Supplementary Figures

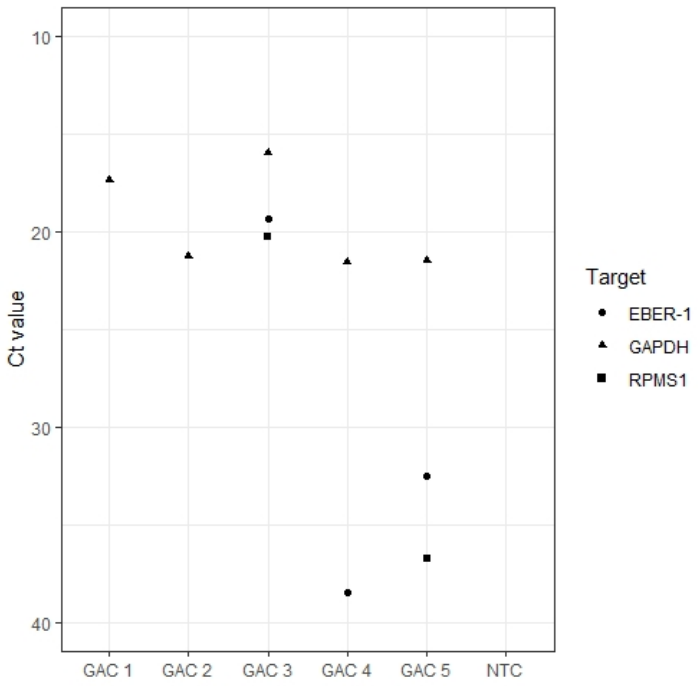

Figure S1. RT-qPCR results. Five gastric adenocarcinomas (GAC1-5) were tested for EBV by RT-qPCR of RPMS1 and EBER-1. GAPDH served as an internal control and nuclease-free water was used as no template control (NTC). Ct values are presented as the mean of technical duplicates. GAC3 was chosen for further processing and referred to in text as GAC.

The following data was generated by using the final annotation comprising all exons. Threshold value is here defined as *reads per transcript variant* and is plotted with a stepwise increment of five. A threshold value of 10, which corresponds to intermediate-confidence classification, means that all transcript variants supported by < 10 reads are disregarded; whereas a threshold value of 100, which corresponds to high-confidence classification, means that all transcript variants supported by < 100 reads are disregarded.

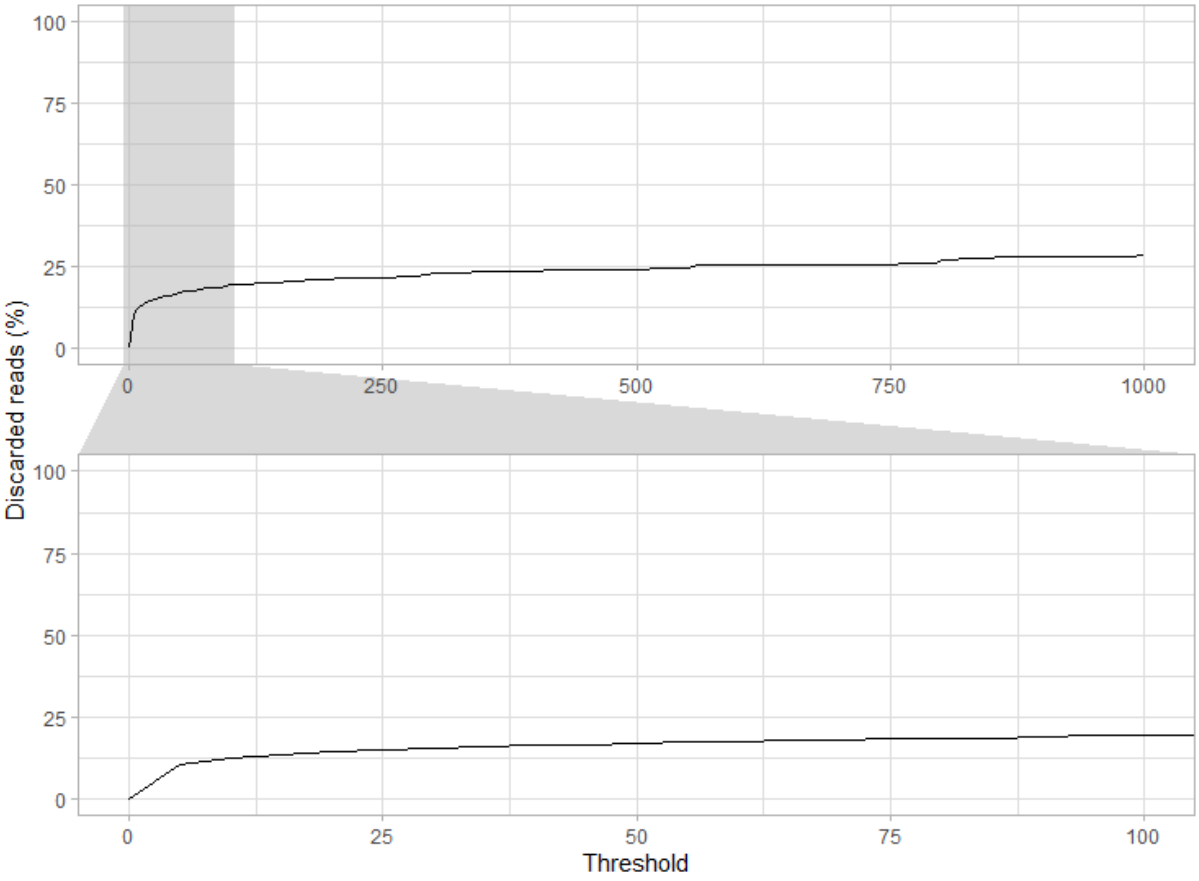

Figure S2A. Discard rate of reads in C666-1. The total number of discarded reads, expressed as percentage of total reads, in relation to threshold value. A threshold value of 10 and 100 corresponds to a discard rate of 13% and 19%, respectively. Hence, at least 81% of the data is considered in subsequent calculations.

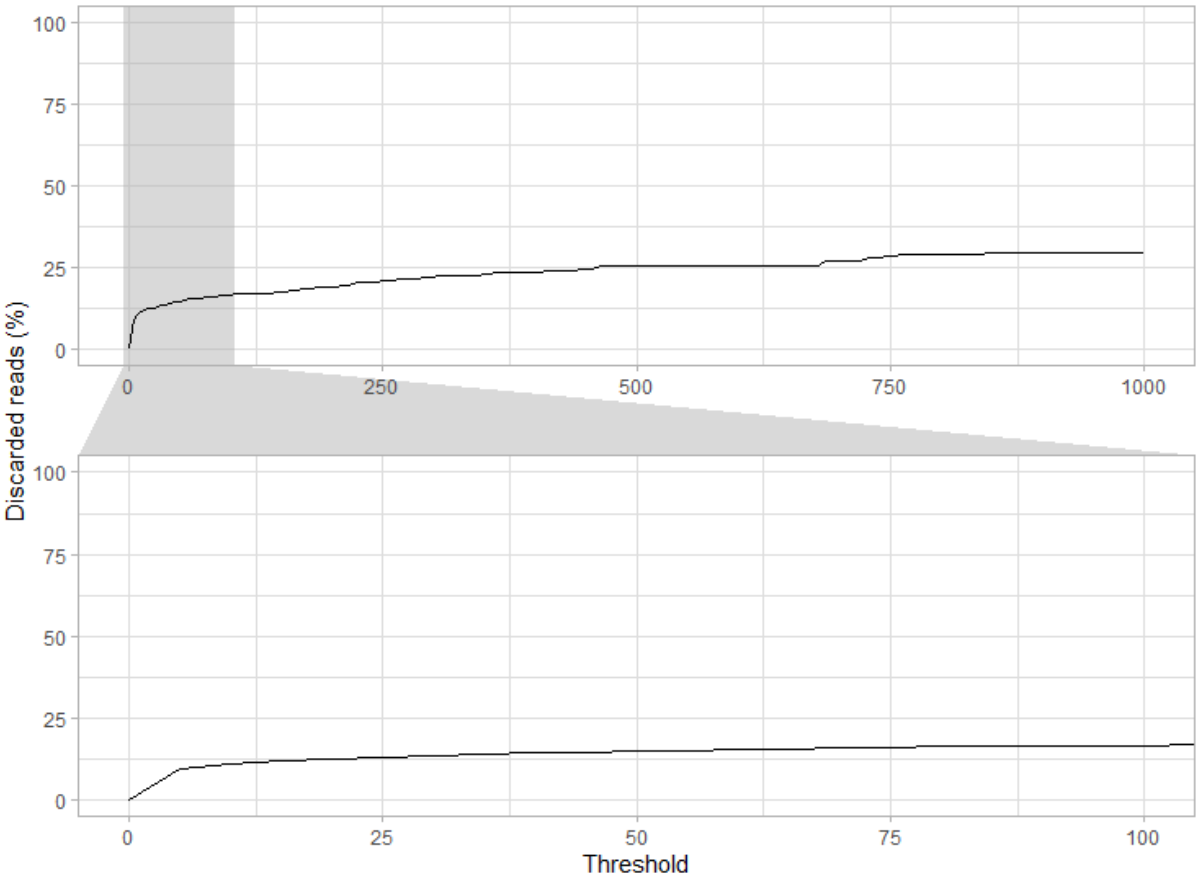

Figure S2B. Discard rate of reads in GAC. The total number of discarded reads, expressed as percentage of total reads, in relation to threshold value. A threshold value of 10 and 100 corresponds to a discard rate of 11% and 17%, respectively. Hence, at least 83% of the data is considered in subsequent calculations.

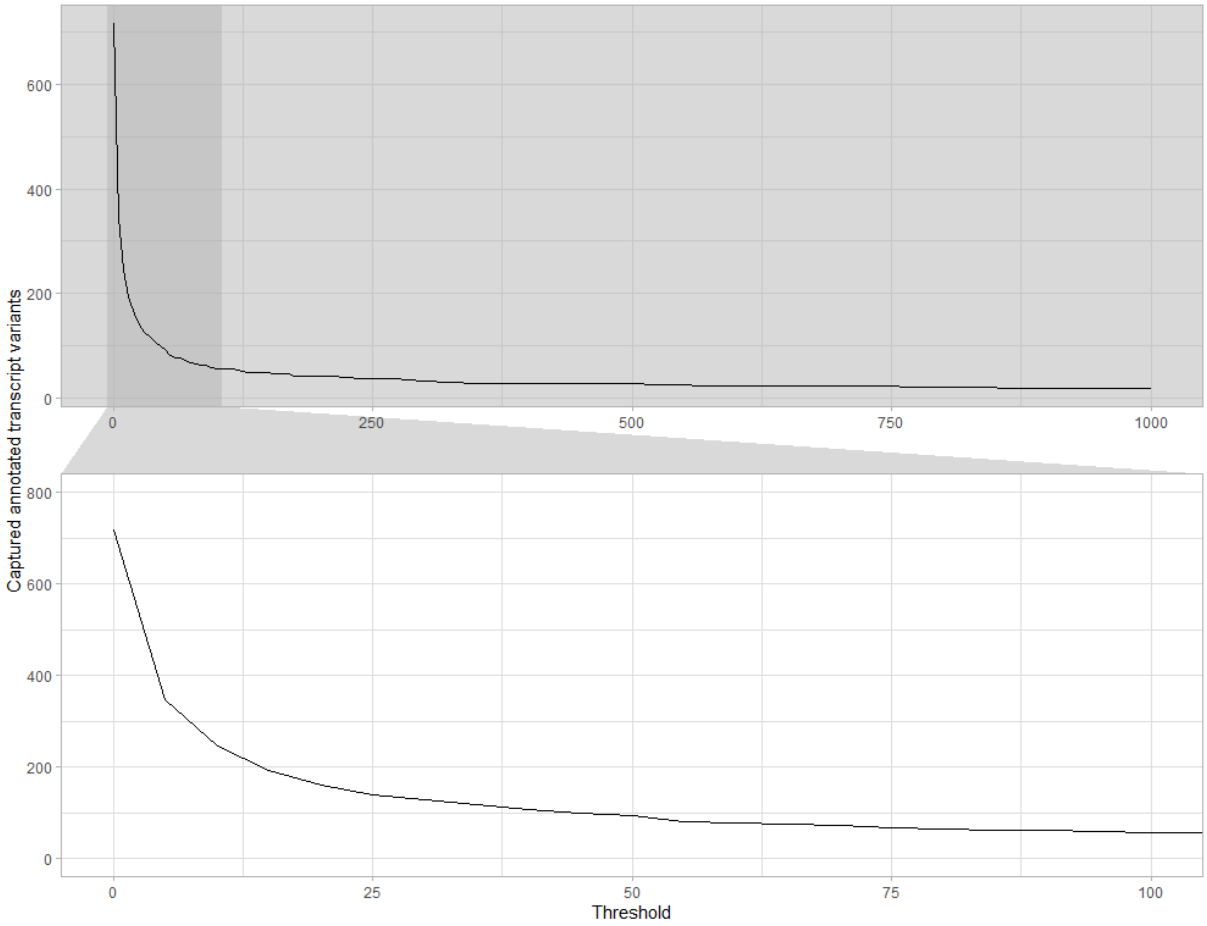

Figure S2C. Capture rate of annotated transcript variants in C666-1. A threshold value of 100 renders 660 annotated transcript variants unavailable for further analysis. However, 72% of these variants are supported by < 10 reads (equivalent to < 0.007% of the total reads).

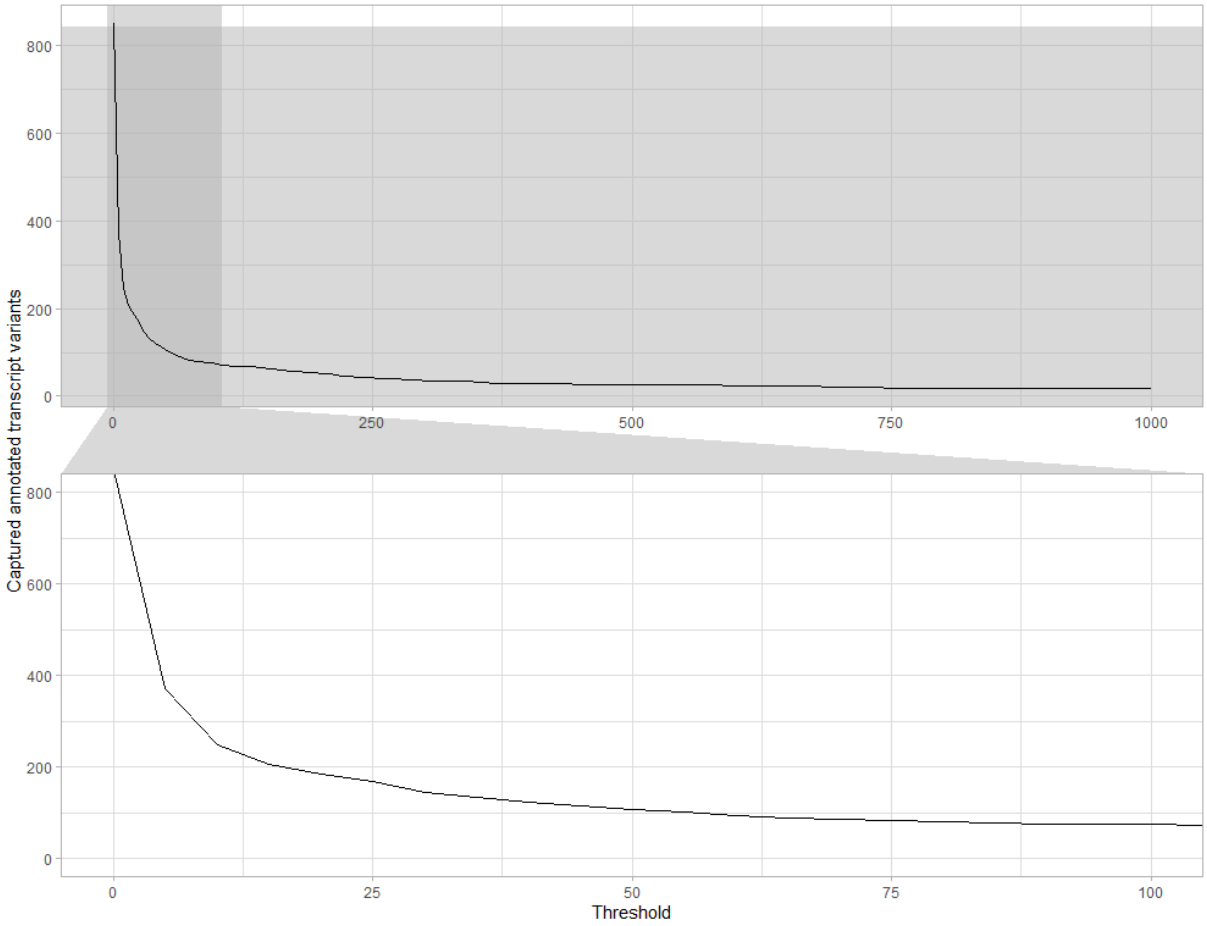

Figure S2D. Capture rate of annotated transcript variants in GAC. A threshold value of 100 renders 776 annotated transcript variants unavailable for further analysis. However, 78% of these variants are supported by < 10 reads (equivalent to < 0.008% of the total reads).

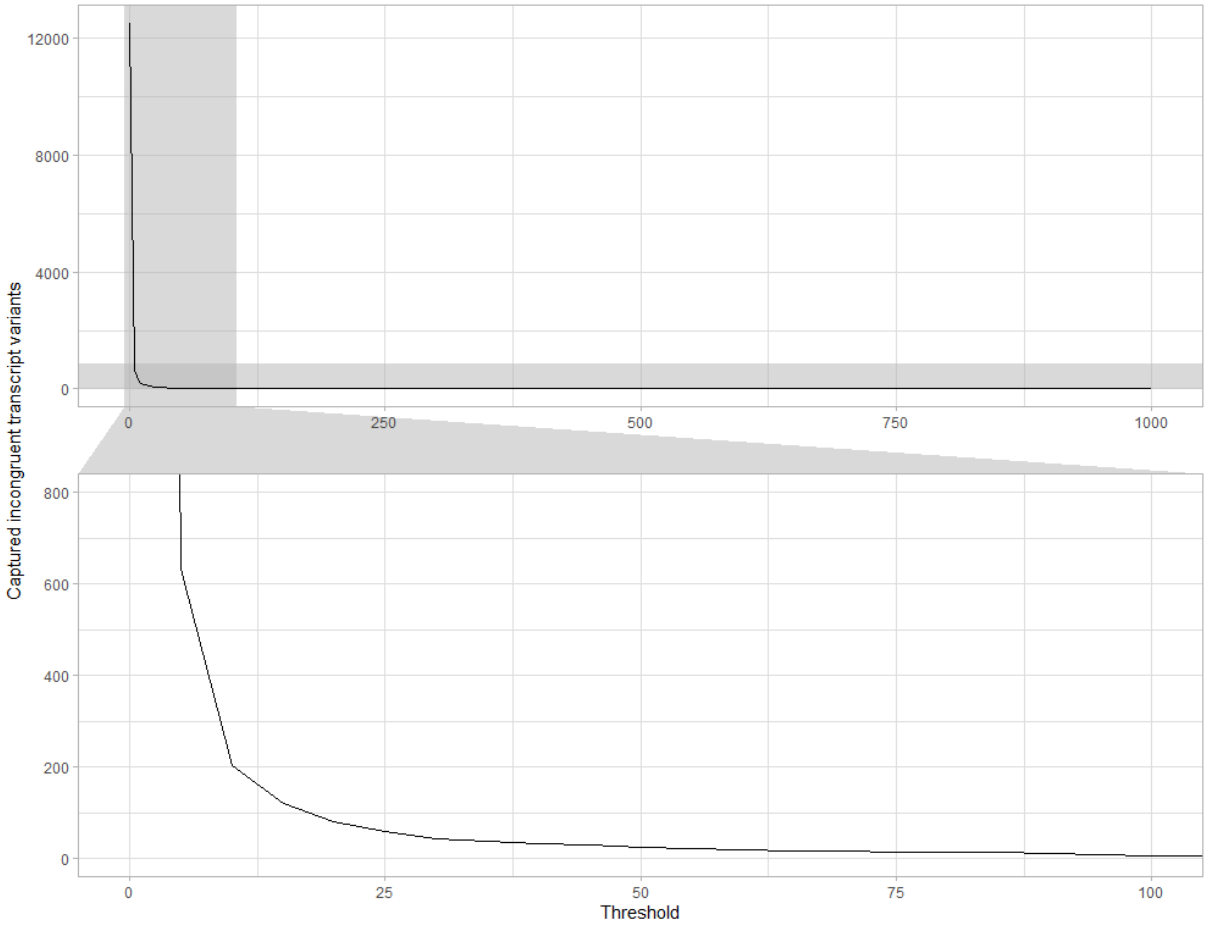

Figure S2E. Capture rate of incongruent transcript variants in C666-1. A threshold value of 100 renders 12 519 incongruent transcript variants unavailable for further analysis. However, 88% of these variants are supported by merely two reads or less. In all probability, the bulk of the disregarded transcript variants are of minor biological relevance.

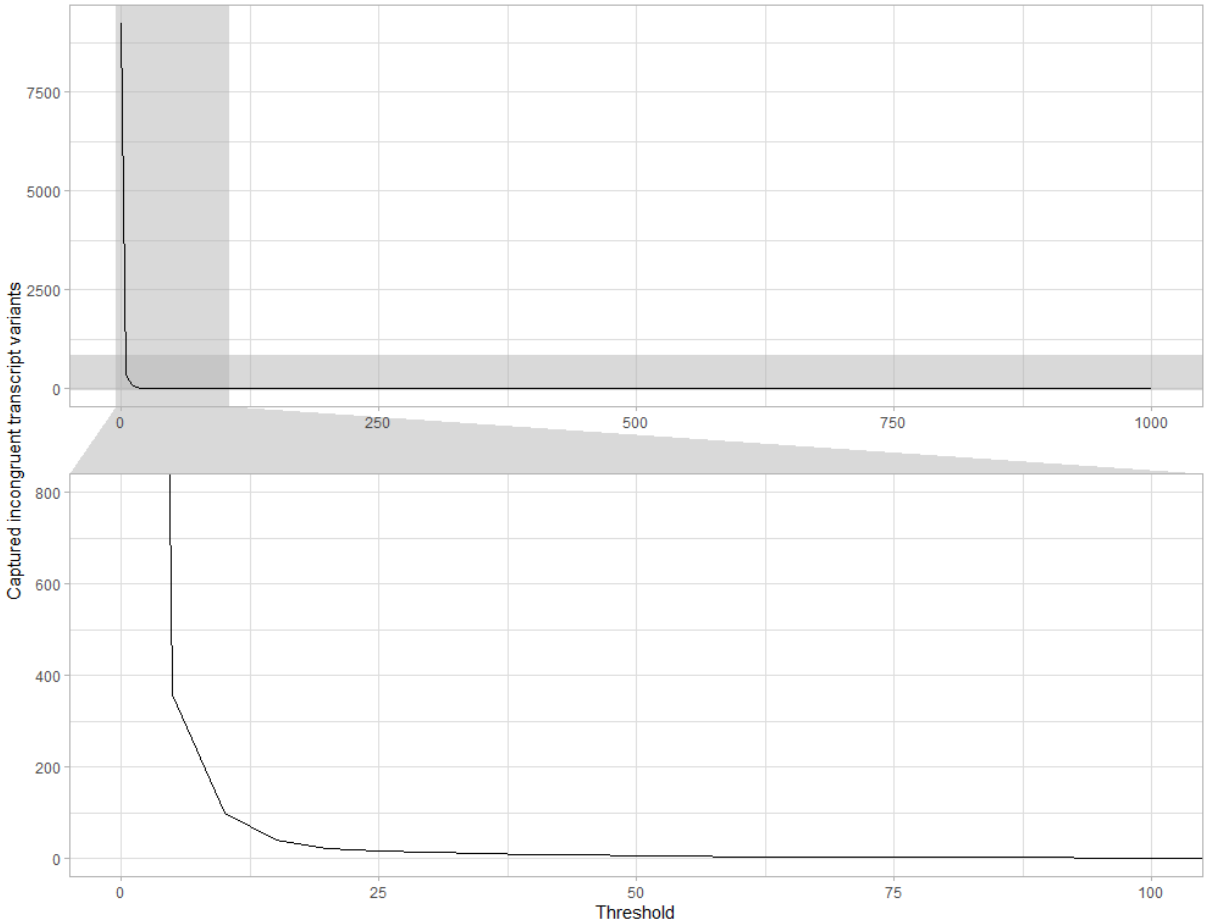

Figure S2F. Capture rate of incongruent transcript variants in GAC. A threshold value of 100 renders 9242 incongruent transcript variants unavailable for further analysis. However, 90% of these variants are supported by merely two reads or less. In all probability, the bulk of the disregarded transcript variants are of minor biological relevance.

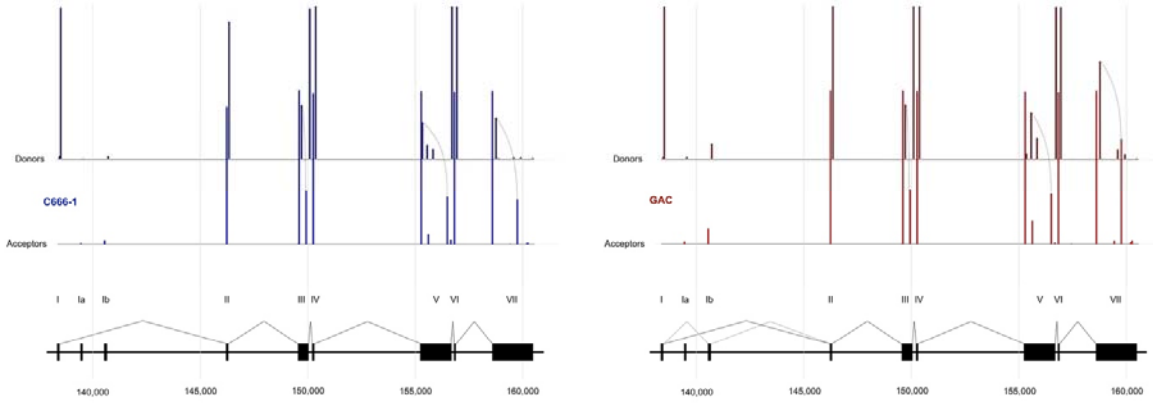

Figure S3. Splice site usage across *RPMS1*. Peaks correspond to the usage frequency of a particular splice site. Dashed lines represent typical alternative splicing events within constitutive exons. The paramount pattern of inter-exonic splice junction connectivity is displayed in the bottom segment. Numbers on the x-axes refer to the global position in the EBV genome.

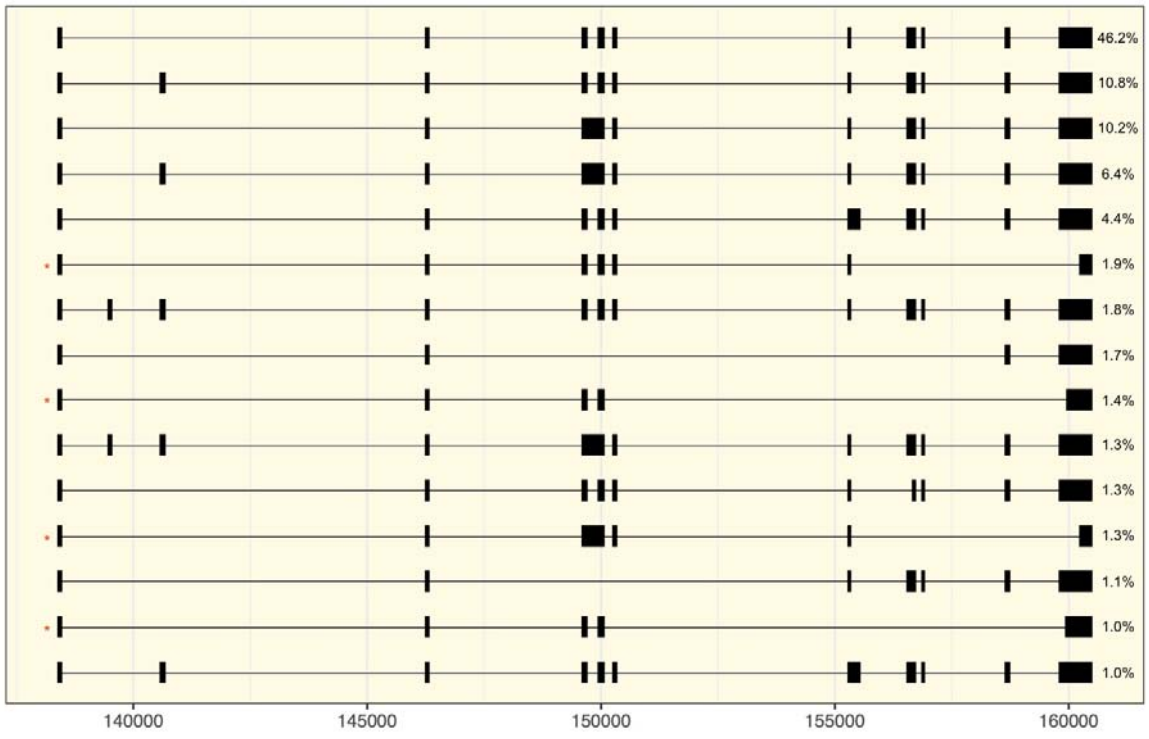

Figure S4. Splice variants of *RPMS1* in Daudi. The 15 most abundant transcript variants of *RPMS1* in Daudi. Numbers indicate the relative abundance, expressed as the percentage of fully annotated and complete high-confidence reads. Asterix (*) marks long-range splice junctions without short-read support. Numbers on the x-axis refer to the global position in the EBV genome.

Figure S5. All splice variants of *RPMS1* called by FLAIR (final annotation). Transcript variants are sorted from top to bottom on the basis of calculated frequency using FLAIR Quantify function.

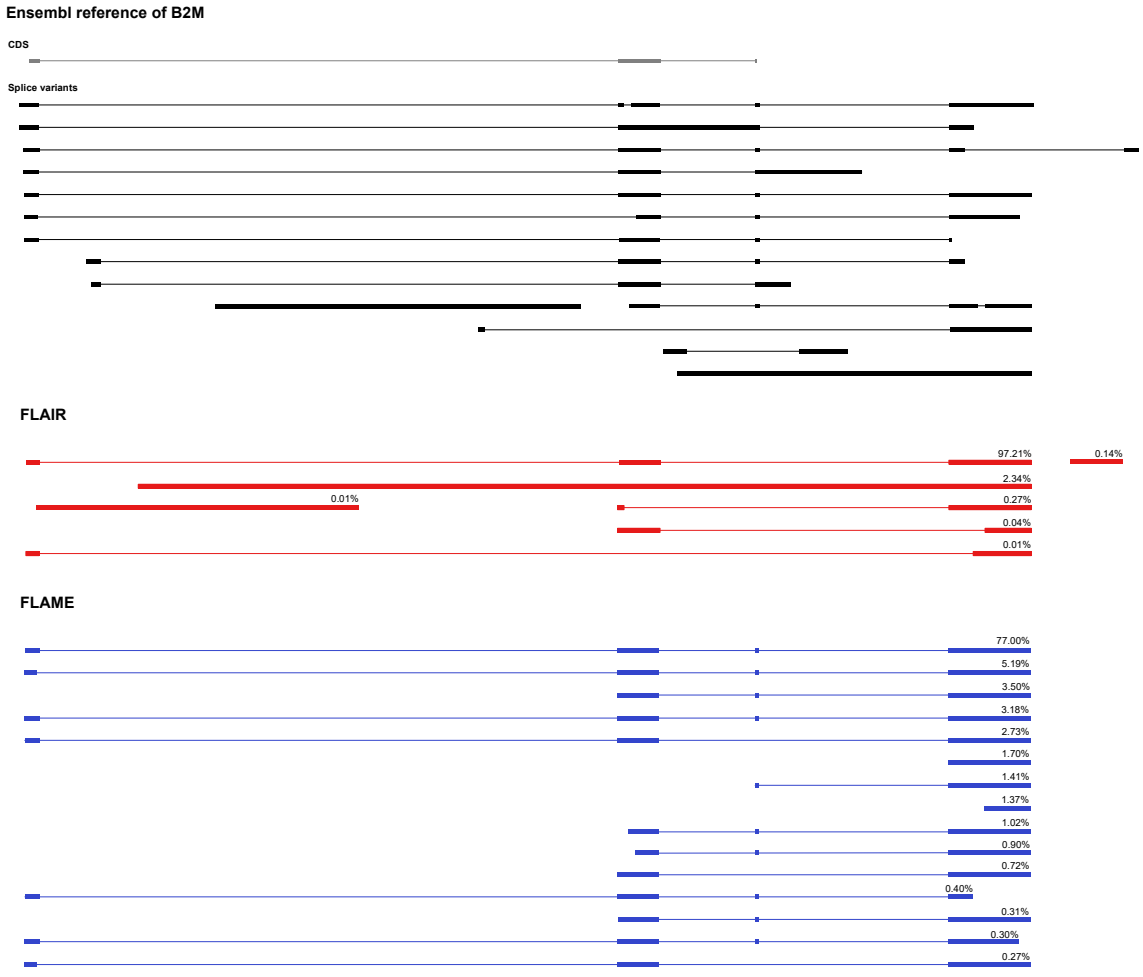

Figure S6. Splicing analysis of *B2M*. The *B2M* reference annotation, as it is defined in the Ensembl genome database project, is delineated in black. The CDS is outlined above in grey. The splice variants assembled by FLAIR are shown in red, whereas the high-confidence splice variants generated by FLAME are shown in blue. Numbers indicate the relative frequency.

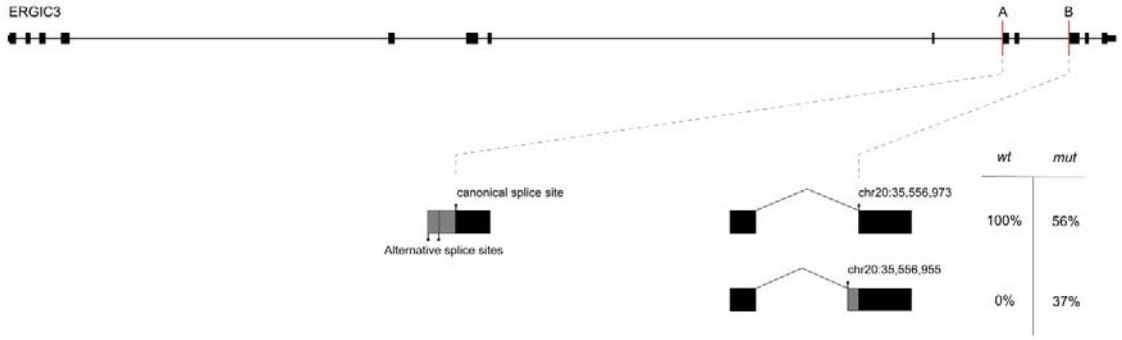

Figure S7. FLAME analysis of alternative 3' splice sites in *ERGIC3*. A and B marks two exons with alternative splice sites within the *ERGIC3* gene. A) Two novel alternative acceptor events (chr20:35,555,972 and chr20:35,555,998) were detected upstream of the canonical chr20:35,556,033. B) The frequency of previously described alternative 3' splice site usage as a result of *SF3B1* mutation. *wt*: wild type *SF3B1*; *mut*: mutant *SF3B1*.