The impact of read depth and read length on RNA-seq splicing analysis

SUPPLEMENTARY MATERIAL

Annika Ladwig[1], Melina Klostermann[1,$], Kathi Zarnack[1,2,$]

Affiliations:

[1] Department of Bioinformatics, Theodor Boveri Institute, Julius Maximilian University Würzburg, Würzburg, Germany

[2] Cluster for Nucleic Acid Sciences and Technologies – NUCLEATE

[$] Corresponding authors

Correspondence:

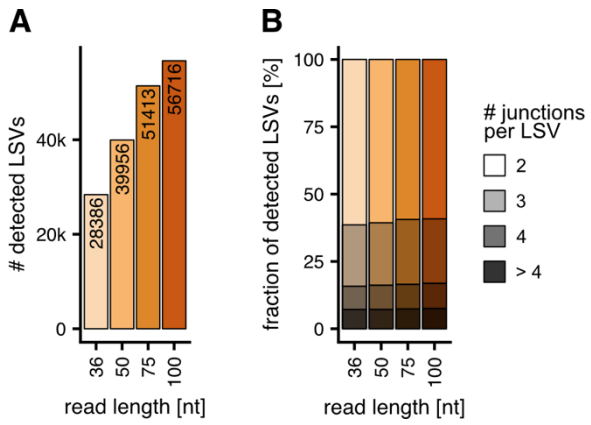
Melina Klostermann (melina.klostermann@uni-wuerzburg.de) & Kathi Zarnack (kathi.zarnack@uni-wuerzburg.de)

	36 nt	50 nt	75 nt	100 nt	125 nt	150 nt
250 M	-	870221	925739	959586	984420	-
200 M	778687	840761	901337	938214	964716	985655
150 M	724971	795289	865355	906873	937546	961110
100 M	638810	721465	805750	856376	890519	915842
75 M	566815	658954	753838	812043	850832	880638
50 M	459568	556604	668029	736906	784978	819056
30 M	321095	416951	537041	620207	675960	719051
25 M	276979	366680	488624	571809	633191	676775
20 M	228381	309350	424258	510273	573053	621582
15 M	172077	240814	344810	429220	492984	544483
10 M	111226	163604	245550	316874	378773	425058
5 M	48404	75568	120705	166198	210196	246260
1 M	7424	11645	18867	27422	34546	41765

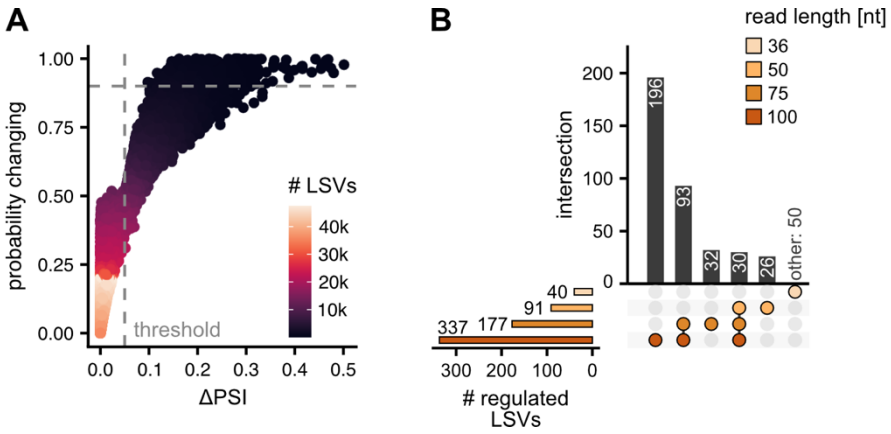
Supplementary Table 1. Numbers of quantifiable introns for combinations of read depth (rows) and read length (columns). Subsampling of an RNA-seq sample in HEK293 cells with originally 380 M reads and a read length of 150 nt is used to estimate the number of introns spanned by at least 10 reads, which we denote as "quantifiable introns".

Read length [nt]	Number of mapped reads
36	245107424
50	255648752
75	265228260
100	267290338
125	258978070
150	241116912

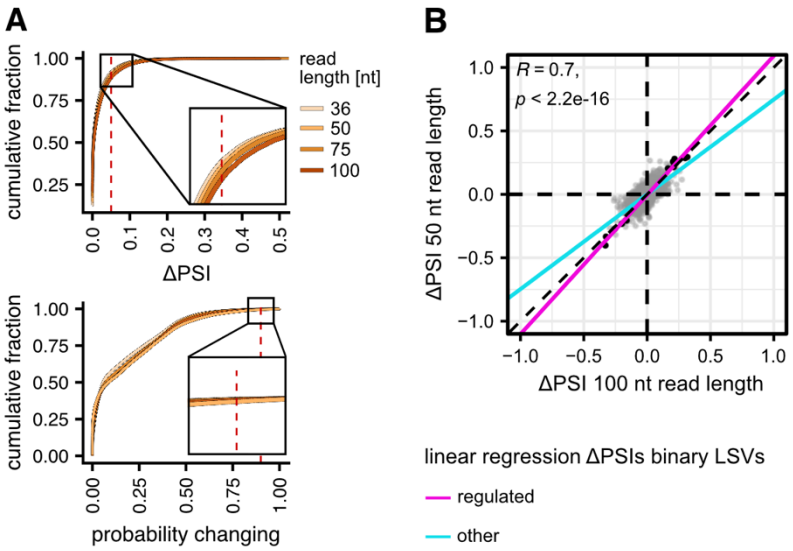
Supplementary Table 2: Number of mapped reads for each read length. Shortening reads of an RNA-seq sample in HEK293 cells with read length of 150 nt to shorter reads influences mapping rates.



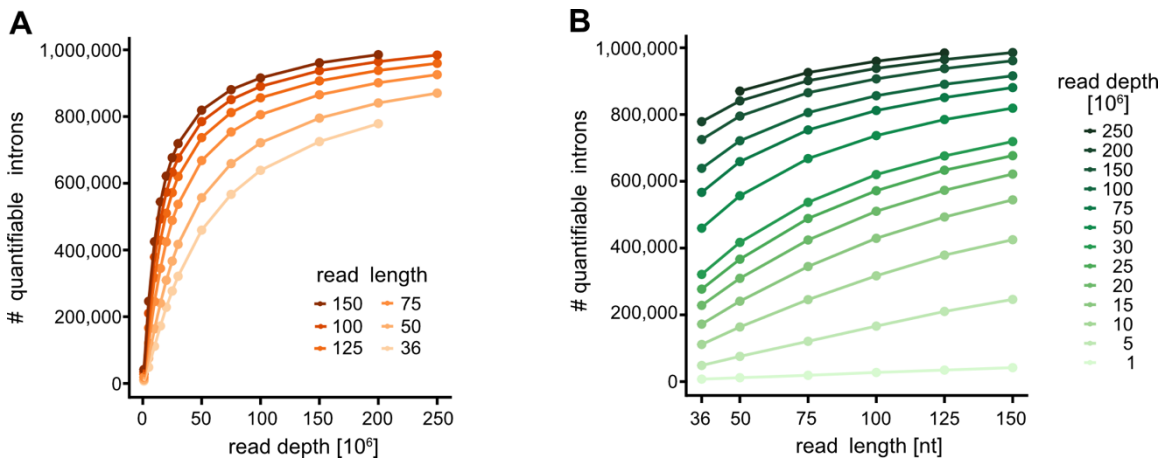
Supplementary Figure 1. Longer reads and a higher read depth allow for better LSV detection. Analyses from the *UPF1* knockdown (KD) experiment. **(A)** Number of LSVs detected by MAJIQ for different read lengths (36–100 nt, light to dark orange). **(B)** LSV complexity. Fraction of LSVs with a given number of junctions therein for different read lengths (36–100 nt, light to dark orange).



Supplementary Figure 2. Differential splicing is identified with greater sensitivity. Analyses from the *UPF1* KD experiment. **(A)** Maximum ΔPSI values are plotted against maximum probability changing (P(|ΔPSI| > 0.05)) for each LSV in the *UPF1* KD experiment. Dashed grey lines show thresholds for an LSV to be considered significantly regulated (ΔPSI ≥ 0.05 and P(|ΔPSI| > 0.05) ≥ 0.9). Colour refers to the point density. **(B)** Impact of read length (36–100 nt, light to dark orange) on the identification of regulated LSVs. UpSet plot shows overlap of regulated LSVs between test data sets.



Supplementary Figure 3. Quantitative estimates are downgraded for non-regulated LSVs. Analyses from the *UPF1* KD experiment. **(A)** Shorter reads lead to smaller ΔPSI estimates. Cumulative fraction of ΔPSI and probability changing values for different read lengths (36–100 nt, light to dark orange. For each LSV, the maximum ΔPSI and probability changing were considered. **(B)** Estimated changes for regulated LSVs are well reproducible. Exemplary comparison of ΔPSI values for shared LSVs between the 100-nt and 50-nt read data sets. LSVs that are regulated in both data sets are shown in black. Magenta and cyan lines represent linear regression of ΔPSI values for regulated and other LSVs, respectively.



Supplementary Figure 4. Theoretical estimation of quantifiable introns in relation to read depth and read length. Subsampling of an RNA-seq sample in HEK cells with originally 380 M reads and 150-nt read length is used to estimate the number of introns spanned by at least 10 reads, which we denote as quantifiable introns. The read depth **(A)** and read length **(B)** are shown against the number of quantifiable introns. Colours mark the different read lengths (A, red & orange scale) or read depth (B, green scale).