Transcriptome analysis of chronic lymphoid leukemia reveals isoform regulation associated with mutations in *SF3B1*

Alejandro Reyes[1], Carolin Blume[2], Vicent Pelechano[1], Petra Jakob[1], Lars M Steinmetz[1,3], Thorsten Zenz[2,4], Wolfgang Huber[1,*]

1 European Molecular Biology Laboratory, Genome Biology Unit, 69117 Heidelberg, Germany

2 Department of Translational Oncology, National Centre for Tumour Diseases and German Cancer Research Centre, Heidelberg, Germany

3 Stanford Genome Technology Center, Stanford University, Palo Alto, CA 94304

4 Department of Medicine V, University Hospital Heidelberg, Heidelberg, Germany

∗ To whom correspondence should be addressed. E-mail: whuber@embl.de

Abstract

Genome sequence studies of chronic lympoid leukemia (CLL) have provided a comprehensive overview of recurrent somatic mutations in coding genes. One of the most intriguing discoveries has been the prevalence of mutations in the HEAT-repeat domain of the splicing factor *SF3B1*. A frequently observed variant is predicted to cause the substitution of a lysine with a glutamic acid at position 700 of the protein (K700E). However, the molecular consequences of the mutations and their contribution to the cancer phenotype are largely unknown. We aimed to investigate the consequences of the K700E mutation in *SF3B1* on transcript isoform regulation in CLL. We sequenced the transcriptomes of six samples: four CLL tumour cells, of which two contained the K700E mutation in *SF3B1*, and CD19 positive cells from two healthy donors. We identified 41 genes that showed differential usage of exons associated with the mutated status of *SF3B1* (false discovery rate 10%). These genes were enriched in pathways including "interferon signaling" and "mRNA splicing". Additionally, we found evidence of differential exon usage of the genes *UQCC* and *RPL31* as a consequence of mutations in *SF3B1* in our CLL data; notably, a similar effect on these genes was described in a previously published study of uveal melanoma. These data provide an initial view of transcript isoform consequences of the SF3B1 (K700E) mutation in CLL, some of which might contribute to the tumourigenesis of CLL. Studies of larger cohorts and model systems are required to extend these findings.

Introduction

Several DNA sequencing studies of chronic lymphocytic leukemia (CLL) revealed that the splicing factor *SF3B1* accumulated somatic point mutations in about 10% of the patients [1, 2]. In most cases the mutations were located in the genomic regions coding for the C-terminal HEAT-repeat domain and in many cases, the mutations gave rise to specific amino acid substitutions. For instance, the substitution of a lysine to a glutamic acid in the amino acid 700 of the protein (K700E) was prevalent in the tumour cells. In addition, the affected amino acids seemed to be clustered spatially in the 3D structure of the protein. These observations suggest that specific changes to the function of the protein could be one of the main drivers of tumour progression in CLL. Additionally, DNA sequencing studies have found recurrent mutations in *SF3B1* in other malignancies, including myelodysplasia (with high incidence in a particular subgroup, RARS) [3, 4] and uveal melanomas [5, 6].

mRNA splicing is the process in which introns are removed from pre-mRNA molecules in order to produce fully mature transcripts. A crucial step of splicing is the recruitment of U2 small nucleolar ribonucleic particle (U2 snRNP) to the branch point sequence: this results in the base pairing between U2 snRNP and the pre-mRNA that allows the first chemical reaction of splicing to occur [7]. This recruitment is preceded by the binding of the protein A2AF to the pyrimidine tract and subsequent

recruitment of SF3b 155 (the protein encoded by *SF3B1*) [8,9]. In fact, either blocking the interaction of SF3b 155 to the pre-mRNA sequences using the anti-tumour drug spliceostatin A (SSA) or the knockdown of *SF3B1* results in unstable recruitment of U2 snRNP, which leads to changes in alternative splicing [10].

Additional studies have shown the relevance of *SF3B1* in the regulation of splicing in different biological contexts. For instance, it has been shown that the interaction between SF3b 155 and the proteins Xfp144 and Rnf2 from the Polycomb group of genes is required for the repression of Hox genes during mouse development [11]. In a similar manner, it was shown that the loss of interaction between the proteins coded by the genes *PQBP1* and *SF3B1* alters alternative splicing in mouse neurons and leads to neurite outgrowth defects [9]. These lines of evidence suggest that the gene *SF3B1* is necessary for the correct splicing of pre-mRNAs.

The presence of mutations in *SF3B1* is correlated with adverse prognosis and shorter survival of CLL patients [12]. But despite their usefulness as clinical markers, the functional consequences of the mutations in *SF3B1* are presently not well understood. It has been hypothesized that the mutations in the HEAT-repeat domain might affect the interaction of SF3b 155 with other co-factors and thus, splicing fidelity. Consistent with that hypothesis, it has been observed that mutations in *SF3B1* are associated with the activation of abnormal 3' acceptor sites of specific genes in CLL tumour cells [1]. In a similar manner, transcriptome analyses of myelodysplastic syndromes and uveal melanomas have identified sets of genes with differential exon usage between tumours with mutations in *SF3B1* and tumours with no mutations in this gene [6,13]. To date, transcriptome analyses have not led to a characterisation of global changes in splicing patterns, but they have pointed to localized splicing differences affecting specific genes [14].

Here, we aimed to investigate differential usage of exons associated with the mutation K700E of *SF3B1* in CLL tumuor cells. We generated RNA-Seq transcriptome data from cells of two tumours with mutations in *SF3B1*, two tumours without mutations in *SF3B1* and from cells from two healthy donors. We identified differences in isoform regulation that were associated with the *SF3B1* mutation in 41 genes. We report interesting examples and possible consequences of the observed altered exon usage pattern in these genes. We compare our results to previous transcriptome analyses of myelodysplastic syndromes and uveal melanomas with mutations in *SF3B1*.

Results

Transcriptome-wide data reveal isoform regulation associated with mutations in *SF3B1* in CLL tumour cells

We isolated RNA from B-CLL cells of four patients, of which two contained mutations in the *SF3B1* gene (predicted to lead to the K700E substitution in the protein), and two had no mutation in *SF3B1* (as confirmed by PCR). In addition, we extracted RNA from CD19-purified cells isolated from the peripheral blood of two healthy donors (see Table S1 for detailed information regarding the samples). We used Illumina HiSeq 2000 to sequence 50 nt paired-end reads using a strand-specific protocol and obtained a total of 275,000,664 sequenced fragments. We mapped these read fragments to the human reference genome (*ENSEMBL* release 68) using *GSNAP* (version 2013-05-09), allowing split alignments for exon-exon junctions [15]. We considered only uniquely mapped fragments for further analysis. In order to observe the expression of the *SF3B1* alleles, we counted the number of mapped fragments in each sample supporting the evidence for the mutation. Based on this, we estimated that when the mutation was present, around half of the transcripts were transcribed from the variant allele (Figure 1). This estimation was consistent with the variant heterogeneity quantification of the tumour DNA, as assessed by 454 genomic sequencing (also around 50%, Table S1), and consistent with no allelic preference of gene expression.

We asked, transcriptome-wide, whether specific differences in isoform regulation in the CLL tumour

cells were associated with the expression of the *SF3B1* K700E variant. Therefore, we used *DEXSeq* to test for differences in exon usage (DEU) [16] between the tumour cells with the mutation K700E in *SF3B1* compared to the tumour cells without the mutation and the healthy donors. Briefly, *DEXSeq* considers, for each exon, the ratio between the number of transcripts originating from the gene that contain the exon and the number of all transcripts originating from the gene. This allowed us to identify changes in relative exon usage independently from the fact that a gene could be differentially expressed. Using this approach, we identified a set of 50 exons in 41 genes with DEU at a false disovery rate (FDR) of 10%.

To explore the functions of the genes whose isoform regulation was associated with the mutant *SF3B1* samples, we mapped these genes to pathways annotated in *REACTOME* [17]. We found a statistically significant over-representation, compared to a background set of genes that were also expressed in these cells, of pathways including "mRNA splicing" and "translation" at a false discovery rate of 0.1 (see Table S2). Interestingly, we also found a significant over-representation of the "interferon signaling" pathway, which is known to inhibit cell proliferation and whose aberrant regulation has been linked to aggressive cases of CLL [18]. These enrichments suggest that the mutations of *SF3B1* could be altering the isoform regulation of genes in specific pathways.

The *SF3B1* mutation is associated with differential exon usage patterns seen both in uveal melanoma and CLL

Next we compared our results with those of two previously published transcriptomes. Notably, these studies used the same sequencing technology, had a similar study design (but in different malignancies) and also used the *DEXSeq* method to test for differences in exon usage.

Furney et al. [6] compared three *SF3B1* mutant and nine *SF3B1* wildtype uveal melanoma tumours and identified 34 exons differentially used in 21 genes (10% FDR). Remarkably, we found a significant overlap between their list of genes and our list of genes (p-value = $2.1 \cdot 10^{-3}$, Fishers' exact test). Specifically, the genes *UQCC* and *RPL31* overlapped with our hits. Furthermore, one out of the two regions with DEU that they reported in the gene *RPL31* was also seen as differentially used in our data (Figure 2). Additionally, three out of the four exons that we detected as significant for the chaperone *UQCC* were also detected to be differentially used in the uveal melanoma study. Its authors reported a decrease in the expression of the 3' end of this gene in uveal melanomas with mutated *SF3B1*, and we observed the same in the CLL tumour cells (Figure 3). Interestingly, this region partly codes for a chaperone domain that is conserved with yeast, where it appears to be required for the assembly of the protein ubiquinol-cytochrome C reductase [19]. In humans, genome-wide association studies have linked this gene to body growth [20]. To summarize this comparison, we found differentially used exons associated with mutations in *SF3B1* in the genes *UQCC* and *RPL31* that are present both in CLL tumour cells and in uveal melanomas. The recurrence of these phenomena in these different biological contexts suggest a strong link between the mutations in *SF3B1* and these effects, which could merit follow-up study.

Visconte et al. reported 423 exons in 350 genes to be differentially used at a FDR of 5% between myelodysplasia patients with mutations in *SF3B1* and one healthy donor [13]. However, the overlap of their list of genes with our list of genes was not larger than what would be expected by chance, and no common pattern was apparent.

Further exploration of results.

We generated a report with plots including those shown in Figures 2 and 3 as a resource to aid the exploration of our results (Supplementary Dataset S1). By exploring this resource, we detected that part of the 5' untranslated region of the gene *FAIM3*, a gene with anti-apoptotic functions in blood cells [21], was included more frequently in transcripts from the mutant samples compared to the tumour cells with wildtype *SF3B1* and the healthy donor cells (Figure 4). We also identified that a region of the

gene *NFAT5*, a transcription factor that was previously shown to be important for normal lymphocyte proliferation [22], was used less in the samples with mutations in *SF3B1* (Figure 5). This region forms part of a transcript that is subject to nonsense mediated decay (as annotated by *ENSEMBL*), hence, a possible interpretation of this result is that in the mutant samples *NFAT5* is subject to nonsense mediated decay less frequently than in tumour cells with wildtype *SF3B1* and in normal cells.

Discussion

To explore the effects on splicing of the expression of mutant *SF3B1* in CLL tumour cells, we generated transcriptome data from two CLL patients harboring the K700E mutation in SF3B1, two CLL patients without the mutation, as well as two healthy donor cells. Our results provide an initial list of DEU events that appear associated with the K700E mutation in *SF3B1* in CLL (see Supporting Dataset S1). Our data rely on a very limited sample of tumours; substantially larger cohorts (e. g. tens of tumours with and without the mutation) will be needed for a more reliable, more comprehensive list of events. A notable result of our analysis is the overlap of events seen here with those in a previous study of uveal melanomas [6], namely, differences in the usage of specific exonic regions of the genes *UQCC* and *RPL31*. These could be a prevalent consequence of the mutations in the HEAT-repeat domain of *SF3B1*. The question of whether or not these gene regulation consequences play a causal role in the tumorigenesis is not directly addressed by our data, but may merit further study.

Materials and Methods

Ethics Statement

Samples were acquired by informed written consent in accordance with the Declaration of Helsinki. Ethical and Institutional Board Review (IRB) approvals were obtained from the University Hospital of Heidelberg.

Sample preparation

Peripheral blood samples from four patients matching standard diagnostic criteria for CLL and featuring a high lymphocyte percentage (median: 98%) were obtained from the University Hospital of Heidelberg. Mononuclear cells (MNCs) were isolated by centrifugation over Ficoll-Paque Premium (GE healthcare, Freiburg). MNCs from buffy coats of healthy donors obtained from the blood bank of the University Hospital of Heidelberg were further CD19-purified by magnetic activated cell sorting (MACS) according to the manufacturer's instructions (Miltenyi Biotech, Bergisch Gladbach) resulting in purities of $\geq 95\%$ CD19+ cells. Clinical and laboratory data are summarized in Supporting Table S1. Exons of *SF3B1*, *TP53*, *BRAF*, *MYD88* and *NOTCH1* containing mutation hot spots were amplified and subjected to next-generation sequencing on the GS Junior 454 platform (Roche, Penzberg) as in [23]. Mutations in SF3B1 were confirmed by conventional Sanger sequencing.

Strand-specific RNA-Seq library preparation

Total RNA was isolated from $1 \cdot 10^8$ to $5 \cdot 10^8$ cells (depending on the sample) via standard trizol extraction. Strand specific RNA-Seq libraries were prepared as described in [24]. Briefly, polyadenylated RNA was isolated from 10 g of total RNA using Dynabeads Oligo (dT)25 (Invitrogen) according to the manufacturer's protocol. The poly(A) enriched RNA was fragmented by incubating the samples at 80°C for 4 minutes in the presence of RNA fragmentation buffer (40 mM Tris-acetate, pH 8.1, 100 mM KOAc, 30 mM MgOAc). The fragmented RNA was purified using 1.8X (v/v) Ampure XP Beads

(Beckman Coulter Genomics) and eluted in 25 μl Elution Buffer (EB) (10 mM Tris-HCl, pH 8) according to manufacturer's protocol. 24 μl of eluted RNA was reverse transcribed using 1 μl of random hexamers (30 ng/μl, Invitrogen). The samples were denatured at 70°C for 5 minutes and transferred to ice. Two μl dNTPs (10 mM), 8 μl 5X first strand buffer (Invitrogen), 4 μl DTT (0.1 M), 0.5 μl actinomycin D (1.25 mg/μl) and 0.5 μl RNaseOut (40 U/μl, Invitrogen) were added to each sample, and the samples were then incubated at 25°C for 2 minutes. Following this, 0.5 μl Superscript III reverse transcriptase (200 U/μl, Invitrogen) was added. The retrotranscription was carried out at 25°C for 10 minutes, at 55°C for 60 minutes, and inactivated at 75°C for 15 minutes. The samples were purified using 1.8X of Ampure XP beads and eluted in 20 μl EB. For producing the second cDNA strand, 19 μl of sample was mixed with 2.5 μl of 10x NEBNext Second Strand Synthesis (dNTP-free) Reaction buffer (NEB), 1.5 μl of dNTPs (containing dUTPs instead of dTTPs, 10 mM), 0.5 μl of RNaseH (10,000 U/ml) and 0.5 μl of E.coli DNA polymerase I (10 U/l, Fermentas). The samples were incubated at 16°C for 2.5 hours, 80°C for 20 minutes and purified with 1.8X Ampure XP beads, and eluted in 17 μl EB. Two μl end repair buffer and 1 μl end repair enzyme mix (NEBNext DNA Sample Prep Master Mix Set 1, NEB) were added, and the samples were incubated at 20°C for 30 minutes. The samples were purified using 1.8x Ampure XP and resuspended in 17 μl EB. Two μl dA tailing buffer (10X NEBuffer 2 from NEB and 0.2 mM dATP) and 1 μl Klenow Fragment 3' - 5' exo (5 U/μl, NEB) were added and the samples incubated at 37°C for 30 minutes. The samples were purified using 1.8x Ampure XP and resuspended in 20 μl EB. 2.5 μl 10X T4 DNA ligase buffer (NEB), 0.5 μl multiplexed PE Illumina adaptors (7 μM, Supporting Table S3) and 2 μl T4 DNA ligase were added (2000 U/μl, NEB) and incubated at 16°C for 1h. The dUTPs of the second strand were hydrolyzed by incubating the samples at 37°C for 15 min with 1 μl USER enzyme (1 U/μl, NEB) and 5 minutes at 95°C. The samples were purified using 0.9 X Ampure XP beads and eluted in 11 μl EB. Enrichment PCR was performed using 5 μl of sample, 25 μl Phusion Master Mix 2x (NEB), 0.5 μl each of oligos PE1.0 and PE2.0 (10 μM, Illumina) and water up to 50μl final. The PCR program was 30 seconds at 98°C, 15 cycles of (10 seconds at 98°C, 30 seconds at 65°C and 30 seconds at 72°C) and 5 minutes at 72°C. The PCR product was size-selected (average of 290bp), and the libraries were submitted for Illumina sequencing.

Bioinformatics

We mapped the read fragments to the human reference genome from *ENSEMBL* (release 68) using GSNAP (version 2013-05-09) [15, 25]. For each sample, we tabulated the number of uniquely aligned fragments that overlapped with exon annotations from *ENSEMBL* release 68 using scripts based on the python HTSeq library (`http://www-huber.embl.de/users/anders/HTSeq`). We used the generalized linear model framework implemented in *DEXSeq* version 1.9.1 to test for differences in exon usage between the samples containing the mutations in *SF3B1* and the wild type allele samples [16].

In order to avoid biases associated to gene expression strength in further enrichment analysis, we generated a background set of genes that contained at least 600 sequenced fragment counts. We mapped the *ENSEMBL* gene identificators to pathways annotated in *REACTOME* [17] and tested for over-representation of our hits compared to the background using Fisher's exact test. We corrected for multiple testing using the method of Benjamini and Hochberg [26]. We used the *ENSEMBL Perl API* to convert protein domain coordinates annotated in *PFAM* to genomic coordinates [27]. Genomic ranges operations were performed using the Bioconductor package *GenomicRanges* [28], and visualizations of the genomic ranges were done using *ggbio* [29]. We visualized the coverage vectors and the expression of variants of *SF3B1* using the Bioconductor package *h5vc* (`http://www.bioconductor.org/packages/2.14/bioc/html/h5vc.html`). We provide Supporting File S1 with a documented *R* session with the code that was used to analyse the RNA-Seq data and to produce the figures. The RNA count data are available in the ArrayExpress database (`www.ebi.ac.uk/arrayexpress`) under accession number E-MTAB-2025 and in the Bioconductor data package *CLL.SF3B1*.

Acknowledgments

We would like to thank EMBL's Genomics Core Facility for the RNA sequencing service and the Information Technology (IT) Core Facility for provision of computational infrastructure. W.H. acknowledges funding from the European Commission through the Collaborative Research Project *Radiant*.

References

1. Quesada V, Conde L, Villamor N, Ordonez GR, Jares P, et al. (2012) Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nat Genet 44: 47–52.

2. Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, et al. (2011) SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. N Engl J Med 365: 2497–2506.

3. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, et al. (2011) Frequent pathway mutations of splicing machinery in myelodysplasia. Nature 478: 64–69.

4. Papaemmanuil E, Cazzola M, Boultwood J, Malcovati L, Vyas P, et al. (2011) Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. N Engl J Med 365: 1384–1395.

5. Harbour JW, Roberson ED, Anbunathan H, Onken MD, Worley LA, et al. (2013) Recurrent mutations at codon 625 of the splicing factor SF3B1 in uveal melanoma. Nat Genet 45: 133–135.

6. Furney SJ, Pedersen M, Gentien D, Dumont AG, Rapinat A, et al. (2013) SF3B1 mutations are associated with alternative splicing in uveal melanoma. Cancer Discov .

7. Newby MI, Greenbaum NL (2001) A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture. RNA 7: 833–845.

8. Gozani O, Potashkin J, Reed R (1998) A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. Mol Cell Biol 18: 4752–4760.

9. Wang Q, Moore MJ, Adelmant G, Marto JA, Silver PA (2013) PQBP1, a factor linked to intellectual disability, affects alternative splicing associated with neurite outgrowth. Genes Dev 27: 615–626.

10. Corrionero A, Minana B, Valcarcel J (2011) Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A. Genes Dev 25: 445–459.

11. Isono K, Mizutani-Koseki Y, Komori T, Schmidt-Zachmann MS, Koseki H (2005) Mammalian polycomb-mediated repression of Hox genes requires the essential spliceosomal protein Sf3b1. Genes Dev 19: 536–541.

12. Oscier DG, Rose-Zerilli MJ, Winkelmann N, Gonzalez de Castro D, Gomez B, et al. (2013) The clinical significance of NOTCH1 and SF3B1 mutations in the UK LRF CLL4 trial. Blood 121: 468–475.

13. Visconte V, Rogers HJ, Singh J, Barnard J, Bupathi M, et al. (2012) SF3B1 haploinsufficiency leads to formation of ring sideroblasts in myelodysplastic syndromes. Blood 120: 3173–3186.

14. Quesada V, Ramsay AJ, Rodriguez D, Puente XS, Campo E, et al. (2013) The genomic landscape of chronic lymphocytic leukemia: clinical implications. BMC Med 11: 124.

15. Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26: 873–881.

16. Anders S, Reyes A, Huber W (2012) Detecting differential usage of exons from RNA-seq data. Genome Res 22: 2008–2017.

17. Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, et al. (2011) Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res 39: D691–697.

18. Tomic J, Lichty B, Spaner DE (2011) Aberrant interferon-signaling is associated with aggressive chronic lymphocytic leukemia. Blood 117: 2668–2680.

19. Shi G, Crivellone MD, Edderkaoui B (2001) Identification of functional regions of Cbp3p, an enzyme-specific chaperone required for the assembly of ubiquinol-cytochrome c reductase in yeast mitochondria. Biochim Biophys Acta 1506: 103–116.

20. Sanna S, Jackson AU, Nagaraja R, Willer CJ, Chen WM, et al. (2008) Common variants in the GDF5-UQCC region are associated with variation in human height. Nat Genet 40: 198–203.

21. Nguyen XH, Lang PA, Lang KS, Adam D, Fattakhova G, et al. (2011) Toso regulates the balance between apoptotic and nonapoptotic death receptor signaling by facilitating RIP1 ubiquitination. Blood 118: 598–608.

22. Go WY, Liu X, Roti MA, Liu F, Ho SN (2004) NFAT5/TonEBP mutant mice define osmotic stress as a critical feature of the lymphoid microenvironment. Proc Natl Acad Sci USA 101: 10673–10678.

23. Hullein J, Jethwa A, Stolz T, Blume C, Sellner L, et al. (2013) Next-generation sequencing of cancer consensus genes in lymphoma. Leuk Lymphoma 54: 1831–1835.

24. Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, et al. (2009) Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic Acids Res 37: e123.

25. Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, et al. (2013) Ensembl 2013. Nucleic Acids Res 41: 48–55.

26. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological) : 289–300.

27. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, et al. (2012) The Pfam protein families database. Nucleic Acids Res 40: 290–301.

28. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, et al. (2013) Software for computing and annotating genomic ranges. PLoS Comput Biol 9: e1003118.

29. Yin T, Cook D, Lawrence M (2012) ggbio: an R package for extending the grammar of graphics for genomic data. Genome Biology 13: R77.

Figure Legends

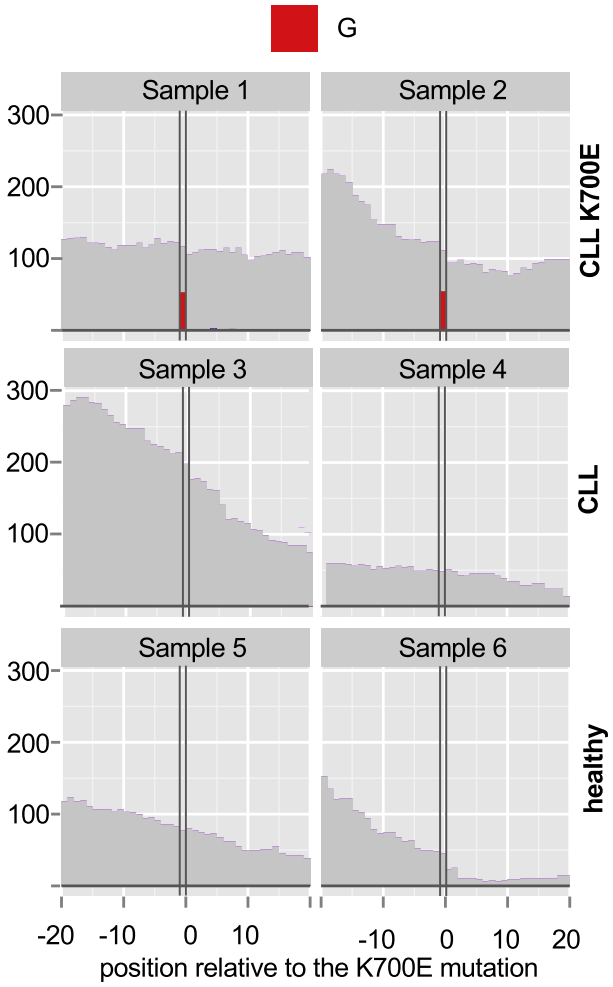

Figure 1. *SF3B1* variant expression. Each panel shows the data for one sample. The y-axis depicts the genomic coverage resulting from the alignment of the RNA-Seq fragments and the x-axis represents the genomic position relative to the position 198,266,834 of chromosome 2 (indicated by the vertical black lines), where the reference genome contains an adenine. The coverage consistent with the reference genome is coloured in gray. The first two samples express the variant that contains a guanine in around half of the transcripts (as indicated by the height of the red bar). This variant is predicted to cause the substitution K700E on the protein coded by *SF3B1*.

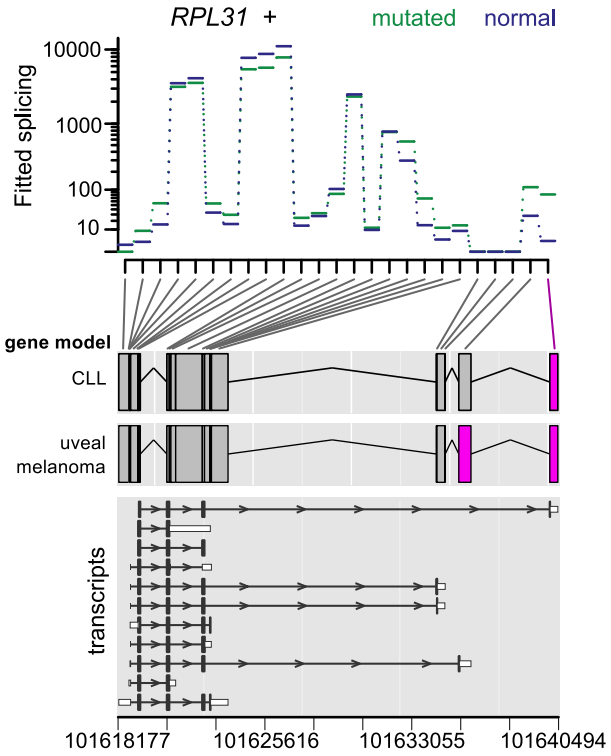

Figure 2. Differential exon usage of gene *RPL31*. The upper panel shows each exon region represented along the x-axis. The y-axis shows the fitted coefficients of the generalized linear model corresponding to the exon usage values. The fitted value from the samples with the mutation in *SF3B1* are coloured in green, and the value from the wild-type samples are coloured in blue. The panels below show the gene model, where the exons detected to be significant for DEU are coloured in magenta. Our results from the CLL samples are presented, as well as the results from the uveal melanoma study [6]. The lower panel depicts the transcripts for *RPL31* annotated in *ENSEMBL*.

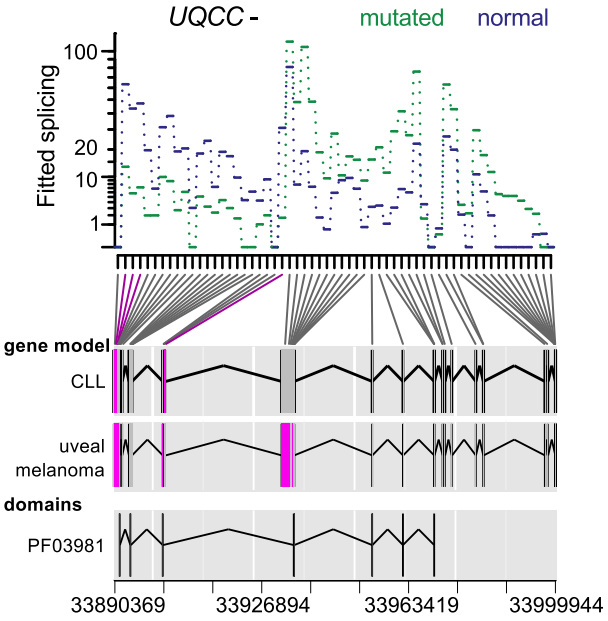

Figure 3. Differential exon usage of gene *UQCC*. Legend as in Figure 3, except that the lower panel presents the protein domain *PF03981* annotated in *PFAM* [27]. This domain is a region of the protein Ubiquinol-cytochrome C chaperone that seems to be conserved in different species clades, including yeast and bacterial species like *Sinorhizobium melilotli*.

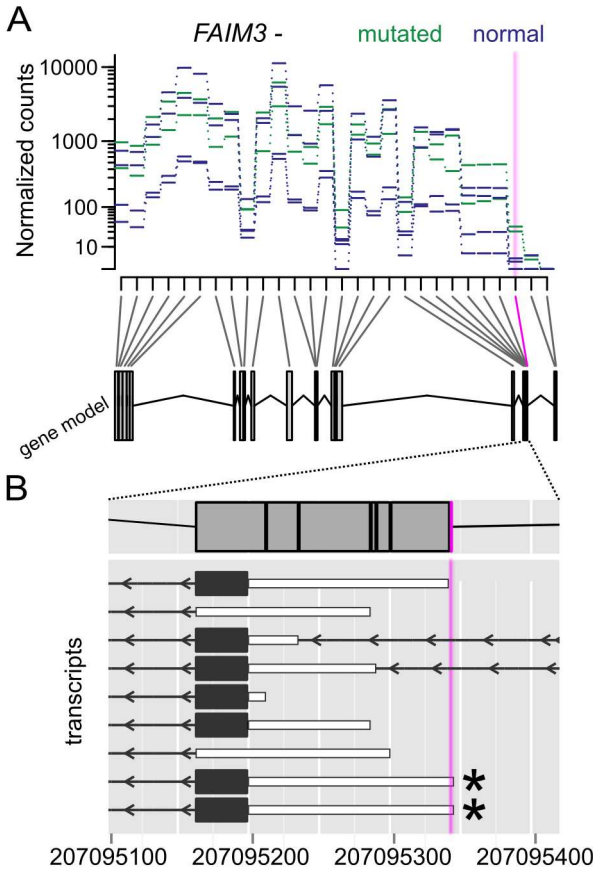

Figure 4. Differential exon usage of gene *FAIM3*. *(A)* As in Figure 2, except that the normalized counts for each sample are plotted along the y-axis. *(B)* Detailed view of the region differentially used, where the transcripts annotated in *ENSEMBL* are also shown. The exonic region that is differentially used overlaps with the 5' untranslated region of two transcripts (as indicated by the asterisks).

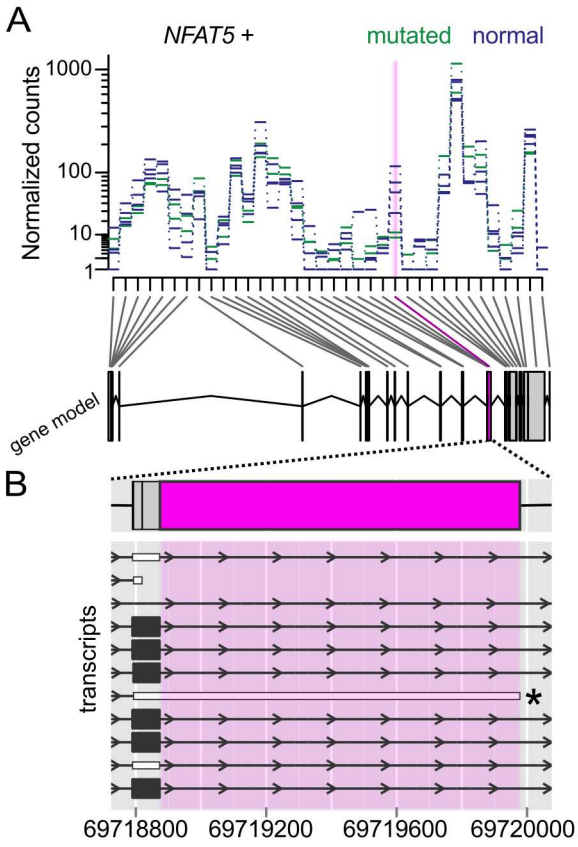

Figure 5. Differential exon usage of gene *NFAT5*. Legend as in Figure 4. The exon region differentially used overlaps with part of the *ENSEMBL* transcript ENST00000567990 (as indicated by the asterisk), which is known to be subject to non-sense mediated decay.

Supporting Tables

Table S1. Clinical and laboratory data of the CLL patients studied.

Table S2. Selected pathways enriched among genes with differential exon usage associated with the *SF3B1* mutation (FDR = 0.1).

Table S3. Oligonucleotide sequences used per sample.

Supporting Datasets

Dataset S1. HTML report of the genes with DEU associated with the mutations in *SF3B1*

Supporting Files

File S1. Documented *R* session with the program code needed to reproduce our analysis of the data and to generate the figures