.

(D) *TERC* expression in CLL samples with *SF3B1* mutation from RNA-seq data prepared from either total RNA or poly(A) selected libraries. Mean (red line) ± SD; n = 3.

(E) *TERC* expression was assessed in independent CLL samples with mutant or WT *SF3B1*, with cDNA prepared with random hexamers.

(F) *TERC* and *TERT* expression was examined in Nalm-6 cells gene-edited to express SF3B1^{K700E} and SF3B1^{H622Q} (compared with SF3B1^{K700K}), with cDNA prepared with random hexamers. Red line indicates mean.

(G) Telomerase activity in Nalm-6 cells expressing either SF3B1^{K700E} or SF3B1^{K700K} was measured with telomeric repeat amplification protocol assay. Shown are representative results from one of three experiments.

See also Figure S3.
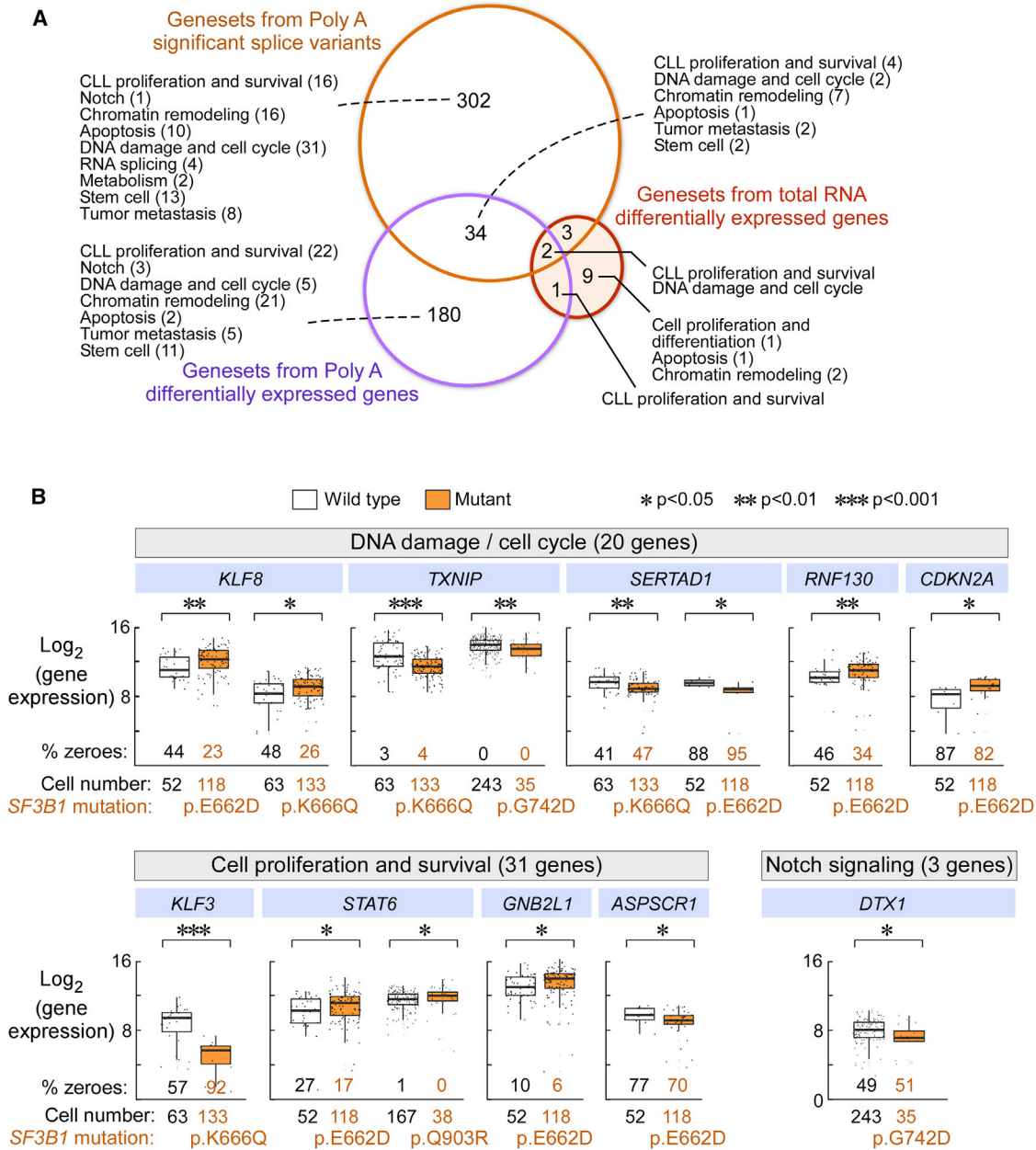
Figure 5. *SF3B1* **Mutations Affect Genes Involved in Multiple CLL-Associated Pathways**
(A) An analysis of overlap of enriched gene sets (all with q value < 0.1) using splice variants and differentially expressed genes associated with mutant *SF3B1* derived from total and poly(A) selected RNA libraries. This analysis includes 1,970 gene sets from MSigDb and 14 additional manually curated gene sets. Each number represents the number of gene sets enriched in the different groups.
(B) Gene expression of single CLL cells from samples with subclonal *SF3B1* mutation that were tested in parallel against a panel of 96 genes, encompassing targets of CLL-associated pathways. Significantly differentially expressed genes associated with expression of *SF3B1* mutation in individual cells associated with the DNA damage response and cell-cycle regulation, proliferation and survival, and Notch signaling are shown. 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, and lines above and below the box show the minimum and maximum. p Values were defined by a two-sided Wilcoxon rank-sum test. "% zeroes" represents the percent of total cells samples with zero expression for that particular genes.
See also Figure S4.

lacking in the majority of cells from CLL032 and CLL096, such that we were only able to confidently call *SF3B1* mutation in 12 cells. As an alternative, we attempted to infer *SF3B1* mutation status of individual cells based on expression of 59 significantly 3′ alternatively spliced isoforms associated with *SF3B1* mutation (per bulk RNA-seq analysis) and with detectable expression in single cells. Still, with this strategy, we could readily distinguish between bulk samples with or without *SF3B1* mutation but
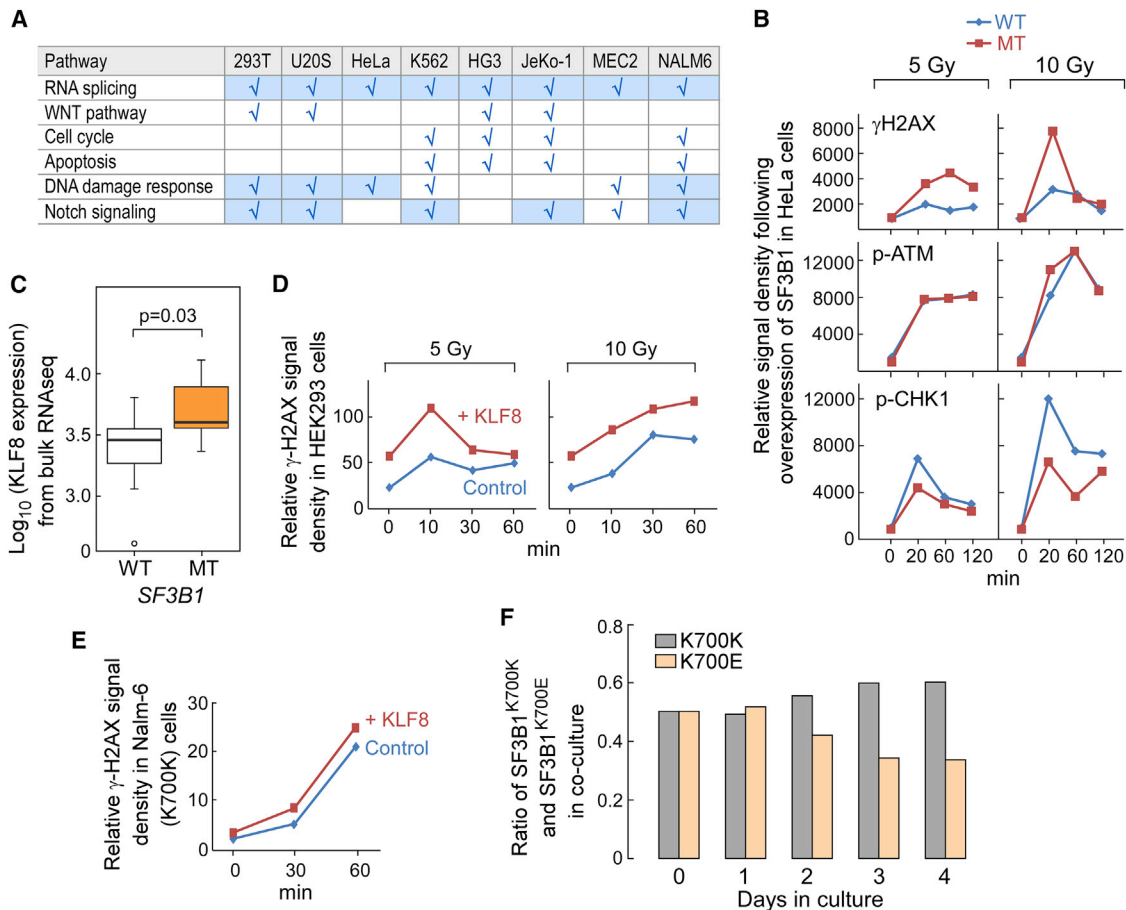
Figure 6. Functional Assessment of SF3B1-K700E on Cellular Processes in CLL and in Cell Lines

(A) Summary of effects of expressing SF3B1-K700E on six CLL-associated pathways examined in different cell lines. Shaded cell lines and pathways for which SF3B1-K700E expression appeared to exert a functional change in the cell type evaluated.

(B) Levels of phosphorylated forms of H2AX, ATM, and CHK1 in HeLa cells upon γ-irradiation were examined from cells overexpressing WT or mutant SF3B1 for 48 hr. Relative signal intensity was plotted based on ImageJ quantification of bands from the raw image.

(C) Box-and-whisker plots of *KLF8* expression in bulk CLL samples with *SF3B1* mutation compared with WT *SF3B1* derived from the RNA-seq dataset was plotted (median-center line within box; bottom and top lines of box represent the 25th and the 75th quartile, respectively, while whiskers above and below the box show the minimum and maximum values).

(D and E) DNA damage response was assessed in HEK293 (D) and Nalm-6 (E) SF3B1^{K700K} cells that were transfected or nucleofected with control or KLF8 expressing constructs upon various doses of γ-irradiation, and protein levels of phosphorylated form of H2AX protein expression was quantified using ImageJ.

(F) Nalm-6 SF3B1^{K700E} and SF3B1^{K700K} cells were co-cultured and the percentage of each cell population was assessed daily. Shown is representative data from two independent culture experiments.

See also Figure S5.

were nonetheless unable to confidently ascertain mutation status on the majority of individual cells (Figure S4, Table S8).

We therefore used a microfluidics-based targeted gene expression and *SF3B1*-mutation detection approach to identify significant changes in genes associated with pathways implicated by GSEA in single cells. We developed a panel of 96 gene expression assays encompassing target genes across the diverse affected pathways (Table S9), against which we interrogated expression of 1,109 *SF3B1* WT and mutant single cells from samples CLL161, 171, 096, or 040. Across the four samples, single cells with *SF3B1* mutation were significantly associated with changes in genes related to poor prognosis in CLL and cancer (*KLF3, TYROBP, DDIT4, FYN, GNBL2, TRABD, STAT6,* and *ZBTB48*), apoptosis (*BIRC3, BCL2, SH3BP1,*

KLH21, and *TIMP1*), DNA damage and cell cycle (*KLF8*, *ATM, FANCD2, CDKN2A, CCND1, CCNE1, TXNIP, RNF130, ANAPC7, KLF3, LRWD1, SKP2, SERTAD1,* and *FRK*) and Notch signaling (*DTX1*) (p < 0.05; Figure 5B and Table S9).

Next we systematically examined evidence of change in other pathways characterized previously as affected in CLL (Landau et al., 2013) through overexpression of WT versus mutated *SF3B1* in cell lines or in the isogenic Nalm-6 cells using established pathway readout assays. We did not detect any impact of overexpressed *SF3B1* mutation on Wnt pathway signaling, or on the cell cycle or apoptosis (Figure 6A). Consistent with the single-cell results, we found evidence of altered DNA damage response with overexpression of mutated *SF3B1* in cell lines. HeLa cells overexpressing mutant SF3B1 demonstrated

greater evidence of DNA damage (Figures 6B and S5A). Since KLF8, a Kruppel-like transcription factor implicated in tumor transformation, progression, and DNA damage repair in solid tumors (Lu et al., 2012), was consistently upregulated in *SF3B1* mutated samples at both single-cell (Figure 5B) and bulk RNA levels (Figure 6C), we tested the effects of overexpression of KLF8 on this pathway. With overexpression of KLF8 in HEK293 and Nalm-6 SF3B1^{K700K} cells, higher levels of DNA damage were consistently induced, as detected by γH2AX expression and attenuated phosphorylation of CHK2 following exposure to γ-irradiation (Figures 6D, 6E, S5B, and S5C). These results support *SF3B1* mutation-associated gene dysregulation as a contributor to altered DNA response. In the Nalm-6 cell lines, we consistently observed a subtle growth disadvantage associated with *SF3B1* mutation when in co-culture with the SF3B1^{K700K}-expressing cells (Figure 6F), although this was not due to mutation-induced cell-cycle arrest or apoptosis (data not shown).

SF3B1 Mutation Affects Notch Signaling through a Splice Variant of *DVL2*

Using a well-characterized Notch luciferase-reporter-assay system (Minoguchi et al., 1997), we detected significantly higher Notch pathway activation induced across myeloid and lymphoid cells lines expressing mutated SF3B1 compared with WT SF3B1 when Notch signaling in these cells was activated by co-expression of an active form of Notch1 (Figure 7A, Figures S6A and S6B). We confirmed that the observed upregulation of Notch signaling was not due to changes in activated Notch1 since its levels were detected at equivalent levels in the cell lines expressing either WT or mutant SF3B1 (Figure S6C).

Given these unexpected observations, we considered whether altered spliced variants associated with SF3B1 mutation could mediate downstream Notch pathway signaling. Focusing on splicing events in genes involved in Notch signaling identified from GSEA of splice variants, we identified an altered splicing event in *DVL2* as a promising candidate target (PSI = 29.6). A core canonical Wnt pathway member, DVL2 has also been reported previously to negatively regulate Notch signaling (Collu et al., 2012). Examination of the RNA-seq level evidence revealed CLL samples with *SF3B1* mutation to exhibit preferential 3′ altered splicing between exons 10 and 11 of *DVL2*, leading to an in-frame 24-amino-acid deletion (Figure 7B). By RT-PCR of a transcript spanning exons 10 and 11, we could detect both normal and altered *DVL2* transcripts in K562 cells overexpressing mutated SF3B1 but only the normal variant in cells overexpressing the WT protein, and this was confirmed by Sanger sequencing of the constitutive and altered products (Figure 7C). Moreover, we confirmed higher expression of altered rather than constitutive *DVL2* in Nalm-6 SF3B1^{K700E} and SF3B1^{H622Q} cells compared with Nalm-6 SF3B1^{K700K} cells (Figure 7D).

A more thorough examination of *DVL2* transcript expression across bulk RNA-seq datasets confirmed the strong association between expression of altered *DVL2* transcript and of *SF3B1* mutation. Altered *DVL2* was absent or minimally expressed across RNA-seq samples from 693 normal tissues within the GTEx collection (GTEx Consortium, 2015), 7 normal B cells, or from 24 CLLs as well as cell lines (K562 and HEK293T) with

WT SF3B1. In contrast, altered *DVL2* expression of >10% PSI was detected in CLLs with mutated *SF3B1* or in cell lines overexpressing mutated *SF3B1* (Figure 7E). We confirmed the higher expression of the *DVL2* splice variant in individual primary CLL cells with *SF3B1* mutation compared with those without, despite equivalent total *DVL2* expression between these two cell populations (Figure 7F). In addition, the protein product of altered *DVL2* was readily identified by immunoblot in the isogenic B cell lines (Nalm-6 SF3B1^{K700E}, SF3B1^{H622Q}), with overexpression of the SF3B1-K700E construct in K562 cells, and in primary CLL samples with *SF3B1* mutation but not in those without (Figures 7G, 7H, and S6D). Expression of altered *DVL2* at transcript and protein levels was stable for at least 96 hr following expression of mutated SF3B1 in cell lines, with similar degradation rates between altered and WT DVL2 after exposure to the protein synthesis inhibitor cycloheximide (Figures S6E–S6G).

To directly test whether altered *DVL2* dysregulates Notch signaling, we subcloned *DVL2* WT and altered cDNA into expression vectors (Figures S7A and S7B), and tested if expression of altered DVL2 could change Notch signaling. Although the Notch pathway was clearly activated in K562 cells following overexpression of full-length SF3B1-K700E (Figures S6A and S6B), we focused on developing a reliable system to interrogate this pathway in a B cell context. We therefore generated stable B cell lines (from the lymphoma OCI-Ly1 line) expressing either WT or altered *DVL2* or both through lentivirus-mediated transduction (Figure 8A), and then activated the Notch pathway through co-culture with OP-9 cells overexpressing the Notch ligand delta 1 ("OP9-DL1 cells"). In the setting of Notch pathway activation, WT DVL2 repressed Notch signaling, as reported previously (Collu et al., 2012). In contrast, expression of altered DVL2 markedly abrogated these repressive effects. Moreover, combined expression of WT and altered DVL2 also reversed these repressive effects, suggesting the dominant impact of altered DVL2 on the WT isoform (Figure 8B). In line with these findings, expression of the downstream Notch pathway target gene *HES1* was higher in the presence of altered DVL2 and of combined altered and WT DVL2 than when WT DVL2 was expressed alone (p < 0.05) (Figure 8C). The impact of altered *DVL2* on Notch pathway activation was independent of an effect on Wnt pathway signaling, since altered and WT *DVL2* demonstrated equivalent potencies for activating the Wnt signaling (Figures S7C and S7D). Altogether, these data identify *DVL2* as a target of mutated *SF3B1* through which alternative splicing modulates Notch signaling activity.

DISCUSSION

Mutations in transcription factors can wield a multitude of effects on cancer cells; mutated *TP53* is a prototypical example of this concept (Bieging et al., 2014). Our comprehensive transcriptome characterization of mutated *SF3B1* in CLL, together with the growing knowledge of the types of transcript alterations arising from other cancer-associated mutated splicing factors and the subsequent effects on cellular transformation (e.g., mutated *U2AF1* and *SRSF2*) (Brooks et al., 2014; Ilagan et al., 2015; Kim et al., 2015; Park et al., 2016; Shirai et al., 2015) suggest that mutated splicing factors in cancer may well behave in an
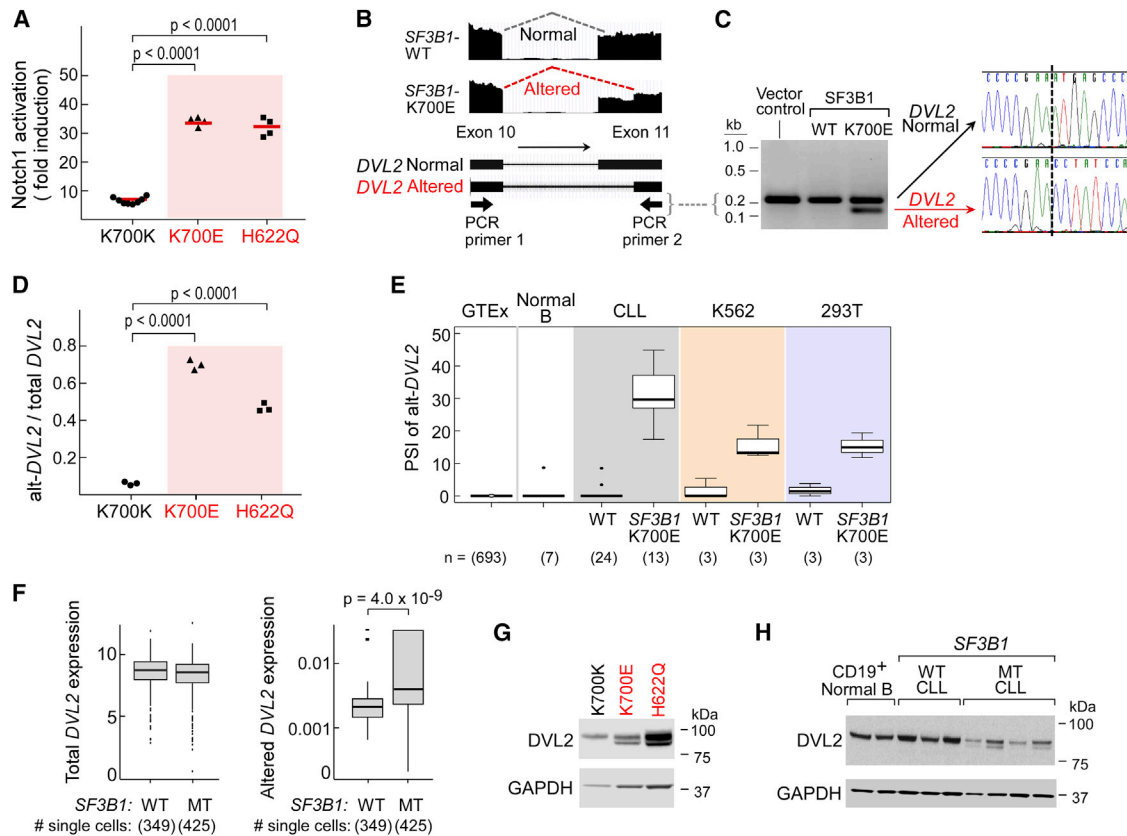
Figure 7. *SF3B1* Mutation Modulates Notch Signaling and Generates Alternative Splicing of *DVL2*

(A) Nalm-6 SF3B1[K700K], SF3B1[H622Q], and SF3B1[K700E] cells were nucleofected with a Notch luciferase reporter along with a GFP-expressing plasmid in the presence or absence of a Notch1-expressing construct. Forty-eight hours after the nucleofection, GFP-positive cells were isolated, and Notch activity in these cells was assessed by measuring luciferase activity.

(B) Visualization of the *DVL2* transcript from RNA-seq of CLL samples with and without *SF3B1* mutation.

(C) Detection of *DVL2* transcript using primers that cross exons 10 and 11 with cDNA derived from K562 cells overexpressing the vector control, or the WT or mutant SF3B1. Sanger sequencing of the two different fragments from mutant samples revealed the same junction as detected in the RNA-seq. The dashed vertical line indicates the junction of RNA splicing.

(D) *DVL2* alternative transcript levels in the Nalm-6[K700E], Nalm6[H622Q], and Nalm-6[K700K] cells.

(E) Expression of the alternative splice isoform of *DVL2* mRNA (alt-*DVL2*) is evaluated in relation to *SF3B1* mutation. PSI of alt-*DVL2* in RNA-seq data from the GTEx consortium (693 samples from blood, brain, breast, lung, and colon), 7 samples of normal B cells, 37 CLL samples and K562 and HEK293T cells over-expressing WT or mutant SF3B1 (median-center line within box; bottom and top lines of box represent the 25th and the 75th quartile, respectively, while whiskers above and below the box show the minimum and maximum values). Outliers are displayed as points.

(F) Total mRNA or the alternative form of *DVL2* expression in single CLL cells with MT or WT *SF3B1*, using samples and an analysis approach per Figure 3D. 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, and lines above and below the box show the minimum and maximum. Outliers displayed as points.

(G and H) Detection of protein expression of alternative DVL2 in Nalm-6 isogenic cell lines (G) and primary CLL samples with or without *SF3B1* mutation (H). See also Figure S6.

analogous fashion. We now demonstrate that mutated *SF3B1* induces hundreds of alterations both through splicing and dysregulated gene expression, with involvement of these RNA changes across diverse cellular processes previously implicated in CLL through somatic mutation characterization.

The majority of *SF3B1* mutation-associated splicing changes have been reported previously to create transcripts with premature stop codons, resulting in truncated proteins or downregulation of gene expression through nonsense-mediated decay (Darman et al., 2015; DeBoever et al., 2015; Quesada et al., 2012). We likewise observed numerous splicing changes in CLL samples with *SF3B1* mutations that would cause a frame-

shift in the resulting protein (69% with out-of-frame changes). However, of the in-frame splicing alterations, we identified an alteration in *DVL2* that increases Notch signaling, a driving pathway in CLL (Puente et al., 2011, 2015; Wang et al., 2011). The Notch pathway was first highlighted as an important CLL pathway upon the discovery of recurrent frameshift *NOTCH1* mutations in 10%–15% of CLL patients from early whole-exome sequencing (WES) studies (Puente et al., 2011; Wang et al., 2011). In CLL, somatic *NOTCH1* mutations have been characterized as pathway-activating (Puente et al., 2011) with the induction of apoptosis resistance (Rosati et al., 2009; Zweidler-McKay et al., 2005), and have been associated with poorer
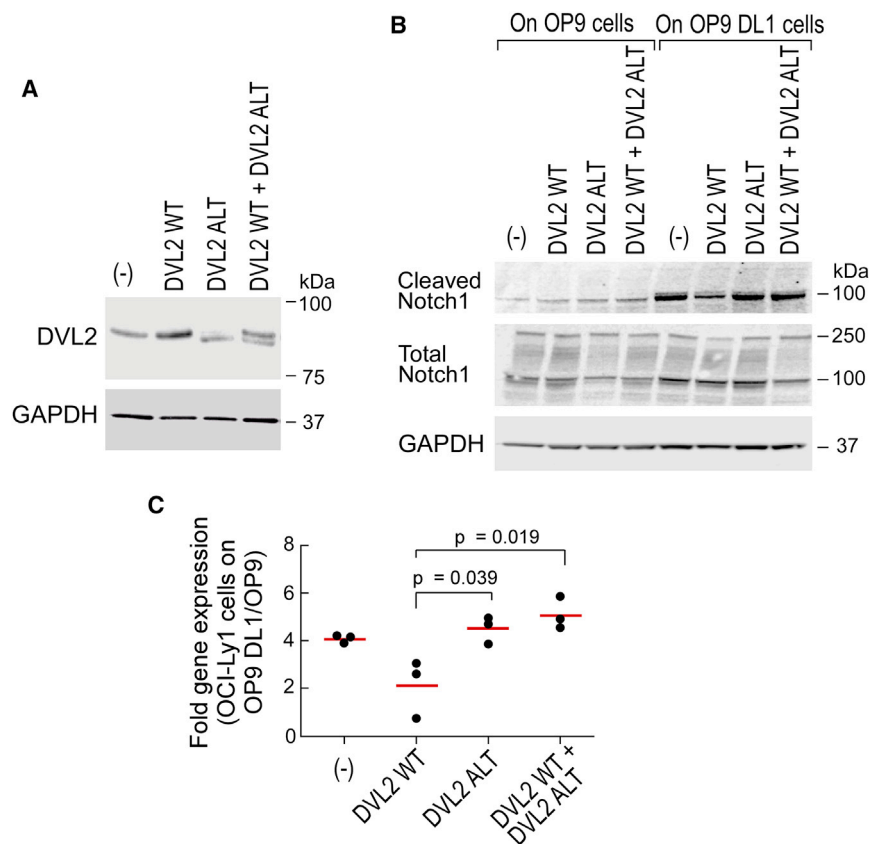
Figure 8. _SF3B1_ Mutation Modulates Notch Signaling through a Splice Variant of DVL2

(A) Detection of DVL2 expression in OCI-Ly1 B cell lymphoma cells with stable expression of WT, altered (ALT), or combined WT and altered forms (WT + ALT) of DVL2 by immunoblot.

(B) Activation of Notch1 in OCI-Ly1 cells expressing different DVL2 isoforms co-cultured with OP9 or OP9-DL1 cells for 48 hr assessed by immunoblot.

(C) Expression of the Notch target gene _HES1_ in OCI-Ly1 cells described in (A) co-cultured with either OP9 or OP9-DL1 cells was measured by qRT-PCR. Red line indicates mean.

See also Figure S7.

prognosis (Puente et al., 2011). Recently, recurrent activating mutations in the 3′ UTR of _NOTCH1_ associated with worse prognosis were identified (Puente et al., 2015). Our findings uncover yet another mode by which Notch signaling is activated in CLL. Of note, among the 229 of 538 CLL samples with unmutated _IGHV_ recently characterized by WES (Landau et al., 2015), only two samples have co-occurring _NOTCH1_ and _SF3B1_ mutations, which is significantly less frequent than expected by chance under an assumption of independence (p = 0.012). Our findings thus support the idea that at least some _SF3B1_ mutation-induced splicing changes include "driver" rather than "passenger" transcript variants, and reinforce the notion that multiple convergent molecular mechanisms can be used by cancer cells to dysregulate core cancer pathways.

While we focused on the functional effects of altered _DVL2_, we noted several other splice variants in mutated _SF3B1_ CLLs with ΔPSI >10% predicted to be pathway-altering in genes linked to Notch signaling (e.g., _DNAJC3_, _TRIP12_, and _HDAC7_). Moreover, our results extend findings from other investigators that have suggested subtle changes in the DNA damage response by _SF3B1_ mutation (Te Raa et al., 2015). A number of altered splice variants in this pathway would be predicted to impose effects on DNA damage response (_CHD1L_, _GAK_, _RAD9A_, and _JMY_). Of the non-polyadenylated transcripts, we found evidence of _TERC_ overexpression associated with _SF3B1_ mutation. Finally, we observed _KLF3_ and _KLF8_ as the most consistently differential expressed genes across the patient samples with _SF3B1_ mutation. Both genes have been associated with oncogenic transformation, cell-cycle regulation, DNA damage response, and cell

differentiation (Lu et al., 2012; Wang and Zhao, 2007). Consistent with this, we observed dysregulated DNA damage responses in cell lines overexpressing _KLF8_. Our aggregate results support the idea that multiple alterations in transcript sequence or expression impact CLL in a concerted fashion across CLL pathways, and our findings provide rich fodder for future in-depth functional studies.

We emphasize that the numerous transcript-level changes induced by mutated _SF3B1_ were subtle overall. The majority of identified variants from both poly(A)-selected and total CLL RNA had ΔPSI values of less than 10%. We focused on spliced transcripts with ΔPSI >10%, but we do not exclude the possibility that lowly expressed altered transcripts could have functional impact. For example, an altered _ATM_ transcript, previously identified as associated with _SF3B1_ mutation (Ferreira et al., 2014), was also found within our dataset, but with a ΔPSI of <4% (data not shown). Regardless of the ΔPSI cutoff, _SF3B1_ mutation appears to exert numerous transcript-level changes on a broad range of genes at relatively low intensity per gene to modulate CLL biology. We speculate that this pattern of activity may affect oncogenesis by allowing the cancer cell to tolerate many changes, even in essential genes, without drastically affecting cell viability, hence allowing these changes to be propagated. At the same time, these broad changes vastly increase the diversity of gene expression in the cancer cell and would be anticipated to enhance its evolutionary capacity. Indeed, this may explain why mutation in _SF3B1_ in otherwise normal hematopoietic cells, recently described as contributing to clonal hematopoiesis (Jaiswal et al., 2014; Xie et al., 2014), does not alone drive cancer in a B cell. However, consistent with _SF3B1_ mutation as a "later" or commonly subclonal CLL event (Landau et al., 2013), mutation in _SF3B1_ in the backdrop of other cancer-driving alterations could push the CLL cell toward a more aggressive phenotype. Recently, another recurrently mutated splicing factor SRSF2 was likewise observed to generate multiple subtle changes in splicing (Zhang et al., 2015), further hinting that this is a general mode of action by mutated cancer-associated splicing factor genes.

EXPERIMENTAL PROCEDURES

Human Samples

Heparinized blood samples were obtained from healthy donors and patients enrolled on clinical research protocols with informed consent, approved by the Human Subjects Protection Committee of the Dana-Farber Cancer Institute (DFCI) and at UCSD (CLL Research Consortium). For 21 samples, *SF3B1* mutation status was confirmed by targeted sequencing.

Full-Length SF3B1 Expression Construct and Gene-Edited Cell Lines

Full-length *SF3B1* was constructed as described in the Supplemental Experimental Procedures. In brief, full-length *SF3B1* was cloned using a partial fragment of SF3B1 cDNA (a gift from Dr. Robin Reed, Harvard Medical School) (Wang et al., 1998), and ligating this to a codon-optimized synthetic fragment encoding the uncloned region of 414 nucleotides (Blue Heron Biotechnology). Pre-B Nalm-6 isogenic cell lines expressing either endogenous *SF3B1-K700E* (SF3B1^{K700E}), *SF3B1-H622Q* (SF3B1^{H622Q}), or *SF3B1-K700K* (SF3B1^{K700K}) was introduced by AAV-mediated homology (provided by H3 Biomedicine) (Darman et al., 2015).

Detection of Activity of *SF3B1* Mutation on CLL Cellular Pathways

Effects of *SF3B1* mutation on CLL cellular pathways were assessed in the isogenic Nalm-6 cell lines or in HEK293T, HeLa, U2OS, K562, HG3, JeKo-1, and MEC2 cells by transiently transfecting or nucleofecting vector control, WT, or mutant *SF3B1* constructs, without or with pathway reporters, in the presence or absence of ligand constructs (Wnt 1, Notch1), depending on the readout. We also interrogated stable cell lines expressing the constitutive and/or altered *DVL2* transcript on OCI-Ly1 cells with or without co-culture with OP9-DL1 cells (Holmes and Zuniga-Pflucker, 2009). See Supplemental Experimental Procedures for more detailed information.

Bulk and Single-Cell RNA-Seq Library Generation and Data Processing

Bulk and single-cell RNA-seq libraries were generated as described previously (Landau et al., 2014). Analysis of alternative splicing was performed using JuncBASE (Brooks et al., 2011) in both bulk and single-cell RNA-seq data. Differential gene expression analysis on bulk samples and single cells was performed using the DESeq2 R package (Love et al., 2014) and SCDE R package (Kharchenko et al., 2014), respectively. The detailed methods regarding total and poly(A) RNA library generation, single-cell RNA-seq, analysis of bulk RNA gene expression, single-cell gene expression, mutation call, and splice variants are provided in the Supplemental Experimental Procedures.

The CLL and normal B cell sequencing data were deposited in the database of Genotypes and Phenotypes (dbGaP: phs000435.v2.p1) and the processed data deposited in GEO (GEO: GSE58889). *SF3B1* codon-optimized cDNA sequences were deposited in GenBank (Genbank: KX881377).

Statistical Analysis

The data in Figures 1F, 2D, 4D, 4E, 7A, and 6D were analyzed using unpaired two-tailed Student's t test. p < 0.05 was considered significant. The data in Figure 8C were analyzed using two-tailed Welch t test.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and nine tables and can be found with this article online at http://dx.doi.org/10.1016/j.ccell.2016.10.005.

AUTHOR CONTRIBUTIONS

L.W., A.N.B., J.F., Y.W., and C.J.W. designed the study. L.W., R.G, K.J.L., S.L., and S.H. performed all the pathway analysis and single-cell study related to *SF3B1* mutation, Y.W., R.G., and W.Z. generated *SF3B1* and *DVL2* constructs, performed splice variants, and *TERC* validation in primary CLL samples and cell lines, and biochemical studies related to incorporation of SF3B1 into the splicing complex. S.Y. performed the telomerase activity detection experiment. D.B.B. performed the immunofluorescence staining of SF3B1 in K562 cells. J.Z.L. and L.F. generated the total RNA libraries, L.Z.R., E.M.G., T.J.K., S.F., and J.R.B. provided the samples. A.N.B. performed the bulk RNA-seq analysis and J.F. performed gene expression and single-cell analyses. M.S. performed branchpoint and alternative splice site-mapping analysis. S.S.F. performed lariat read mapping. D.K., D.A.L., S.S., K.C., C.L.C., and D.S.D. provided computational help. D.N. performed statistical analysis. S. B., P.G.S., J.C.A., R.R., G.G., M.M.M., and P.V.K. provided reagents and constructive suggestions. C.J.W. supervised the study. All co-first authors prepared the manuscript with help from all co-authors.

REFERENCES

Adiconis, X., Borges-Rivera, D., Satija, R., DeLuca, D.S., Busby, M.A., Berlin, A.M., Sivachenko, A., Thompson, D.A., Wysoker, A., Fennell, T., et al. (2013). Comparative analysis of RNA sequencing methods for degraded or low-input samples. Nat. Methods *10*, 623–629.

Alsafadi, S., Houy, A., Battistella, A., Popova, T., Wassef, M., Henry, E., Tirode, F., Constantinou, A., Piperno-Neumann, S., Roman-Roman, S., et al. (2016). Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. Nat. Commun. *7*, 10615.

Bieging, K.T., Mello, S.S., and Attardi, L.D. (2014). Unravelling mechanisms of p53-mediated tumour suppression. Nat. Rev. Cancer *14*, 359–370.

Brooks, A.N., Yang, L., Duff, M.O., Hansen, K.D., Park, J.W., Dudoit, S., Brenner, S.E., and Graveley, B.R. (2011). Conservation of an RNA regulatory map between Drosophila and mammals. Genome Res. *21*, 193–202.

Brooks, A.N., Choi, P.S., de Waal, L., Sharifnia, T., Imielinski, M., Saksena, G., Pedamallu, C.S., Sivachenko, A., Rosenberg, M., Chmielecki, J., et al. (2014). A pan-cancer analysis of transcriptome changes associated with somatic mutations in U2AF1 reveals commonly altered splicing events. PLoS One *9*, e87361.

Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat. Rev. Mol. Cell Biol. *10*, 741–754.

Collu, G.M., Hidalgo-Sastre, A., Acar, A., Bayston, L., Gildea, C., Leverentz, M.K., Mills, C.G., Owens, T.W., Meurette, O., Dorey, K., and Brennan, K. (2012). Dishevelled limits Notch signalling through inhibition of CSL. Development *139*, 4405–4415.

Darman, R.B., Seiler, M., Agrawal, A.A., Lim, K.H., Peng, S., Aird, D., Bailey, S.L., Bhavsar, E.B., Chan, B., Colla, S., et al. (2015). Cancer-associated

SF3B1 hotspot mutations induce cryptic 3' splice site selection through use of a different branch point. Cell Rep. *13*, 1033–1045.

DeBoever, C., Ghia, E.M., Shepard, P.J., Rassenti, L., Barrett, C.L., Jepsen, K., Jamieson, C.H., Carson, D., Kipps, T.J., and Frazer, K.A. (2015). Transcriptome sequencing reveals potential mechanism of cryptic 3' splice site selection in SF3B1-mutated cancers. PLoS Comput. Biol. *11*, e1004105.

Ferreira, P.G., Jares, P., Rico, D., Gomez-Lopez, G., Martinez-Trillos, A., Villamor, N., Ecker, S., Gonzalez-Perez, A., Knowles, D.G., Monlong, J., et al. (2014). Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia. Genome Res. *24*, 212–226.

GTEx Consortium (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science *348*, 648–660.

Holmes, R., and Zuniga-Pflucker, J.C. (2009). The OP9-DL1 system: generation of T-lymphocytes from embryonic or hematopoietic stem cells in vitro. Cold Spring Harb. Protoc. *2009*, http://dx.doi.org/10.1101/pdb.prot5156.

Ilagan, J.O., Ramakrishnan, A., Hayes, B., Murphy, M.E., Zebari, A.S., Bradley, P., and Bradley, R.K. (2015). U2AF1 mutations alter splice site recognition in hematological malignancies. Genome Res. *25*, 14–26.

Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P.V., Mar, B.G., Lindsley, R.C., Mermel, C.H., Burtt, N., Chavez, A., et al. (2014). Age-related clonal hematopoiesis associated with adverse outcomes. N. Engl. J. Med. *371*, 2488–2498.

Jeromin, S., Weissmann, S., Haferlach, C., Dicker, F., Bayer, K., Grossmann, V., Alpermann, T., Roller, A., Kohlmann, A., Haferlach, T., et al. (2014). SF3B1 mutations correlated to cytogenetics and mutations in NOTCH1, FBXW7, MYD88, XPO1 and TP53 in 1160 untreated CLL patients. Leukemia *28*, 108–117.

Kesarwani, A.K., Ramirez, O., Gupta, A.K., Yang, X., Murthy, T., Minella, A.C., and Pillai, M. (2016). Cancer-associated SF3B1 mutants recognize otherwise inaccessible cryptic 3' splice sites within RNA secondary structures. Oncogene *279*, 1–11.

Kharchenko, P.V., Silberstein, L., and Scadden, D.T. (2014). Bayesian approach to single-cell differential expression analysis. Nat. Methods *11*, 740–742.

Kim, E., Ilagan, J.O., Liang, Y., Daubner, G.M., Lee, S.C., Ramakrishnan, A., Li, Y., Chung, Y.R., Micol, J.B., Murphy, M.E., et al. (2015). SRSF2 mutations contribute to myelodysplasia by mutant-specific effects on exon recognition. Cancer Cell *27*, 617–630.

Landau, D.A., Carter, S.L., Stojanov, P., McKenna, A., Stevenson, K., Lawrence, M.S., Sougnez, C., Stewart, C., Sivachenko, A., Wang, L., et al. (2013). Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell *152*, 714–726.

Landau, D.A., Clement, K., Ziller, M.J., Boyle, P., Fan, J., Gu, H., Stevenson, K., Sougnez, C., Wang, L., Li, S., et al. (2014). Locally disordered methylation forms the basis of intra-tumor methylome variation in chronic lymphocytic leukemia. Cancer Cell *26*, 813–825.

Landau, D.A., Tausch, E., Taylor-Weiner, A.N., Stewart, C., Reiter, J.G., Bahlo, J., Kluth, S., Bozic, I., Lawrence, M., Bottcher, S., et al. (2015). Mutations driving CLL and their evolution in progression and relapse. Nature *526*, 525–530.

Levin, J.Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D.A., Friedman, N., Gnirke, A., and Regev, A. (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat. Methods *7*, 709–715.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lu, H., Hu, L., Li, T., Lahiri, S., Shen, C., Wason, M.S., Mukherjee, D., Xie, H., Yu, L., and Zhao, J. (2012). A novel role of Kruppel-like factor 8 in DNA repair in breast cancer cells. J. Biol. Chem. *287*, 43720–43729.

Mercer, T.R., Clark, M.B., Andersen, S.B., Brunck, M.E., Haerty, W., Crawford, J., Taft, R.J., Nielsen, L.K., Dinger, M.E., and Mattick, J.S. (2015). Genome-wide discovery of human splicing branchpoints. Genome Res. *25*, 290–303.

Minoguchi, S., Taniguchi, Y., Kato, H., Okazaki, T., Strobl, L.J., Zimber-Strobl, U., Bornkamm, G.W., and Honjo, T. (1997). RBP-L, a transcription factor related to RBP-Jkappa. Mol. Cell Biol. *17*, 2679–2687.

Park, S.M., Ou, J., Chamberlain, L., Simone, T.M., Yang, H., Virbasius, C.M., Ali, A.M., Zhu, L.J., Mukherjee, S., Raza, A., and Green, M.R. (2016). U2AF35(S34F) promotes transformation by directing aberrant ATG7 Pre-mRNA 3' end formation. Mol. Cell *62*, 479–490.

Puente, X.S., Pinyol, M., Quesada, V., Conde, L., Ordonez, G.R., Villamor, N., Escaramis, G., Jares, P., Bea, S., Gonzalez-Diaz, M., et al. (2011). Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature *475*, 101–105.

Puente, X.S., Bea, S., Valdes-Mas, R., Villamor, N., Gutierrez-Abril, J., Martin-Subero, J.I., Munar, M., Rubio-Perez, C., Jares, P., Aymerich, M., et al. (2015). Non-coding recurrent mutations in chronic lymphocytic leukaemia. Nature *526*, 519–524.

Qi, X., Rand, D.P., Podlevsky, J.D., Li, Y., Mosig, A., Stadler, P.F., and Chen, J.J. (2015). Prevalent and distinct spliceosomal 3'-end processing mechanisms for fungal telomerase RNA. Nat. Commun. *6*, 6105.

Quesada, V., Conde, L., Villamor, N., Ordonez, G.R., Jares, P., Bassaganyas, L., Ramsay, A.J., Bea, S., Pinyol, M., Martinez-Trillos, A., et al. (2012). Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nat. Genet. *44*, 47–52.

Rosati, E., Sabatini, R., Rampino, G., Tabilio, A., Di Ianni, M., Fettucciari, K., Bartoli, A., Coaccioli, S., Screpanti, I., and Marconi, P. (2009). Constitutively activated Notch signaling is involved in survival and apoptosis resistance of B-CLL cells. Blood *113*, 856–865.

Rossi, D., Bruscaggin, A., Spina, V., Rasi, S., Khiabanian, H., Messina, M., Fangazio, M., Vaisitti, T., Monti, S., Chiaretti, S., et al. (2011). Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. Blood *118*, 6904–6908.

Shirai, C.L., Ley, J.N., White, B.S., Kim, S., Tibbitts, J., Shao, J., Ndonwi, M., Wadugu, B., Duncavage, E.J., Okeyo-Owuor, T., et al. (2015). Mutant U2AF1 expression alters hematopoiesis and Pre-mRNA splicing in vivo. Cancer Cell *27*, 631–643.

Taggart, A.J., DeSimone, A.M., Shih, J.S., Filloux, M.E., and Fairbrother, W.G. (2012). Large-scale mapping of branchpoints in human pre-mRNA transcripts in vivo. Nat. Struct. Mol. Biol. *19*, 719–721.

Te Raa, G.D., Derks, I.A., Navrkalova, V., Skowronska, A., Moerland, P.D., van Laar, J., Oldreive, C., Monsuur, H., Trbusek, M., Malcikova, J., et al. (2015). The impact of SF3B1 mutations in CLL on the DNA-damage response. Leukemia *29*, 1133–1142.

Wang, X., and Zhao, J. (2007). KLF8 transcription factor participates in oncogenic transformation. Oncogene *26*, 456–461.

Wang, C., Chua, K., Seghezzi, W., Lees, E., Gozani, O., and Reed, R. (1998). Phosphorylation of spliceosomal protein SAP 155 coupled with splicing catalysis. Genes Dev. *12*, 1409–1414.

Wang, L., Lawrence, M.S., Wan, Y., Stojanov, P., Sougnez, C., Stevenson, K., Werner, L., Sivachenko, A., DeLuca, D.S., Zhang, L., et al. (2011). SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. N. Engl. J. Med. *365*, 2497–2506.

Weng, N.P., Palmer, L.D., Levine, B.L., Lane, H.C., June, C.H., and Hodes, R.J. (1997). Tales of tails: regulation of telomere length and telomerase activity during lymphocyte development, differentiation, activation, and aging. Immunol. Rev. *160*, 43–54.

Xie, M., Lu, C., Wang, J., McLellan, M.D., Johnson, K.J., Wendl, M.C., McMichael, J.F., Schmidt, H.K., Yellapantula, V., Miller, C.A., et al. (2014). Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat. Med. *20*, 1472–1478.

Zhang, J., Lieu, Y.K., Ali, A.M., Penson, A., Reggio, K.S., Rabadan, R., Raza, A., Mukherjee, S., and Manley, J.L. (2015). Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. Proc. Natl. Acad. Sci. USA *112*, E4726–E4734.

Zweidler-McKay, P.A., He, Y., Xu, L., Rodriguez, C.G., Karnell, F.G., Carpenter, A.C., Aster, J.C., Allman, D., and Pear, W.S. (2005). Notch signaling is a potent inducer of growth arrest and apoptosis in a wide range of B-cell malignancies. Blood *106*, 3898–3906.

Supplemental Information

Transcriptomic Characterization of *SF3B1* Mutation

Reveals Its Pleiotropic Effects in Chronic

Lymphocytic Leukemia

Lili Wang, Angela N. Brooks, Jean Fan, Youzhong Wan, Rutendo Gambe, Shuqiang Li, Sarah Hergert, Shanye Yin, Samuel S. Freeman, Joshua Z. Levin, Lin Fan, Michael Seiler, Silvia Buonamici, Peter G. Smith, Kevin F. Chau, Carrie L. Cibulskis, Wandi Zhang, Laura Z. Rassenti, Emanuela M. Ghia, Thomas J. Kipps, Stacey Fernandes, Donald B. Bloch, Dylan Kotliar, Dan A. Landau, Sachet A. Shukla, Jon C. Aster, Robin Reed, David S. DeLuca, Jennifer R. Brown, Donna Neuberg, Gad Getz, Kenneth J. Livak, Matthew M. Meyerson, Peter V. Kharchenko, and Catherine J. Wu

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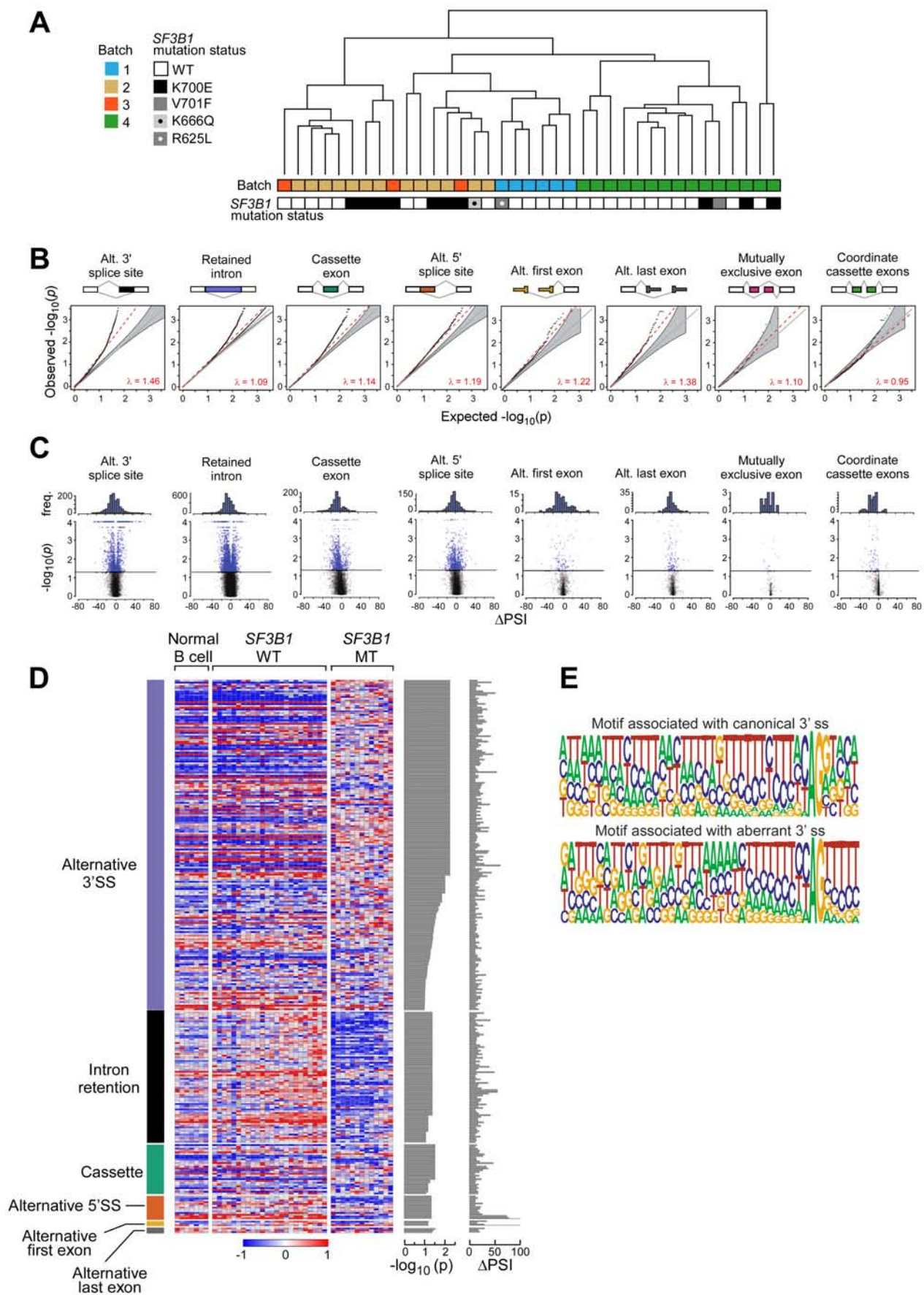
Figure S1, related to Figure 1. Batch effect and alternative splicing characterization from 37 samples used in the study.

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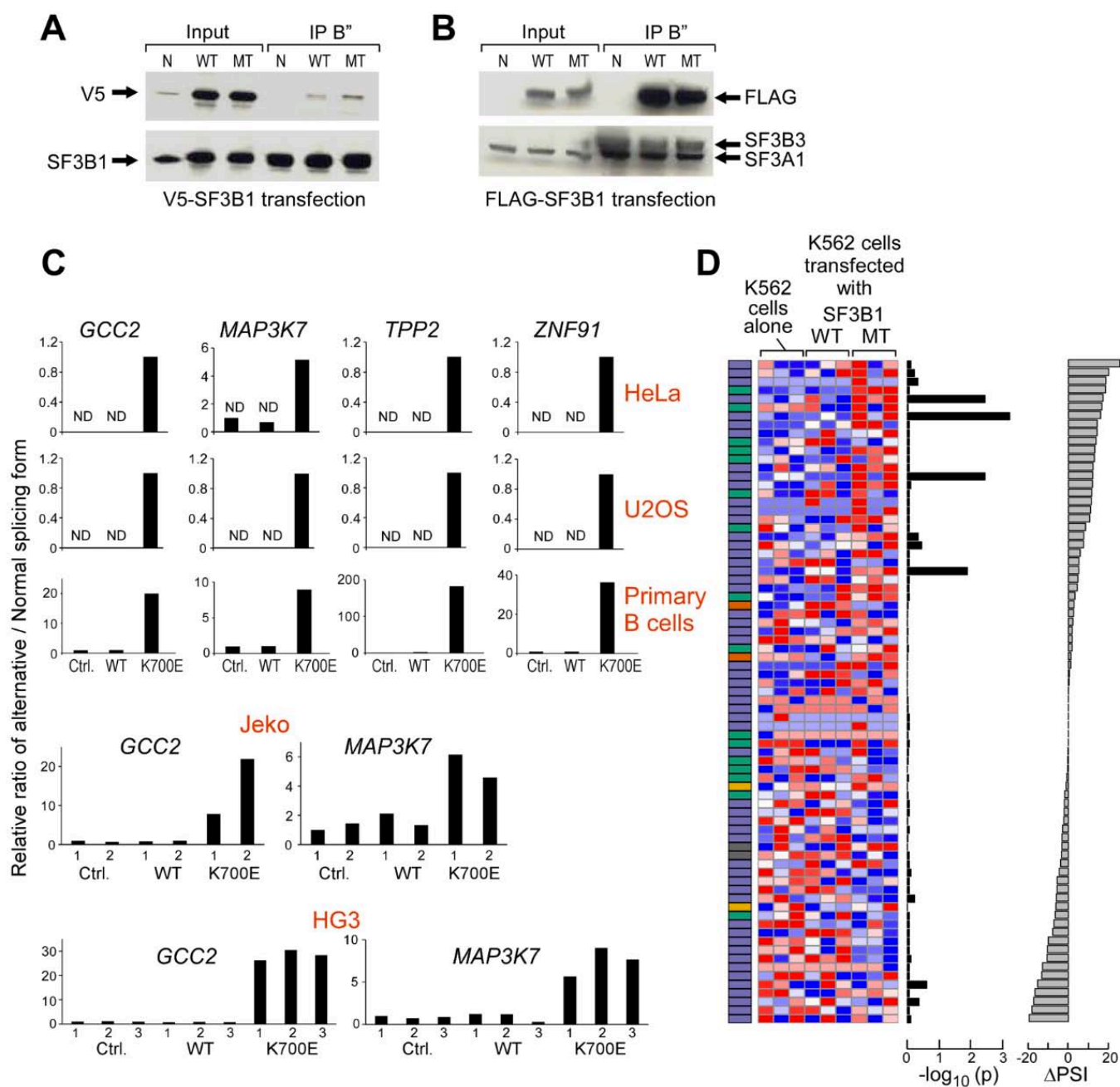
Figure S2, related to Figure 2. Expression of SF3B1-K700E using N-terminal-tagged SF3B1 construct in different cell lines is associated with altered splicing.

(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.

3

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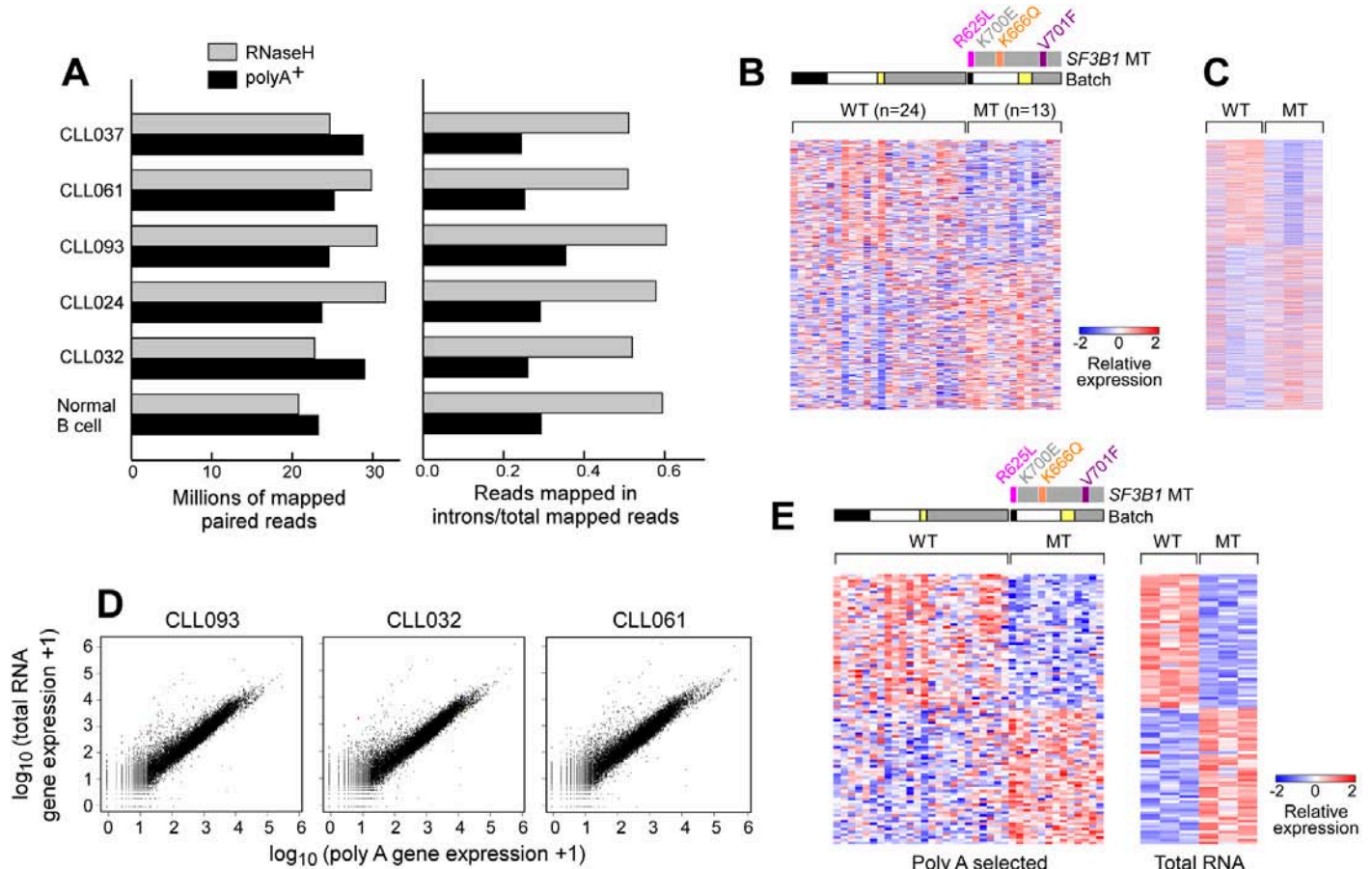
Figure S3, related to Figure 4. Comparisons between poly-A selected and total RNA libraries.

(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.
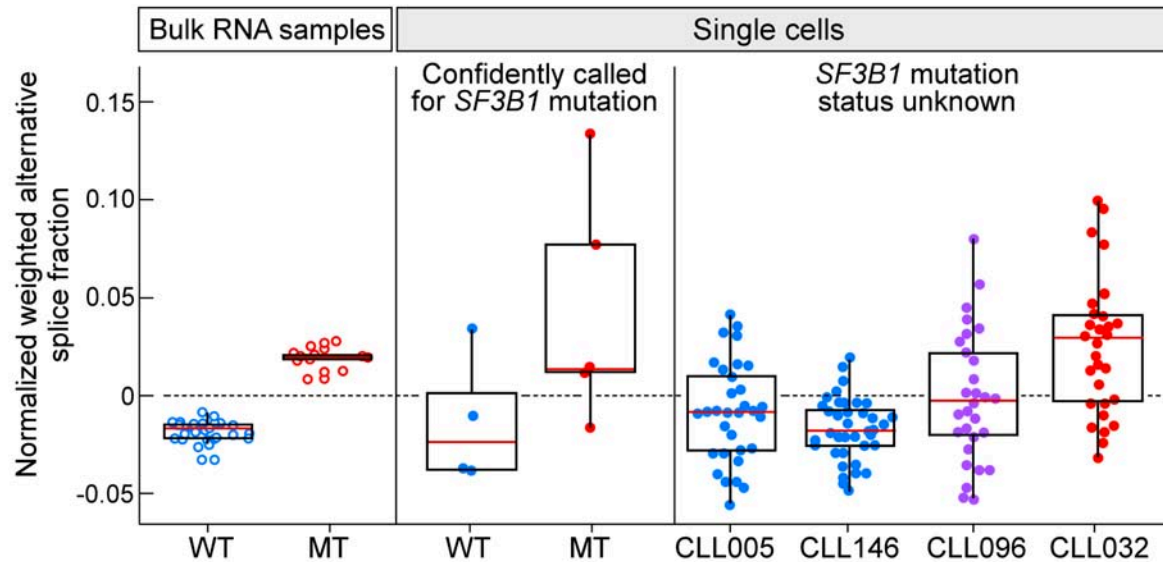

Figure S4, related to Figure 5. Inference of single cell *SF3B1* mutation status by splice variants.

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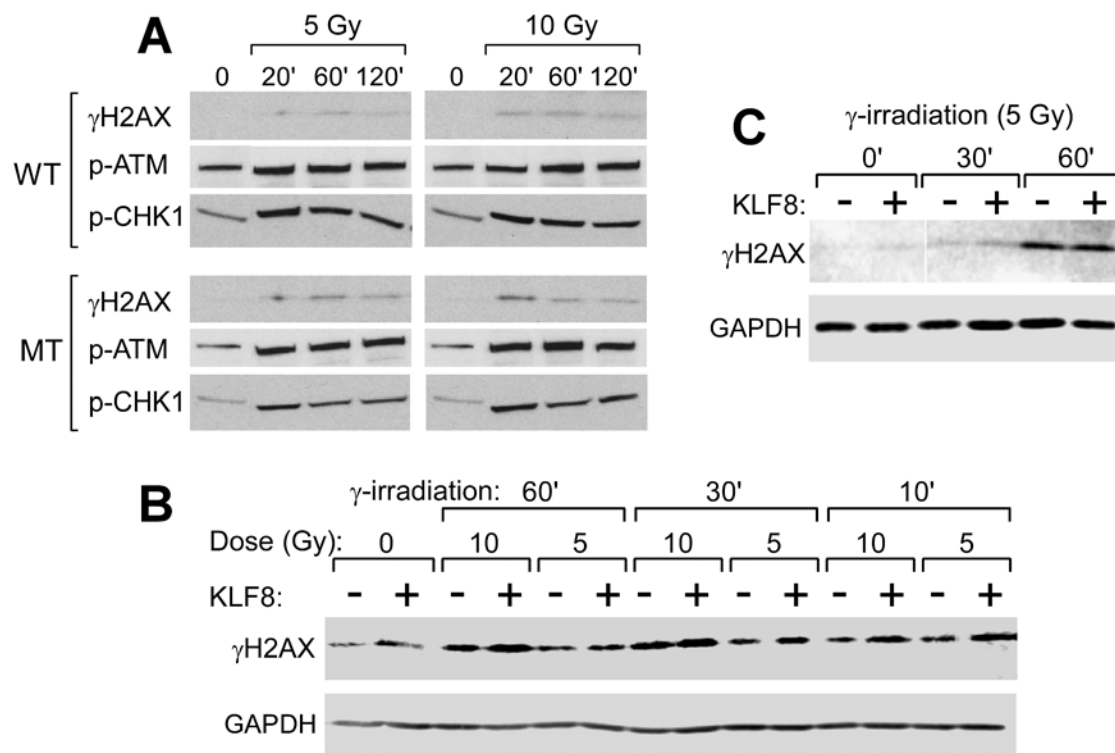
Figure S5, related to Figure 6. Effects of *SF3B1* mutation DNA damage pathway.

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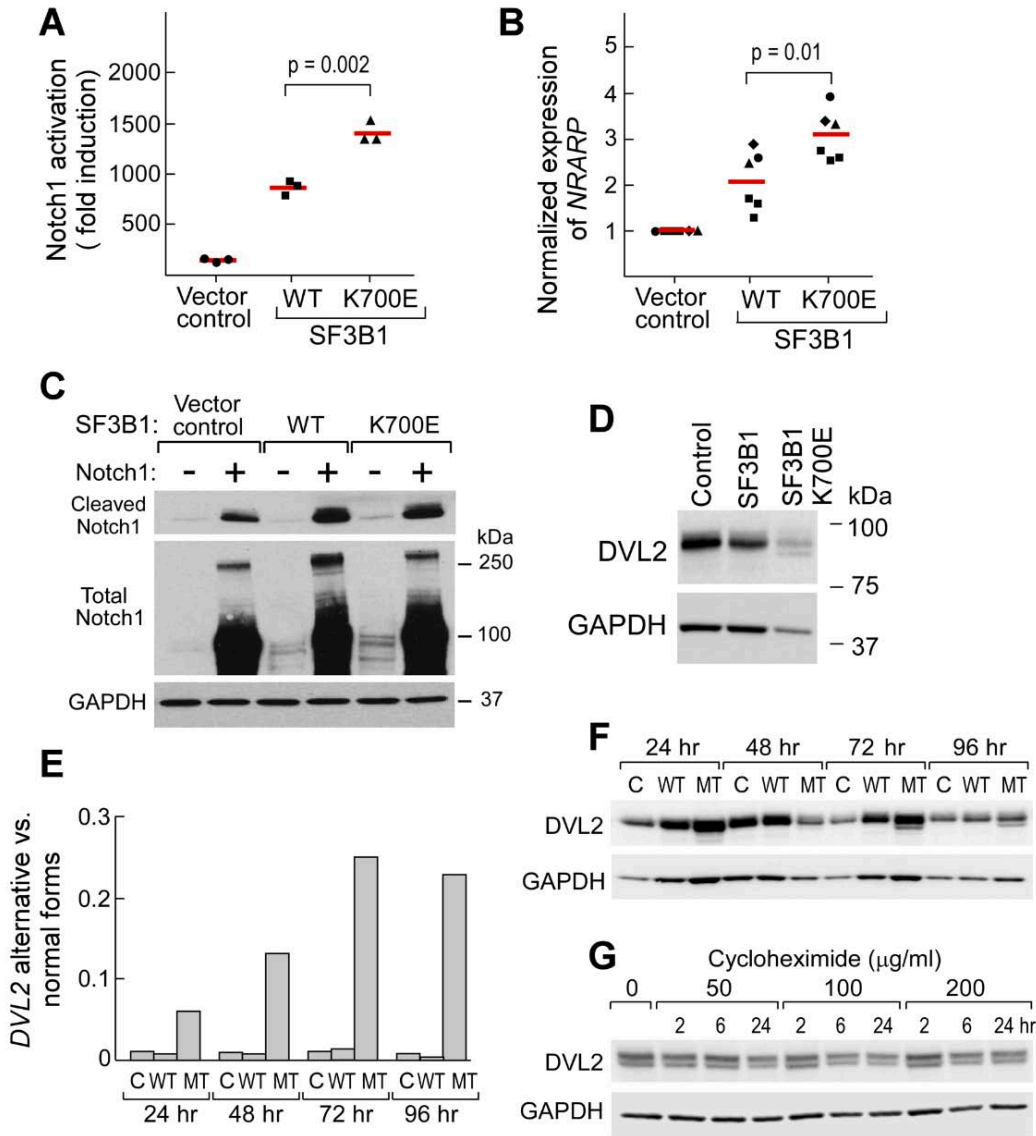
Figure S6, related to Figures 7. *SF3B1* **mutation impacts Notch signaling and** *DVL2* **splicing.**

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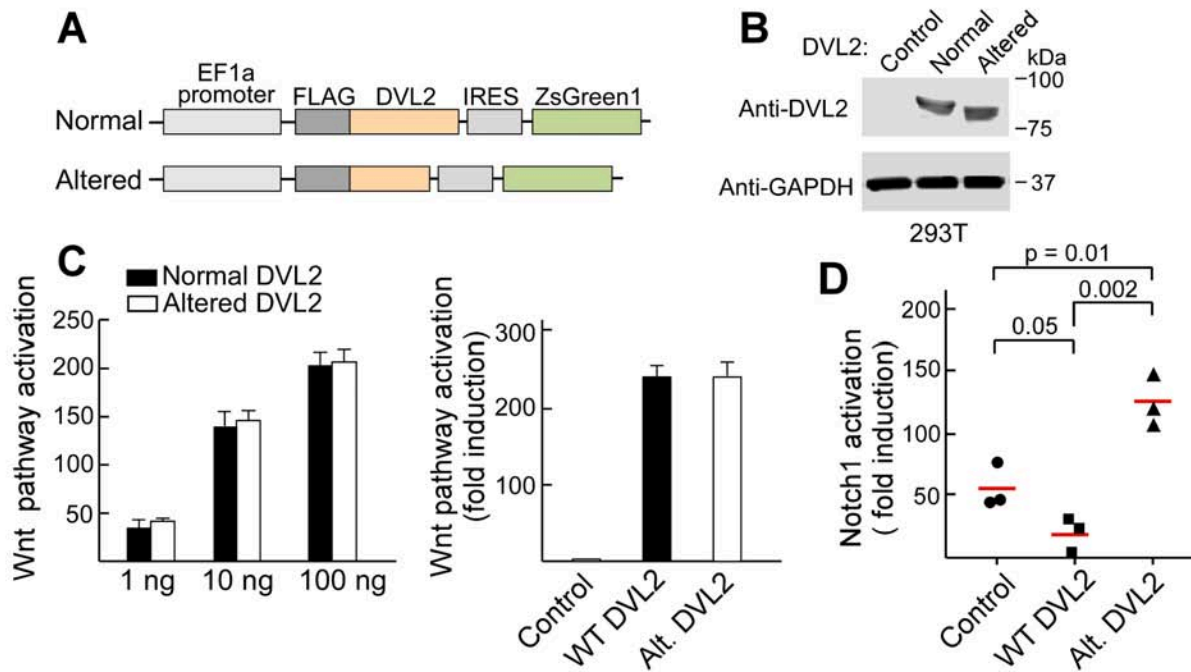
Figure S7, related to Figure 8. *SF3B1* **mutation impacts Notch signaling through splicing of** *DVL2***.**

(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 $< 1 \times 10^6$, mode mapping quality = 0, passed filter aligned bases $< 1.5 \times 10^8$) 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 $\gamma-{}^{32}P-ATP$, 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

14

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.