Identification of differentially expressed splice variants by the proteogenomic pipeline Splicify
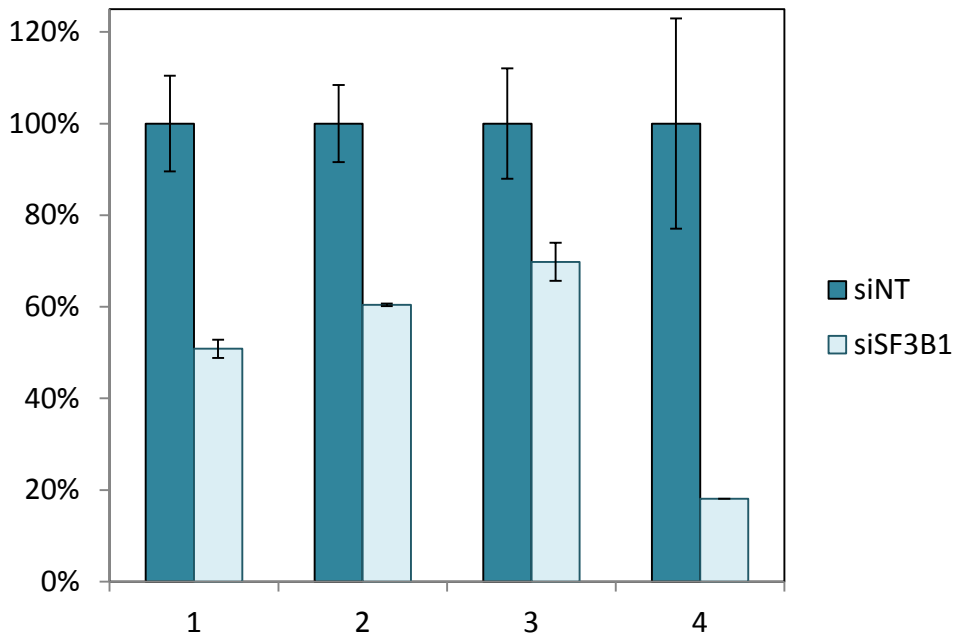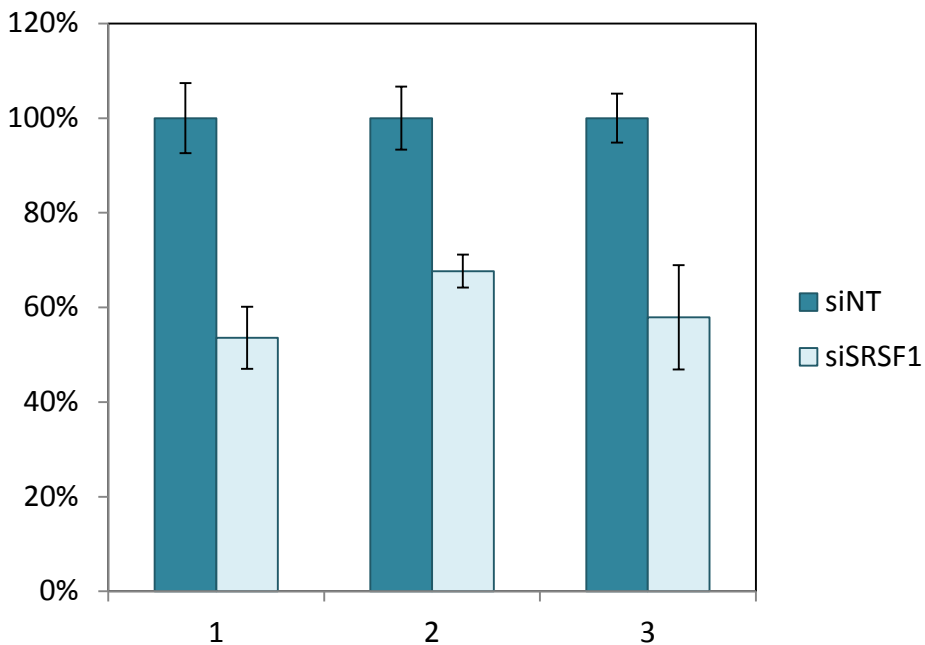
Supplementary Figures

**Supplementary Figure 1. Knock-down efficiency of SF3B1 and SRSF1**
A Mean relative expression of SF3B1 upon siRNA-mediated down-modulation (siSF3B1) compared to the non-targeting control (siNT). RNA was harvested in 48 hours after transfection. The experiment was performed three times in technical duplicates. Then experiment was repeated the forth time in technical replicate for the control and as a single experiment for the siSF3B1 **B** Mean relative expression quantified by RT-qPCR of SRSF1 upon siRNA-mediated down-modulation (siSRSF1) compared to the control (siNT). RNA was harvested in 72 hours after transfection. The experiment was performed three times in technical duplicates.
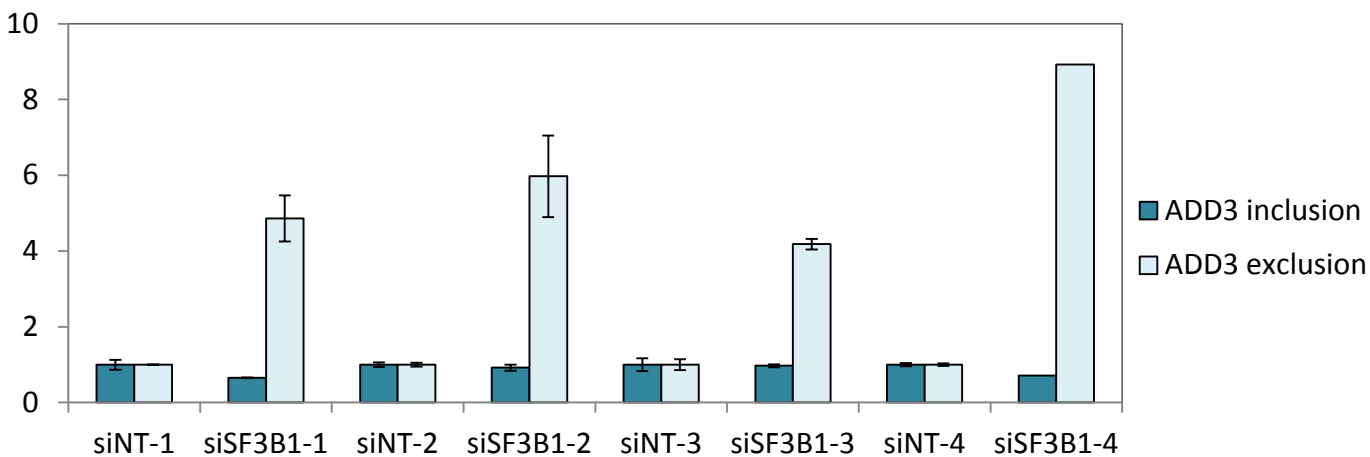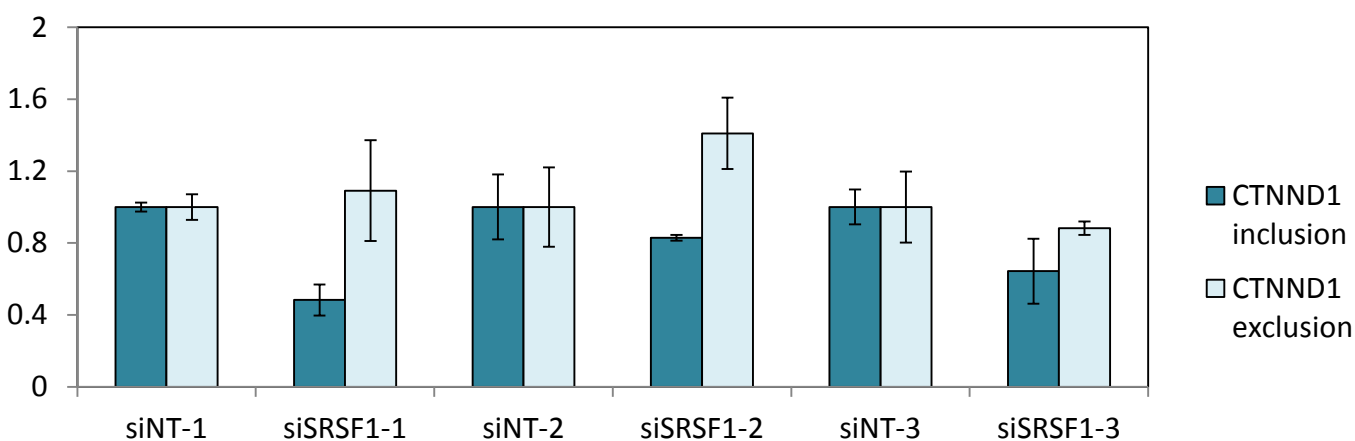
Supplementary Figure 2 Positive controls of alternative splicing
A RT-qPCR quantification of the inclusion and exclusion variants of ADD3 exon 14 in siSF3B1 compared to the siNT control. Upon-down modulation of SF3B1 there is higher expression of the ADD3 isoform in which exon 14 is excluded. **B** RT-qPCR quantification of the inclusion and exclusion variants of CTNND1 exon 20 in siSRSF1 compared to the siNT control. Upon down-modulation of SRSF1 there is higher expression of the CTNND1 isoform in which exon 20 is excluded.
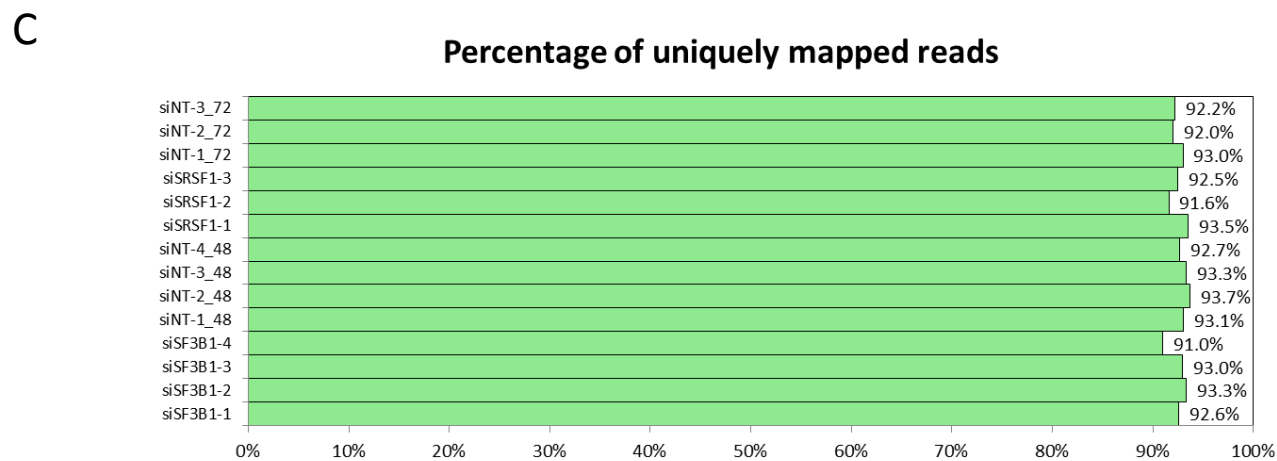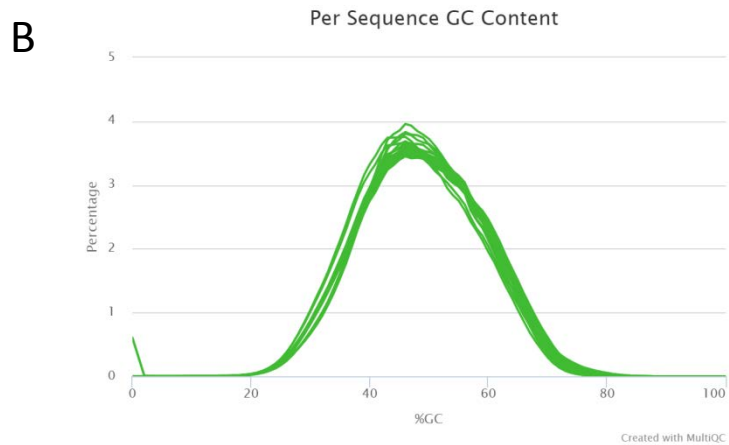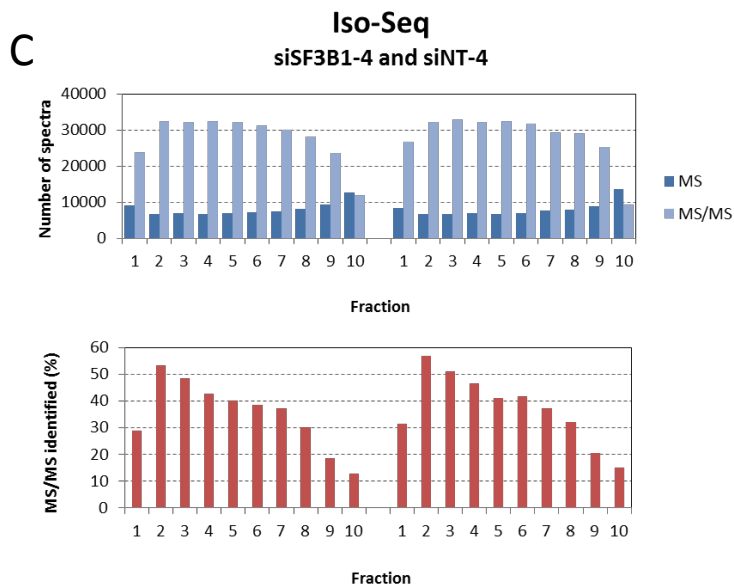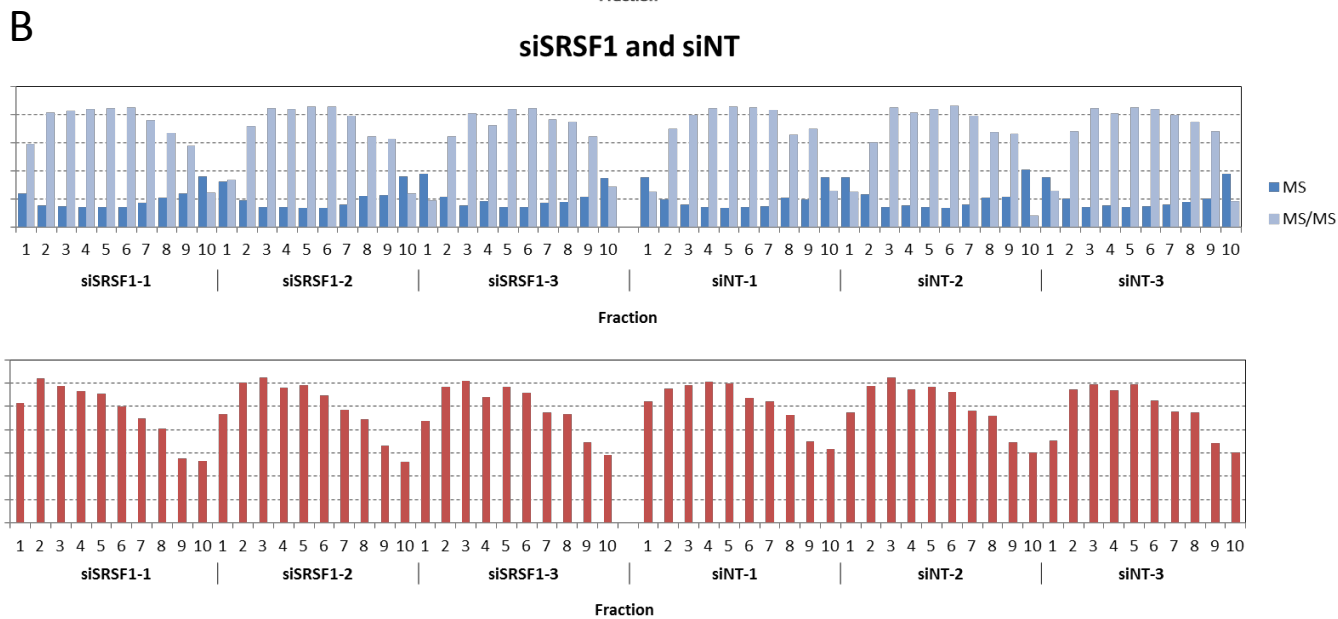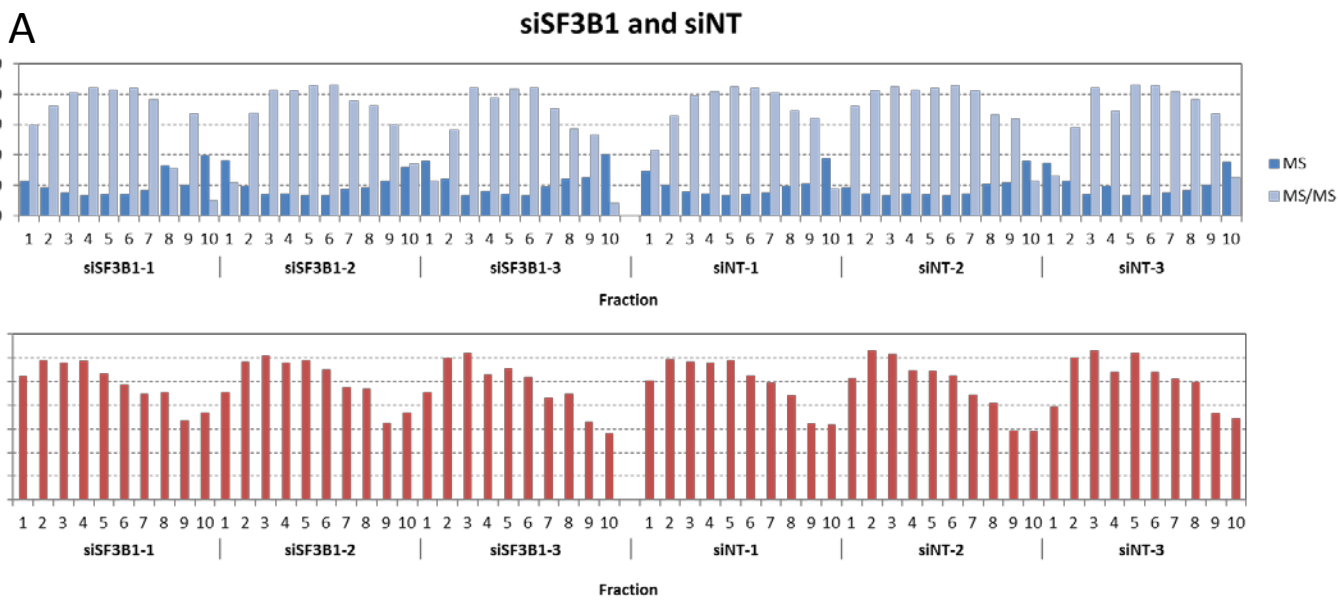
Supplementary Figure 3 Quality checks of RNA-seq data
A The mean phred score value across each base position in the read was calculated with FastQC. Each line represents a sample that was subjected to RNA-seq. Plot was produced with MultiQC. **B** The average GC content of all the reads was calculated with FastQC. Each line represents a sample that was subjected to RNA-seq. Plot was produced with MultiQC. **C** Percentage of uniquely mapped RNA reads per sample were obtained from STAR output. The controls for siSF3B1 were labelled "siNT-x_48 " and the controls for siSRSF1 were labelled "siNT-x_72".
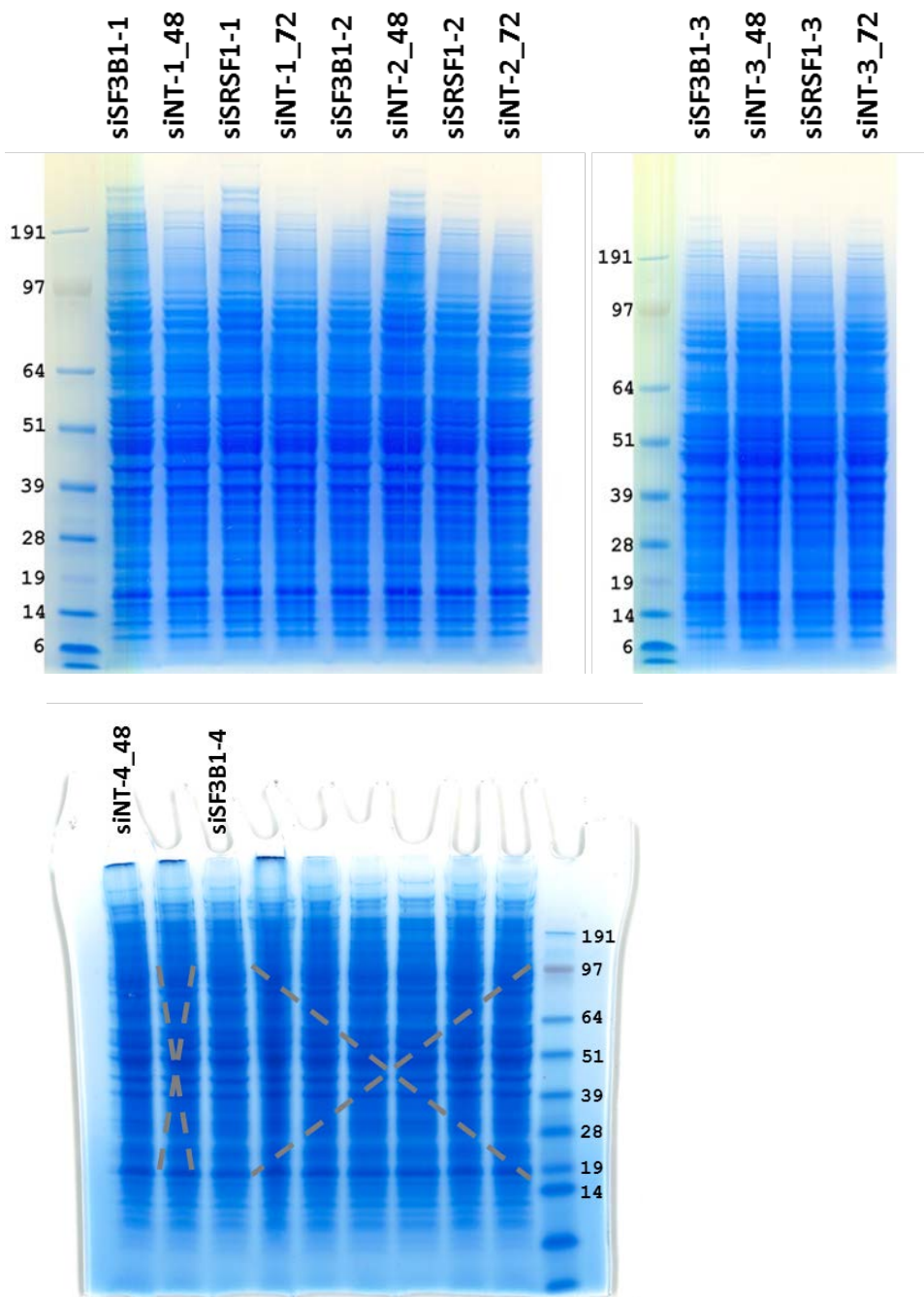
**Supplementary Figure 4 The Coomassie-stained gels**

Coomassie-stained gels present the protein band pattern of all the samples subjected to mass spectrometry, indicating equal protein loads. Sample names are shown above each gel lane. Gel lanes not used are not annotated and are crossed out. The controls for siSF3B1 were labelled "siNT-x_48 " and the controls for siSRSF1 were labelled "siNT-x_72".

Supplementary Figure 5 Quality checks of mass spectrometry database search
Bar plots represent the number of MS and MS/MS spectra identified and the percentage of MS/MS spectra identified in the database search performed by MaxQaunt. Numbers were obtained from MaxQuant summary files produced per each search; siSF3B1 and siNT **(A)**, siSRSF1 and siNT **(B)** and Iso-Seq experiment **(C)**.
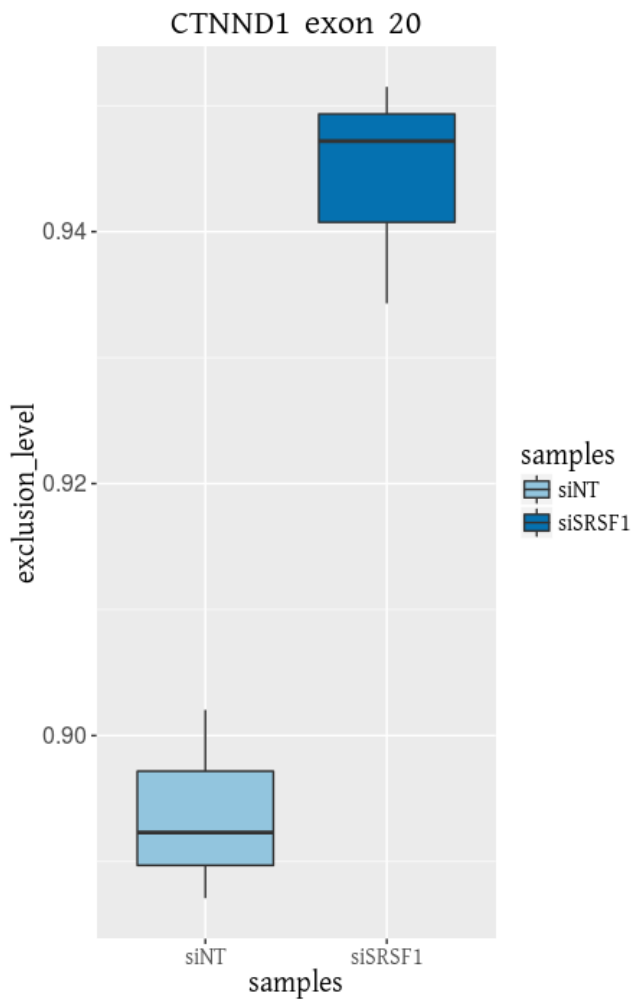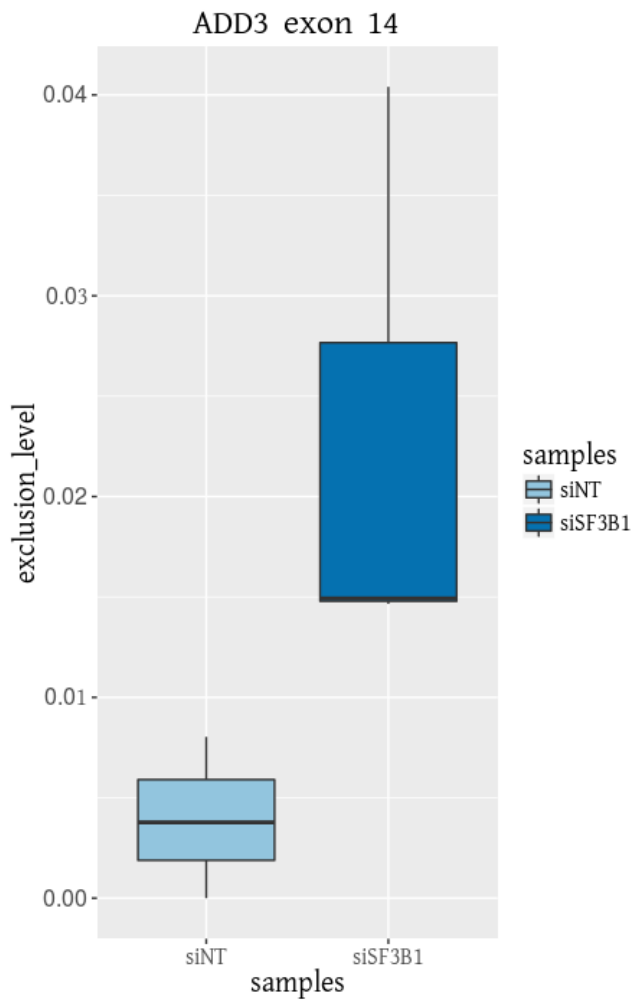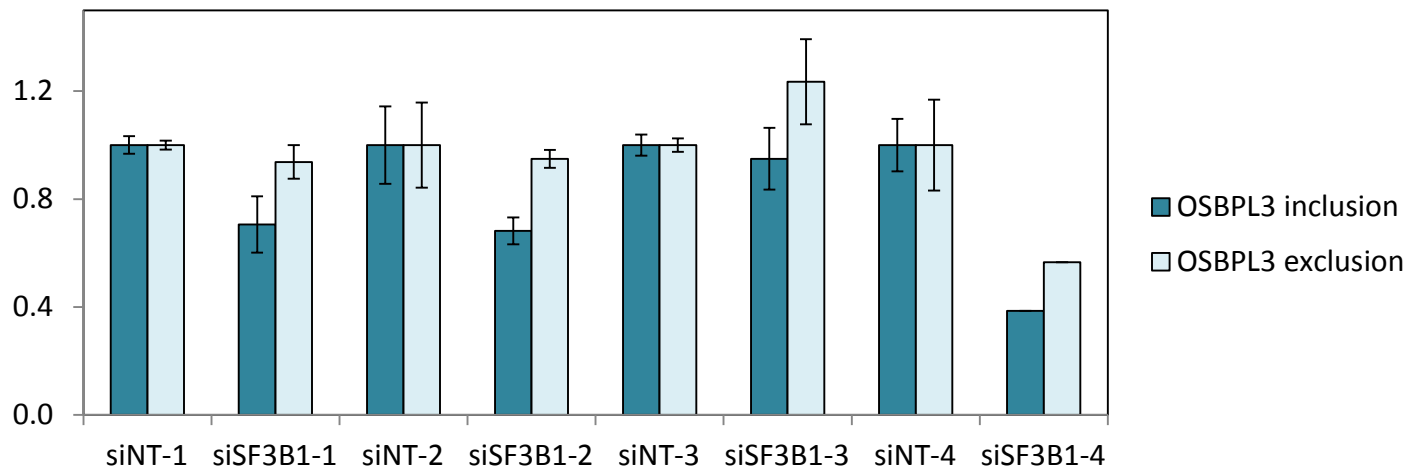
Supplementary Figure 6 Positive controls of alternative splicing identified with RNA-seq data.

Exclusion isoforms for ADD3 exon 14 in siSF3B1 and its control (siNT) and for CTNND1 exon 20 for siSRSF1 and its control (siNT). Here, exclusion level is higher in siSF3B1 and siSRSF1 *versus* siNT, respectively, for both exclusion of exon 14 in ADD3 and exclusion of exon 20 in CTNND1. Exclusion level was calculated based on exclusion spanning reads divided by the sum of inclusion and exclusion spanning reads.
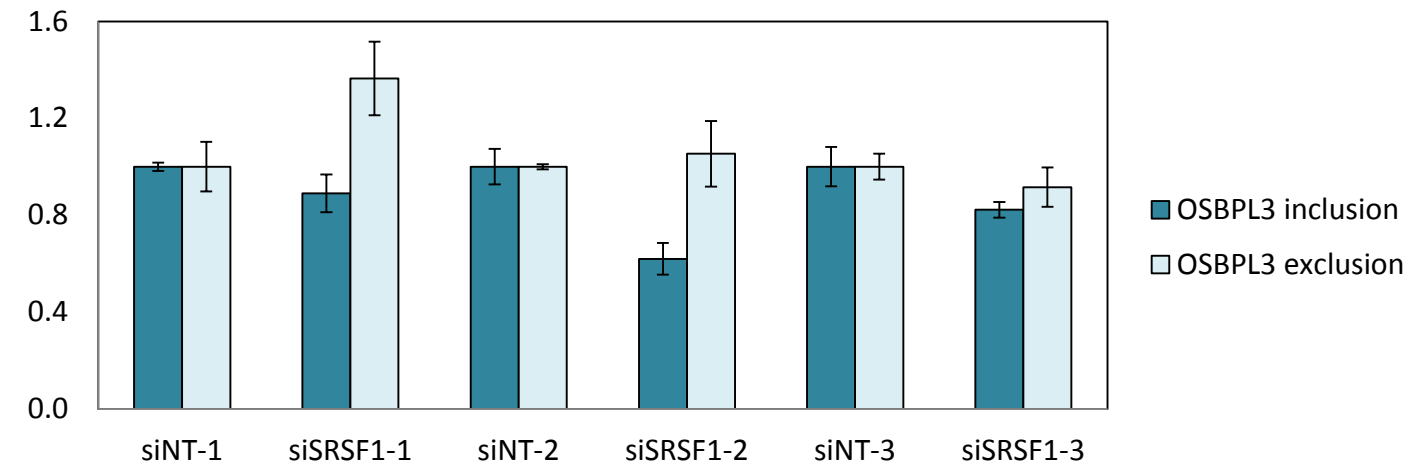
Supplementary Figure 7 RT-qPCR quantification of the inclusion and exclusion variants of OSBPL3 exon 9 upon knock-down of siSF3B1 and siSRSF1 compared to the control (siNT). Upon both down-modulation of SF3B1 or SRSF1 there is higher expression of the OSBPL3 isoform where exon 9 is excluded.

Mean relative expression of MKI67 exon 7
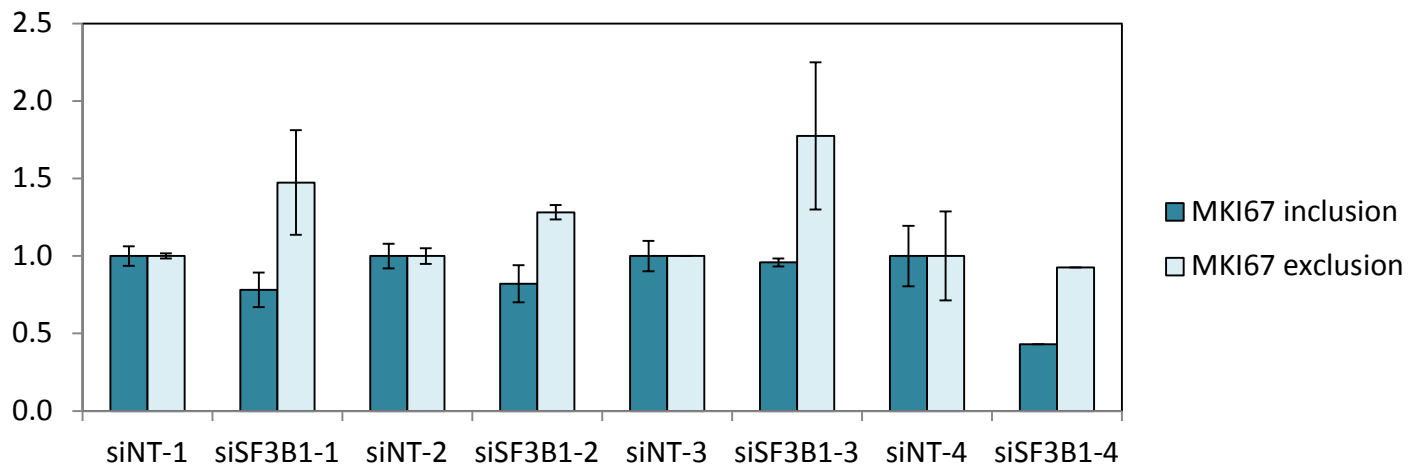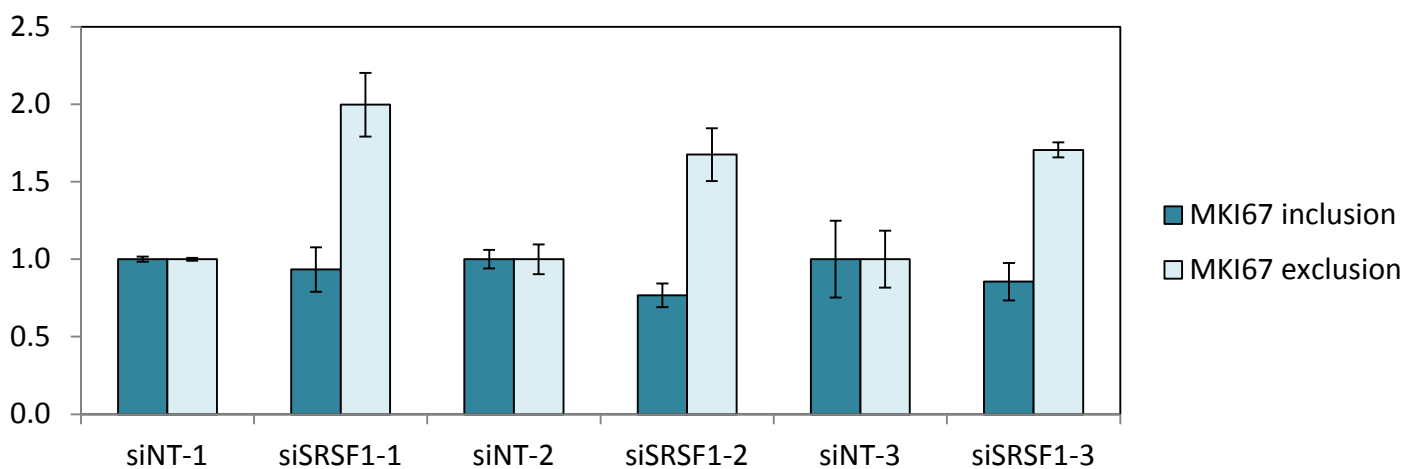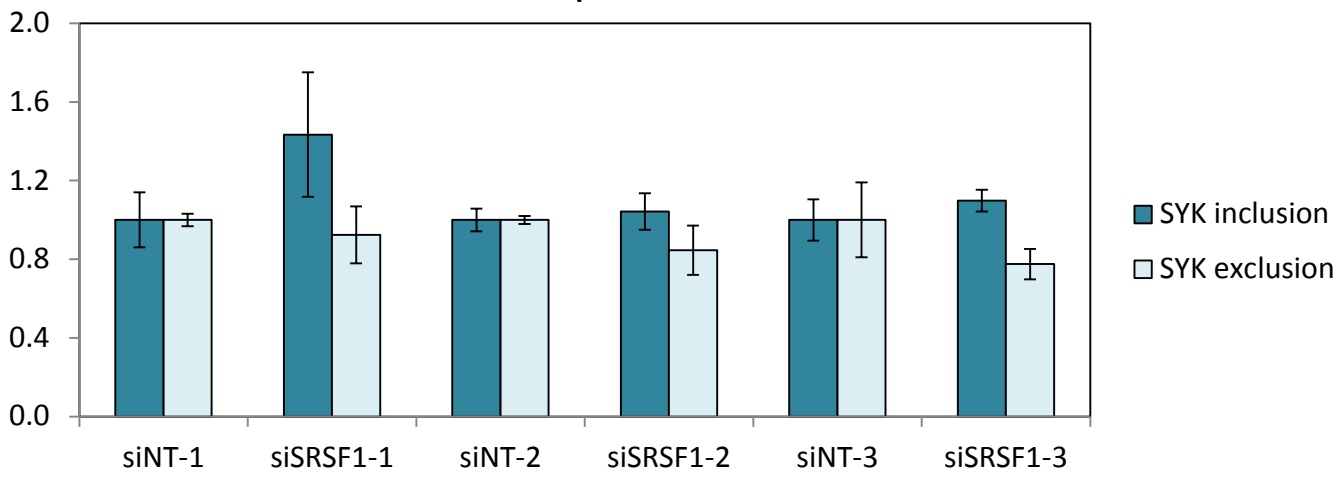siSF3B1 vs siNT

**Mean relative expression of MKI67 exon 7**
siSRSF1 vs siNT

Supplementary Figure 8 RT-qPCR quantification of the inclusion and exclusion variants of MKI67 exon 7 in siSF3B1 and siSRSF1 in comparison to the control (siNT). Upon both down-modulation of SF3B1 or SRSF1 there is higher expression of the MKI67 isoform where exon 7 is excluded.

A

Mean relative expression of SYK exon 7



B

Mean relative expression of RAC1 exon 4
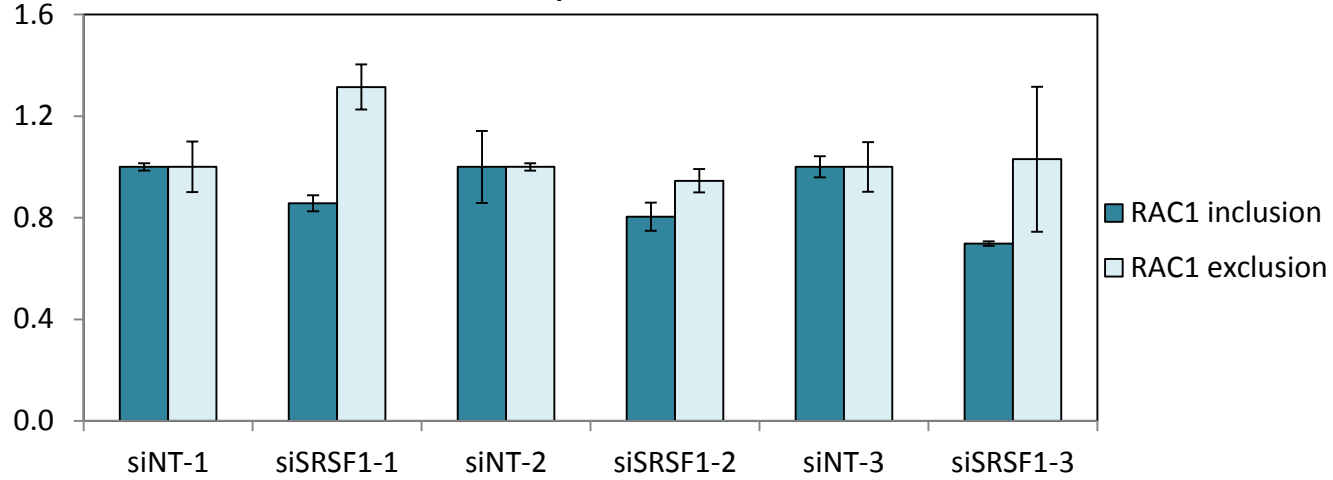
Supplementary Figure 9 RT-qPCR validation of skipped exon in SYK and RAC1 upon down-modulation **of** **SRSF1**
A RT-qPCR quantification of the inclusion and exclusion variants of SYK exon 7 in siSRSF1 in comparison to the control (siNT). Upon down-modulation of SRSF1 there is higher expression of the SYK isoform where exon 7 is included. **B** RT-qPCR quantification of the inclusion and exclusion variants of RAC1 exon 4 in siSRSF1 in comparison to the control (siNT). Upon down-modulation of SRSF1 there is higher expression of the RAC1 isoform where exon 4 is excluded.
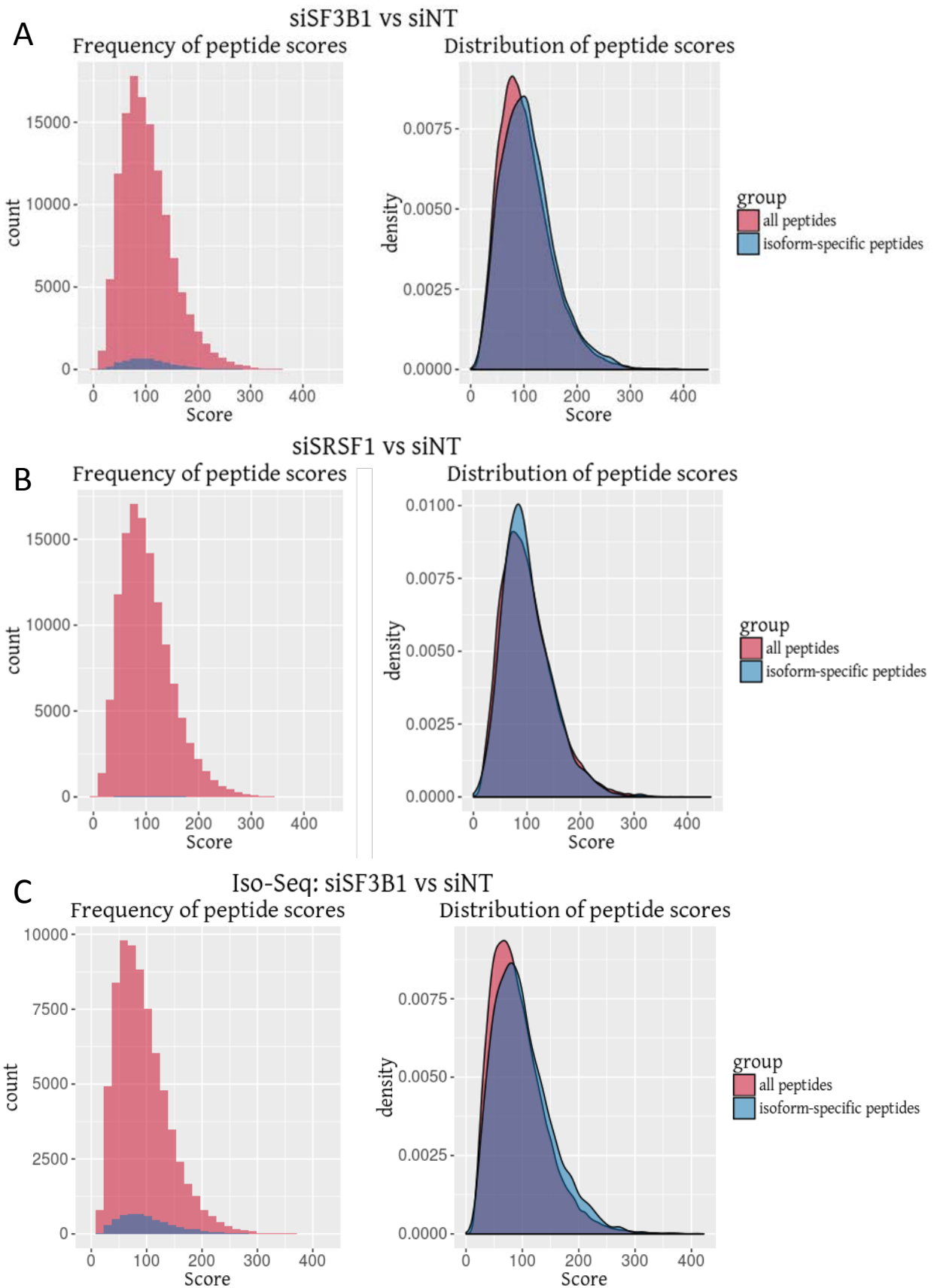
Supplementary Figure 10 Comparison of peptide scores of isoform-specific peptides to all identified peptides

Peptide scores (Andromeda scores) were calculated by MaxQuant and obtained from the peptide output file for each database search; siSF3B1 and siNT **(A)**, siSRSF1 and siNT **(B)** and Iso-Seq experiment **(C)**. The histogram represents the frequency (count) of each peptide score. Kernell density was calculated to obtain the distribution of peptide scores. The figures show that isoform-specific peptides are not scoring better or worse than the standard peptides, indicating that they were correctly identified.
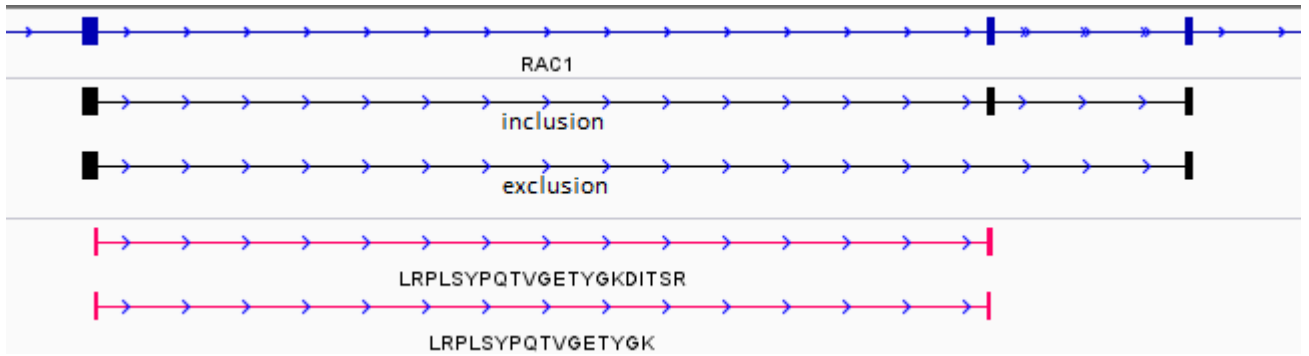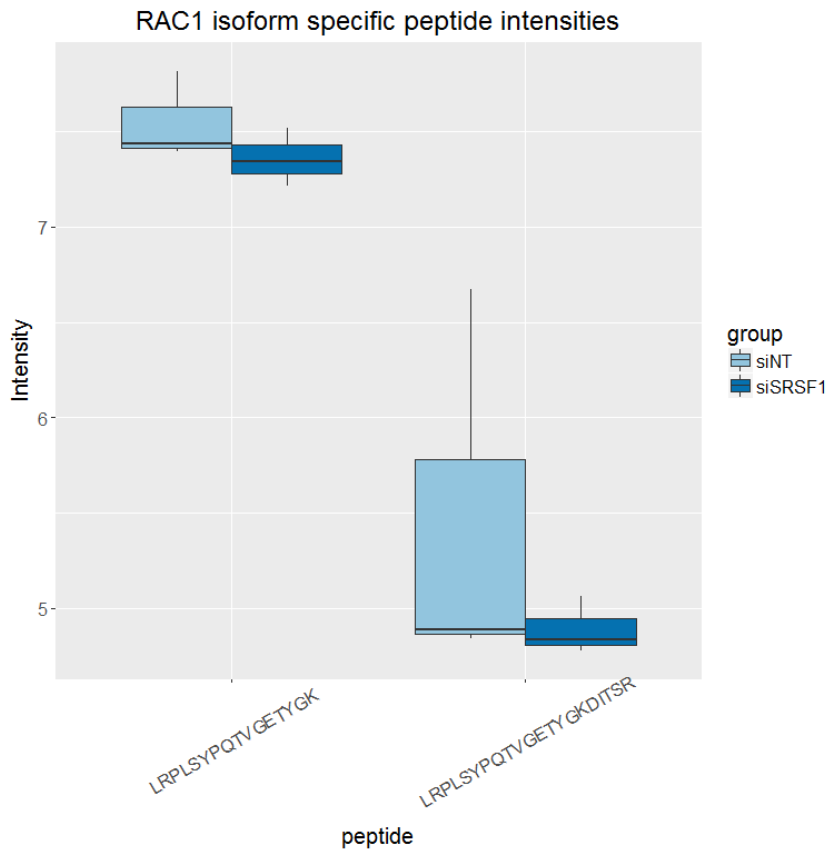
Supplementary Figure 11 A An IGV screenshot of a fragment of RAC1 gene; in blue – RefSeq Gene, in black – skipped exon inclusion and exclusion variants identified in the RNA-seq data, in pink – split peptides supporting inclusion on the exon 4 in RAC1 gene. **B** Differences in peptide intensities between down-modulation of siSRSF1 and the control for the inclusion specific peptides for RAC1 isoform. Expression differences indicate that inclusion isoform is higher expressed in the control. Peptide expression confirms differential isoform expression obtained from the RNA-seq data.