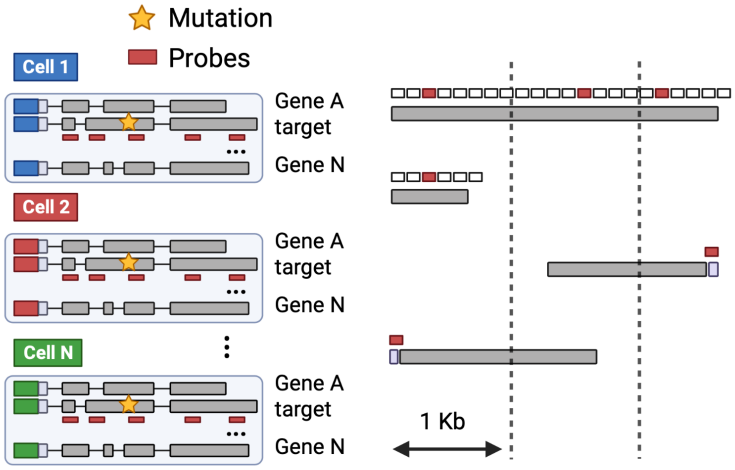
Supplementary Data for:

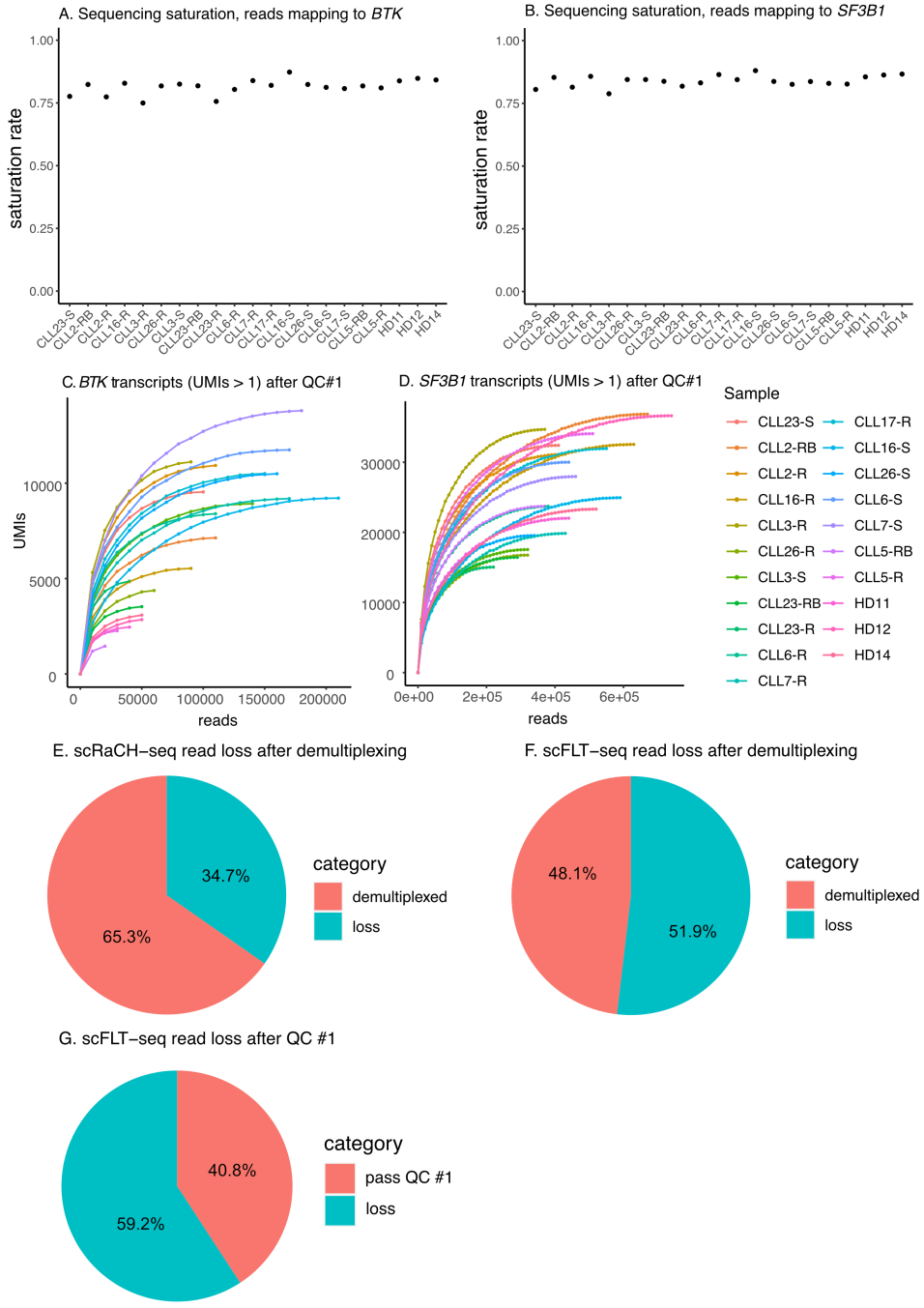
Single-cell Rapid Capture Hybridization sequencing reliably detects isoform usage and coding mutations in targeted genes

Supplemental Figure S1



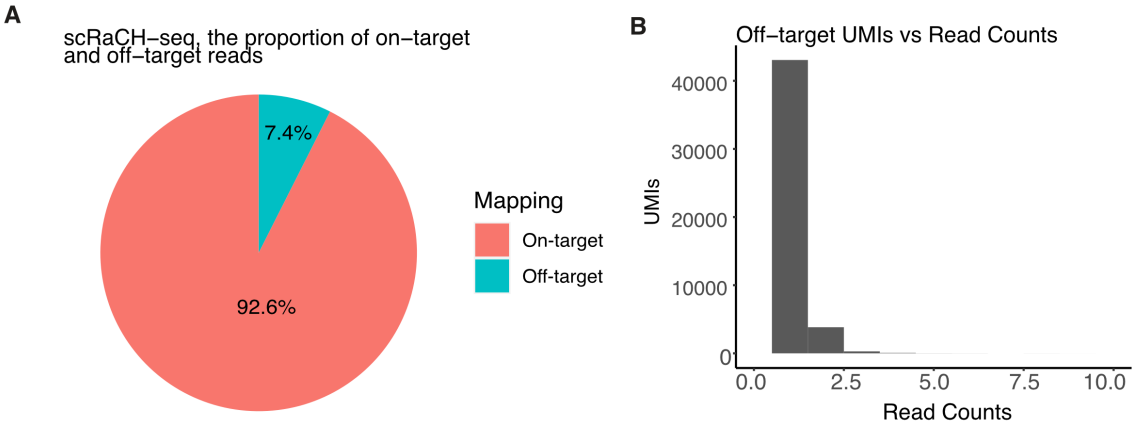
Supplemental Fig S1. Illustration of the strategy for probe panel design. Probes were designed to target the gene of interest (left panel). The probes (red and white) were designed to be 120 bp. One probe (red) per kilobase of the target gene exon was selected based on their GC content. If the exons were longer than 1 kilobase, multiple probes were selected per exon. For exons shorter than 120 base pairs, the exons were aligned to create sequences longer than 120 base pairs. This strategy was employed to ensure that each annotated exon was covered by at least one probe.

Supplemental Figure S2



Supplemental Fig S2. scRaCH-seq achieves high sequencing saturation and has lower read loss after demultiplexing and read integrity checks.
A) Sequencing saturation rates of reads mapping to *BTK* in all samples.
B) Sequencing saturation rates of reads mapping to *SF3B1* in all samples.
C) Sequencing saturation of transcripts of *BTK* with UMIs appearing more than twice across all samples in scRaCH-seq data.
D) Sequencing saturation of transcripts of *SF3B1* with UMIs appearing more than twice across all samples in scRaCH-seq data.
E) The proportion of read loss in scRaCH-seq data after demultiplexing using *FLAMES.*
F) The proportion of read loss in scFLT-seq data after demultiplexing using *FLAMES.*
G) The proportion of read loss in scFLT-seq data after read integrity checks.

Supplemental Figure S3



C

Similar sequences shared by *SF3B1* and *RPL13* (a top off-target gene)

Range 1: 497 to 703 Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
123 bits(136)	3e-31	159/216(74%)	9/216(4%)	Plus/Plus

```
Query  1633  GGAGATCGAGACCATCCTGGCTAACGCGGTGAAGTCCCGTCTACTAAAAATACAAAAATT  1692
             ||||| || ||||||| ||||||| |||     | |||     |||||||||||| | |||||
Sbjct  497   GGAGTTCAAGACCAGCCTGGCCAAC------ATGTCAGAACTACTAAAAATAAAGAAATC  550

Query  1693  AGCCAGGCGTGGTGGCGGGCGCCTGTAGTCCCAGCTACTAGGGAGGCTGAGGCAGGAGAA  1752
             ||||| || |||||  |  | | ||||||  ||||| || |||||||  ||||  |  | |
Sbjct  551   AGCCATGCTTGGTGCTGCACACTTGTAGTTGCAGCTCCT-GGGAGGCAGAGG-TGAGGGA  608

Query  1753  TGGCTTGAACCCGGGAGGCGGAGCTTGCAGTGAGCCGAGATAGCGCTGCTACACTCCAGC  1812
             |  ||| ||||| ||||||| |||  |||| ||||||  |||  |||  || ||||| |||
Sbjct  609   TCACTT-AACCCAGGAGGCAGAGGCTGCACTGAGCCAGGATCACGCCACTGCACTCTAGC  667

Query  1813  CTGGGCGACAGAGACTCCGTCTCaaaaaaaaaaaaa  1848
             ||||||| |||| ||  | ||||||||||||||||||
Sbjct  668   CTGGGCAACAGTGAGACTGTCTCAAAAAAAAAAAAA  703
```

Range 2: 1545 to 1606 Graphics ▼ Next Match ▲ Previous Match ▲ First Match

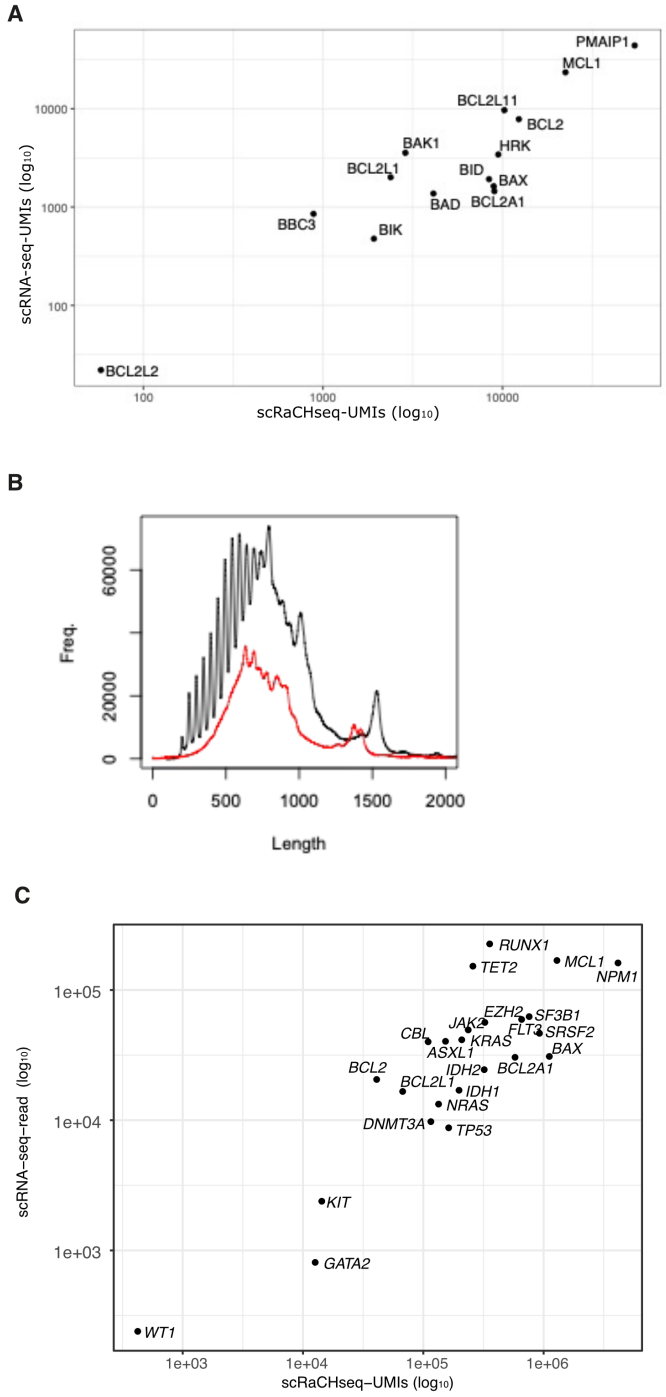
Score	Expect	Identities	Gaps	Strand
73.4 bits(80)	4e-16	55/64(86%)	2/64(3%)	Plus/Plus

```
Query  1562  GGCCGGGTGTGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGCGGA  1621
             |||||||||  |||||||| |||||||||||| | |||  |||||||| |||||||| |||
Sbjct  1545  GGCCGGGTGCAGTGGCTCACGCCTGTAATCCTAACAC--TGGGAGGCCGAGGCGGGTGGA  1602

Query  1622  TCAC  1625
             ||||
Sbjct  1603  TCAC  1606
```

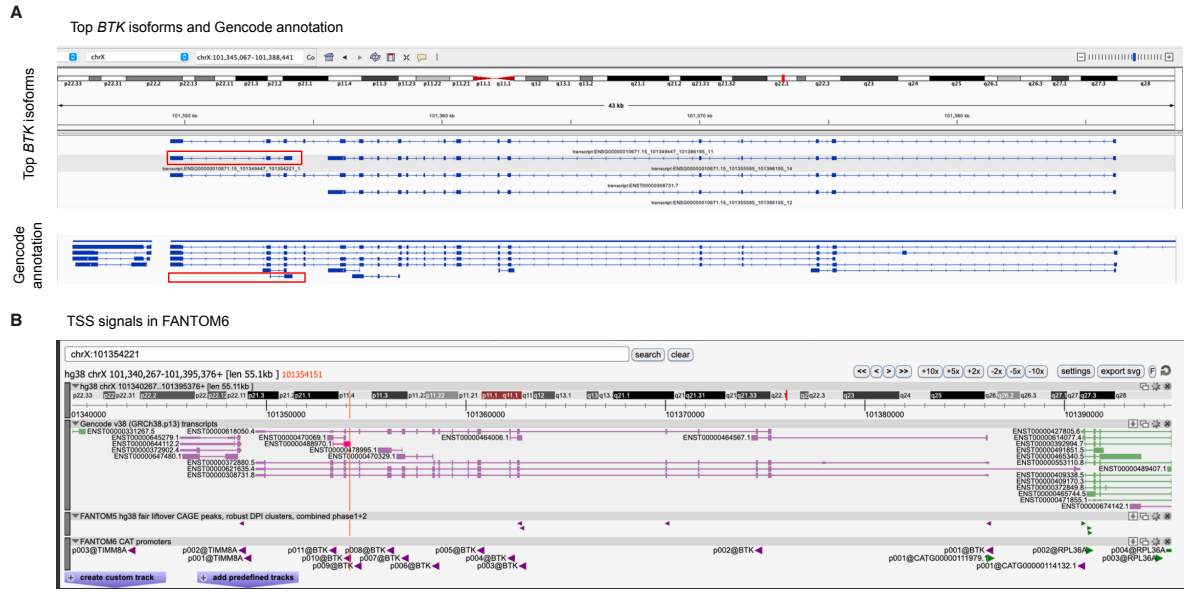
Supplemental Fig S3. A low proportion of total reads is affected by off-target reads.
A) The proportion of total reads affected by on-target and off-target reads.
B) Using sample CLL2-RB as an example, for the off-target transcripts, the read count on the X-axis was plotted against the UMI count on the Y-axis.
C) Shared sequences between *SF3B1* and *RPL13,* a top off-target gene. The alignment was conducted using BLASTn on BCBI.

Supplemental Figure S4



Supplemental Fig S4. The unique molecule identifiers (UMIs) captured by scRNA-seq and scRaCH-seq are correlated.
A) scRaCH-seq was performed with a probe panel of 292 probes targeting 17 BCL2 family genes. The dots represent the UMIs of target genes detected by scRaCH-seq (X-axis) and scRNA-seq (Y-axis) for a CLL sample.
B) A graph showing the read lengths from 0 to 2,000 base pairs from scRaCH-seq data of an AML sample, before (black line) and after (red line) demultiplexing and integrity checks.
C) scRaCH-seq was performed with a probe panel of 1157 probes targeting 24 genes frequently mutated in AML. The dots represent the UMIs of target genes detected by scRaCH-seq (X-axis) and scRNA-seq (Y-axis) for an AML sample.

Supplemental Figure S5

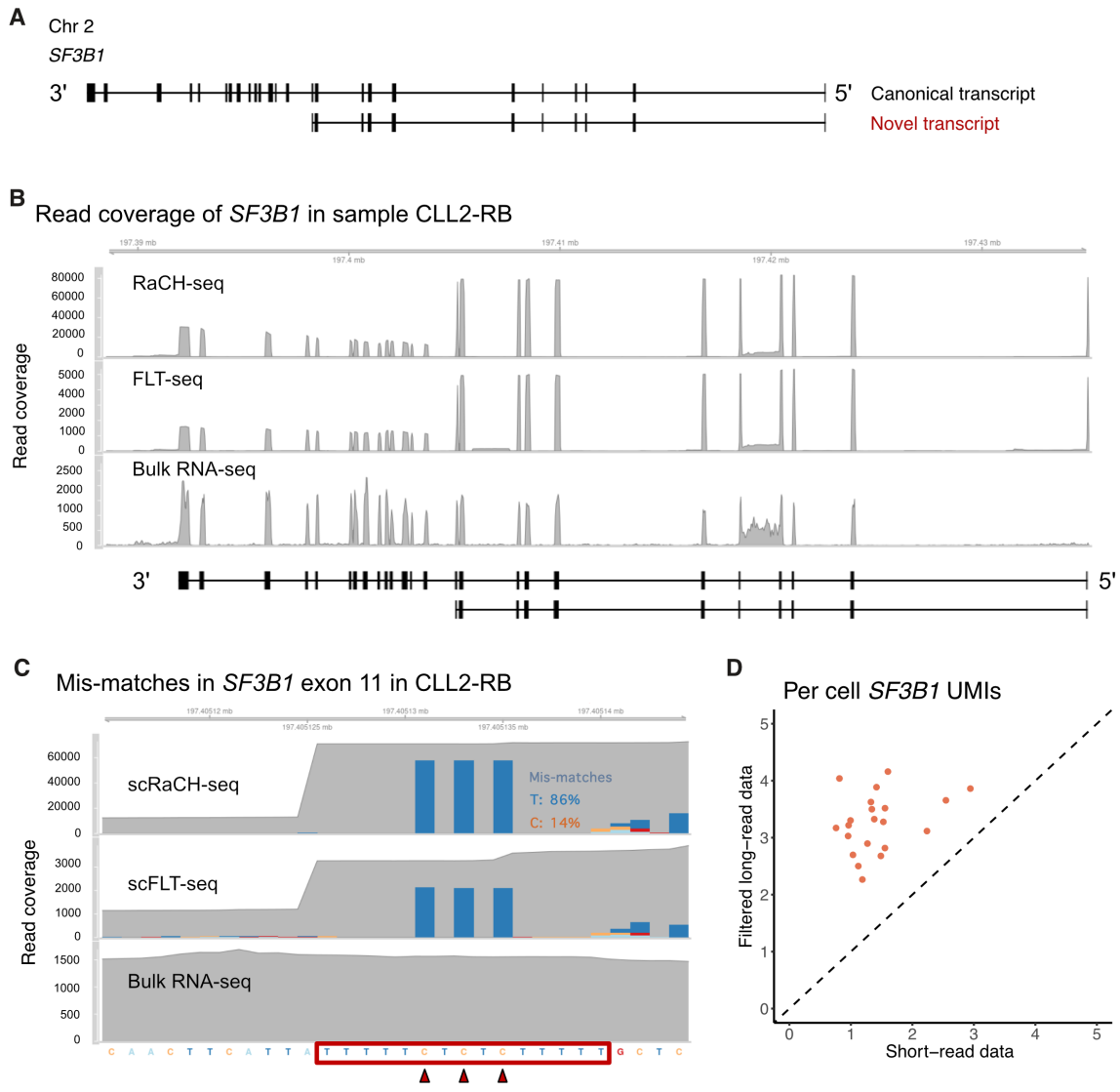


Supplemental Fig S5. *BTK* transcript #5 shares a nearby transcript start site (TSS) with an annotated *BTK* isoform

A) IGV screenshot showing the top 5 *BTK* transcripts identified in scRaCH-seq data (top panel) and the Gencode annotated isoforms (bottom panel). *BTK* transcript #5 and the annotated isoform sharing the same TSS are highlighted by the red squares.

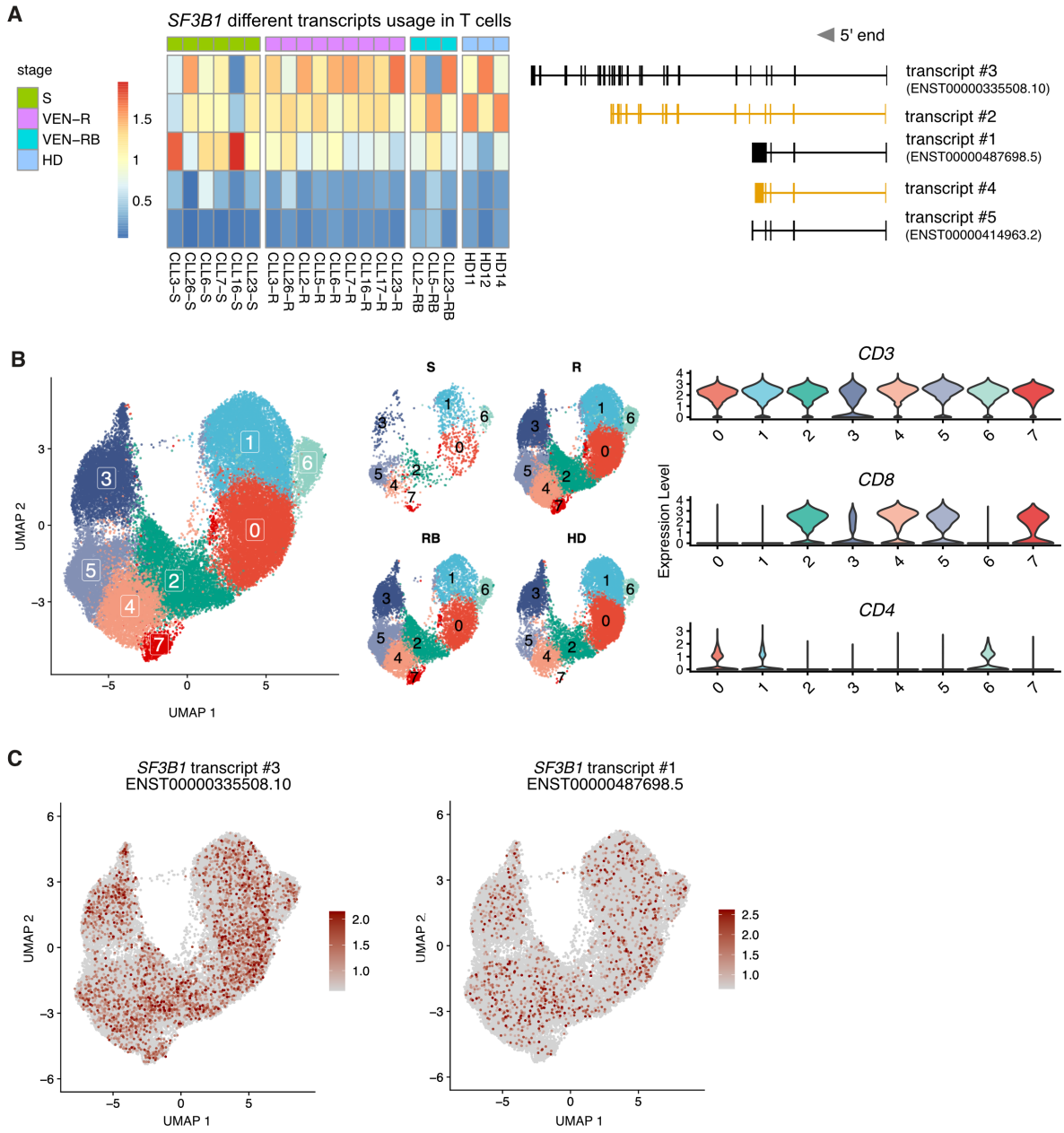
B) The TSS signals recorded in the CAGE database (FANTOM6). The orange line highlights the TSS signal around the start site of *BTK* transcript #5.

Supplemental Figure S6



Supplemental Fig S6. Identification of an *SF3B1* artefact introduced by 10x Genomics.
A) Graph showing the *SF3B1* read coverage per different sequencing method for sample CLL2-RB.
B) Graph showing the mismatches C→T mapped to *SF3B1* exon 11 in scRaCH-seq, scFLT and bulk RNA-seq data.
C) Dot plot showing the per cell *SF3B1* UMIs captured in scRNA-seq (X-axis) and scRaCH-seq (Y-axis) data after artefact removal.

Supplemental Figure S7



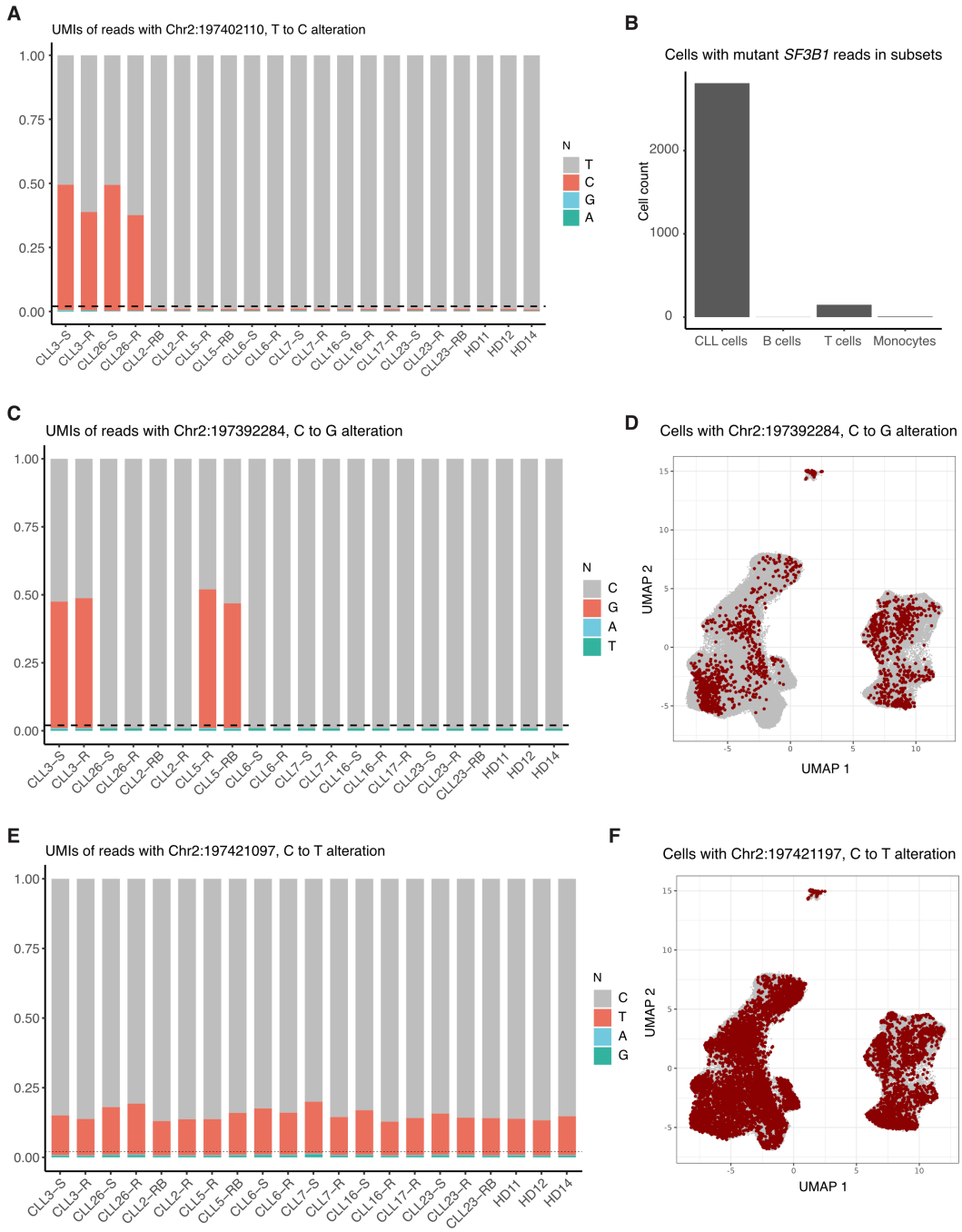
Supplemental Fig S7. *SF3B1* isoform usage by T cells from CLL patients

A) Heatmap showing *SF3B1* isoform usage (rows) per sample (columns) in the T cells. The top 5 *SF3B1* transcripts are illustrated on the right with novel transcripts highlighted in yellow. Samples are grouped in screening (S; green), venetoclax relapsed (VEN-R/R; pink), venetoclax relapsed and subsequently on BTKi (VEN-RB/RB; blue) and healthy donor (HD; purple).

B) UMAP projection of T cells from CLL patients and healthy donors and clustering based on short-read gene expression data. Middle panel showing the UMAP projection of T cells per treatment group. Violin plot showing the expression of the T cell marker (CD3), CD8 T cell marker (CD8) and CD4 T cell marker (CD4) per cluster.

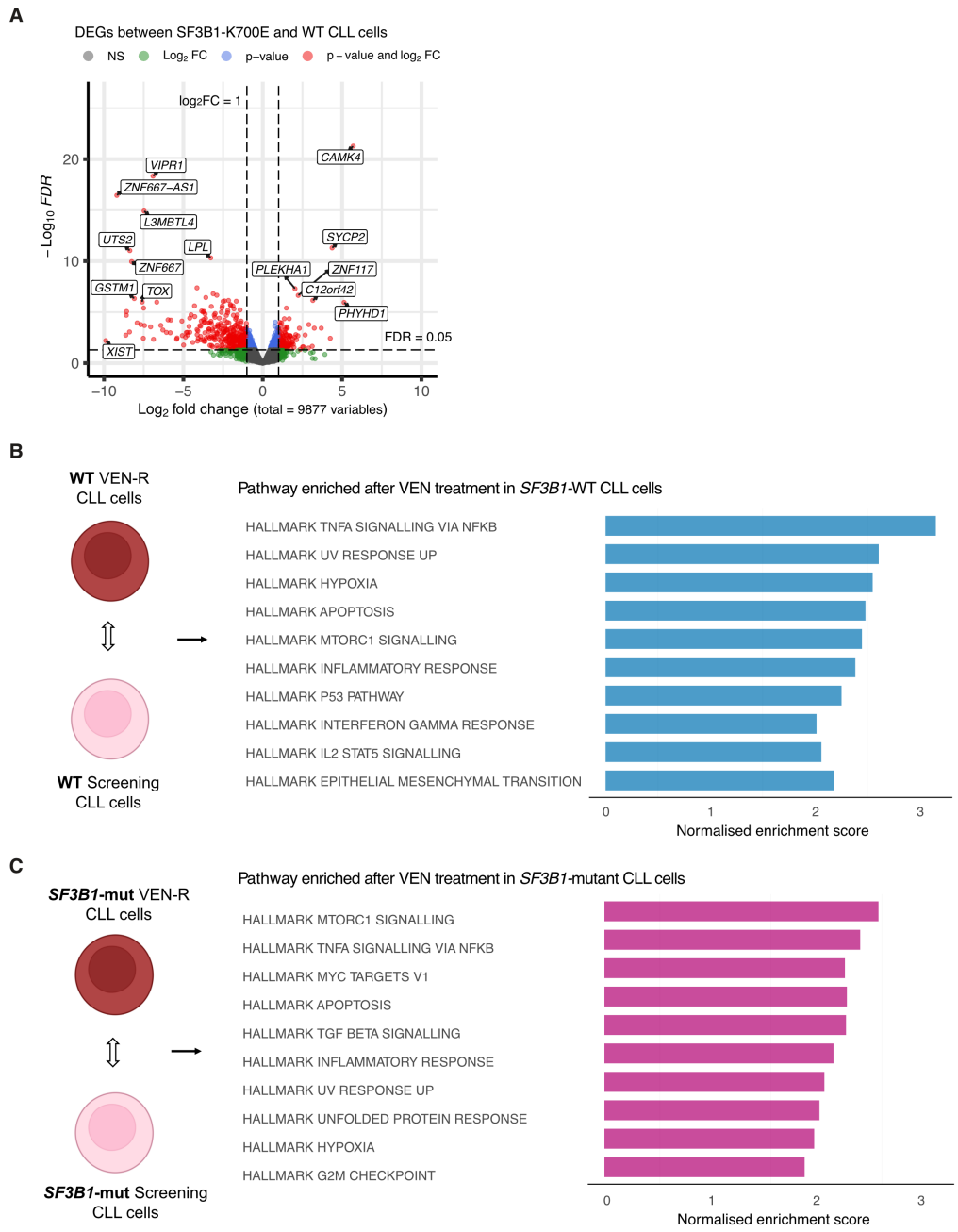
C) UMAP projection of *SF3B1* transcript #1 and #3 gene-level expression (red dots) in the T cells.

Supplemental Figure S8



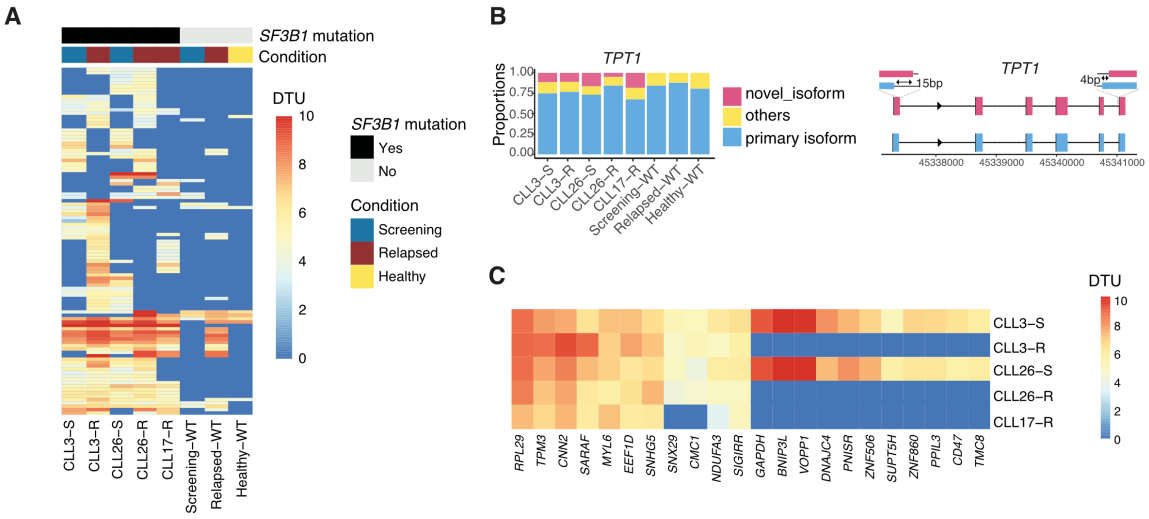
Supplemental Fig S8. *SF3B1* altered transcript expression in CLL and B cells.
A) Bar plot showing the abundance of nucleotide (A, T, C, G) at Chr2:197402110 per sample. The dashed horizontal line indicates the baseline sequencing error (3%) of nanopore sequencing.
B) Bar plot showing the quantification of cells with T-to-C alterations at Chr2:197402110 in CLL and non-CLL clusters.
C) Bar plot showing the abundance of nucleotide (A, T, C, G) at Chr2:197392284 per sample. The dashed horizontal line indicates the baseline sequencing error (3%) of nanopore sequencing.
D) UMAP projection of cells carrying the Chr2:197392284 C-to-G alteration (red dots)
E) Bar plot showing the abundance of nucleotide (A, T, C, G) at Chr2:197421097 per sample. The dashed horizontal line indicates the baseline sequencing error (3%) of nanopore sequencing.
F) UMAP projection of cells carrying the Chr2:197421097 C-to-G alteration (red dots).

Supplemental Figure S9



Supplemental Fig S9. Altered transcriptional expression in SF3B1 K700E mutated cells.
A) Volcano plot showing the differentially expressed genes (DEGs; FDR<0.05) found by pseudo-bulk DE analysis between SF3B1 K700E mutant and wild-type CLL cells.
B) Graph showing the enriched pathways in wild-type *SF3B1* venetoclax-relapsed CLL cells (blue). CLL cells with >10 wild-type *SF3B1* transcripts were used for differential gene expression analysis between the venetoclax-relapsed (VEN-R) group to the screening group. The differentially expressed genes were used as input for gene set enrichment analysis.
C) Graph showing the enriched pathways in SF3B1 K700E mutant venetoclax-relapsed CLL cells (pink). CLL cells with >2 mutant *SF3B1* transcripts were used for differential gene expression analysis between the venetoclax-relapsed (VEN-R) group to the screening group. The differentially expressed genes were used as input for gene set enrichment analysis.

Supplemental Figure S10



Supplemental Fig S10. Novel transcript usage in *SF3B1* mutated CLL samples.
A) Heatmap showing differential transcript usage (DTU) compared to primary isoform for genes (rows) per sample (columns). The wild-type *SF3B1* CLL cells from patients at screening (blue), at venetoclax-relapsed (red) and the healthy donor B cells (yellow) are combined.
B) Bar plot showing the novel isoform usage of the *TPT1* gene selected from the heatmap in Figure 8A. Novel spliced isoform is highlighted in pink and illustrated in the right panel.
C) Heatmap showing differential transcript usage (DTU) compared to primary isoform for selected genes (columns) per *SF3B1*-mutated sample (rows). The genes are selected based on differences between the screening and venetoclax-relapsed groups.