EQP Supplementary Information

EQP fragment assignment criterion

The evidence-based assignment criterion of EQP requires that the alignment of a (paired-end) read is compatible with the considered feature. In the following we give a precise definition of this criterion for each feature type. Consider an alignment A of a (single) read against the genome and assume that A consists of the (maximal) intervals $I_1,\ldots,I_n$ ordered from left to right along the genomic axis.

- We say A is compatible with exon E if either $n = 1$ and the genomic interval I_E of E contains I_1 or $n > 1$ and one of the three following conditions is met:
    1. I_E equals one of the intervals I_j for $1 < j < n$.
    2. I_E contains I_1 and the right end point of I_E equals the right end point of I_1.
    3. I_E contains I_n and the left end point of I_E equals the left end point of I_n.
- We say A is compatible with gene G if G contains exons $E_1,\ldots,E_n$ which are compatible with A and whose union contains A.
- We say A is compatible with the exon-exon junction between exons E_1 and E_2 if A is compatible with E_1 and E_2 and the two intervals I_i and I_j of A that are contained in E_1 and E_2 are consecutive, that is, $j = i + 1$ or vice versa.

Now assume that A consists of the alignments A_1 and A_2 of two reads which are aligned as a read pair.

- We say A is compatible with exon E if A_1 or A_2 are compatible with E.
- We say A is compatible with gene G if A_1 and A_2 are compatible with G.
- We say A is compatible with junction J if A_1 or A_2 are compatible with J.

Example for BED intersection

We illustrate how EQP generates the BED file which is the intersection of the BED file derived from the combined SAM file and the BED file containing the exon coordinates. Assume that we have three reads R1, R2, and R3 which are aligned against the transcripts, junctions, and the genome resulting in the following BED entries:

```
ENST00000289963     1268    1368    R1      1       +
ENSG00000132507-junction-09/2-10/3:01:      39      139     R2      0       +
11      64823465        64823565        R3      255     -
```

- For read R1 the best alignments are against the transcripts and one of them is against the interval [1269,1368] on the forward strand of transcript ENST00000289963.
- For read R2 the best alignments are against the junctions and one of them is against the interval [40,139] on the forward strand of the Fasta sequence of length 200bp with identifier "ENSG00000132507-junction-09/2-10/3:01:". The identifier ENSG00000132507-junction-09/2-10/3:01: denotes the junction between exons 09/2 and 10/3 of gene ENSG00000132507 (the

exon numbering is generated by EQP; consecutive sequences of overlapping exons are numbered with a "/"). Exon 09/2 is too short (shorter than 100bp) and padded downstream by exon 01; exon 10/3 is long enough (otherwise the numbers of padding exons would be listed after the second colon of the identifier).

- For read R3 the best alignments are against the genome and one of them is against the interval [64823466,64823565] on the reverse strand of chromosome 11.

The above BED entries are intersected with the entries of a BED file which contains corresponding entries for the exons, for instance,

```
ENST00000289963 1229 1355 8/22527392/22527517/+ 0 +
ENST00000289963 1355 1427 8/22528506/22528577/+ 0 +
ENSG00000132507-junction-09/2-10/3:01: 26 100 17/7311578/7311651/+0 +
ENSG00000132507-junction-09/2-10/3:01: 100 734 17/7311822/7312455/+0 +
11 64823386 64824529 11/64823387/64824529/- 0 -
```

These have the same semantic as the BED entries for the reads except that we now consider exons. Exon identifiers are given by their genomic coordinates, for instance, 8/22527392/22527517/+. By intersecting the two BED files we obtain information about the overlap between the reads and the exons. Here are the intersection BED entries for the reads and the exons:

```
ENST00000289963 1268 1368 R1 1.0 + ENST00000289963 1229 1355
8/22527392/22527517/+ 0 + 87
ENST00000289963 1268 1368 R1 1.0 + ENST00000289963 1355 1427
8/22528506/22528577/+ 0 + 13
ENSG00000132507-junction-09/2-10/3:01: 39 139 R2 0.0 + ENSG00000132507-
junction-09/2-10/3:01: 26 100 17/7311578/7311651/+ 0 + 61
ENSG00000132507-junction-09/2-10/3:01: 39 139 R2 0.0 + ENSG00000132507-
junction-09/2-10/3:01: 100 734 17/7311822/7312455/+ 0 + 39
11 64823465 64823565 R3 255.0 - 11 64823386 64824529
11/64823387/64824529/- 0 - 100
```

So, for instance, the alignment of read R1 overlaps the location of exon 8/22527392/22527517/+ on transcript ENST00000289963 by 87bp and of exon 8/22528506/22528577/+ by 13bp. From this information the (spliced) genomic alignment of read R1 can be easily computed.

Paired-end reads

Paired-end reads impact two steps of the quantification process. Firstly, the information about read pairs is used in the alignment of the reads against the reference sequences. Here, alignments of a read pair in the correct orientation and at a distance which is compatible with the mean insert size should be preferred over alignments with either the wrong orientation or a very large distance between the two reads. Of course, the distance criterion can only be applied to the mapping of the reads against the transcripts (as performed by EQP or Tophat2) but not to a splice-aware mapping against the genome (as

performed by STAR or SpliceMap) since the distance between two reads can also depend on the comparatively large intron sizes if they map to two different exons.

Secondly, the information about read pairs can be used in the quantification step by counting read pairs instead of reads; that is, if both reads of a pair are compatible with a feature, the read pair is counted only once (instead of twice). This is implemented in all considered tools except for QuasR.

Aligner command-line calls

Tophat2 call (version v2.0.8b)

```
tophat2 <options> <Bowtie2 genome index> <Fastq file 1> <Fastq file1>
```

Options:

```
-p 6 --max-multihits 100 --mate-inner-dist 0 -o <output dir> --b2-
sensitive --microexon-search --transcriptome-index <Bowtie2 transcript
index> --prefilter-multihits
```

The Bowtie2 transcript index is a prebuilt index created by running Tophat2 on the Ensembl v76 GTF file augmented with GTF entries for the custom junction entries created by EQP to provide the greatest amount of reference information for Tophat2.

STAR call (version 2.3.0)

```
STAR --genomeDir <STAR genome index> --runThreadN 6 --readFilesIn
<Fastq file 1> <Fastq file 2> --genomeLoad LoadAndRemove --
outFilterMultimapNmax 100 --outFilterMismatchNmax 6 --
outFilterMismatchNoverLmax 0.2
```

The STAR genome index was built using a GTF file:

```
STAR --runMode genomeGenerate --genomeDir <STAR genome index> --
genomeFastaFiles <genome Fasta file> --sjdbGTFfile <Ensembl GTF file>
--sjdbOverhang 99
```

We only used the Ensembl GTF file. If junction GTF entries which are created by EQP are added to the Ensembl GTF file, then the size of the STAR index increases to ~35GB. We tested the alignment with and without the junction GTF entries on one sample (SEQC-A-BC01-s_1):

	Aligned reads	Exonic reads	Num. introns	Soft masked bases
Ensembl GTF file	9902168	8600456	219140	24194087
Ensembl and junction GTF file	9883014	8588842	260612	24176005

It is surprising that the number of aligned reads as well as the number of reads that map to exons is reduced if the junction GTF entries are added, on the other hand, the sensitivity for spliced is clearly increased.

SpliceMap call

We make use of the SpliceMap implementation available in QuasR and call SpliceMap directly from R:

```
qAlign(<Fastq annotation file>, genome=<genome Fasta file>,
alignmentsDir=<output SAM dir>, cacheDir=<temp dir>,
splicedAlignment=TRUE)
```

Comparison of uniquely aligned reads

Since SpliceMap considers only uniquely aligned reads, we also present data on the percent of uniquely aligned and uniquely aligned exonic reads for the other aligners.

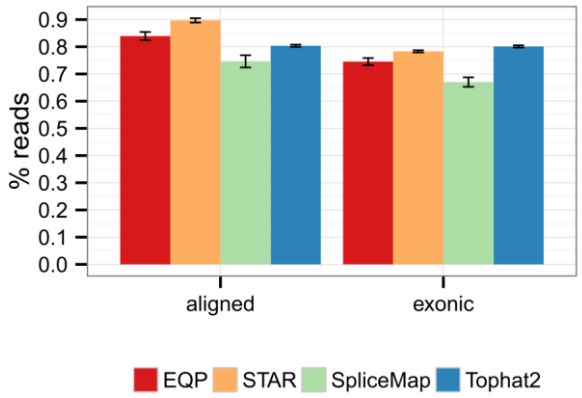


Figure 1 The percent of uniquely aligned and uniquely aligned exonic reads for the aligners Bowtie2/EQP, STAR, SpliceMap, and Tophat2. The numbers for SpliceMap are the same as previously shown.

As can be seen the numbers are generally lower, yet SpliceMap still displays the least sensitivity as compared to the other aligners.

Aligner execution times

The following figure shows the run times of the different alignment tools across the 64 Fastq files of the SEQC data set.

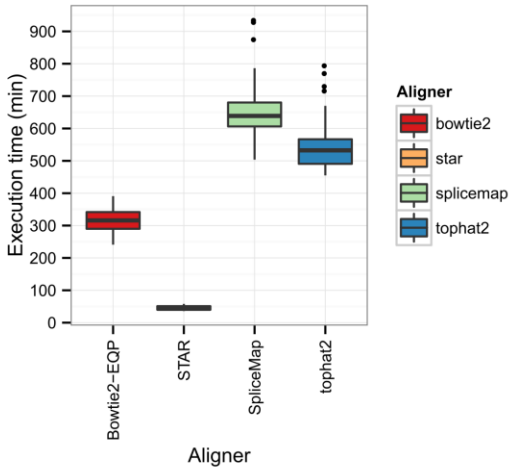


Figure 2 The run time of the four different alignment tools (Bowtie2-EQP, STAR, SpliceMap, Tophat2) in minutes when being run with six cores in the Novartis cluster environment (for SpliceMap, eight cores were used). The alignment time of Bowtie2-EQP is the sum of the alignment times against the genome, the transcripts, and the junctions.

As can be seen from Figure 2 STAR is by far the fastest aligner with a mean of about 45min, followed by Bowtie2-EQP with a mean runtime of about 5h 15min, Tophat2 with about 9h and, finally, SpliceMap with about 10h 50min.

Quantifier calls

Gene quantification

htseq-count call for mode union, intersection-nonempty, or intersection-strict (v. HTSeq 0.6.1.p2)

```
htseq-count -m <mode> -i gene_id -a 0 -s no  <SAM file> <GTF file>
```

featureCounts call (v1.4.5-p1)

```
featureCounts -a <GTF file>  -g gene_id  -p -o <count file> <BAM file>
```

Cufflinks call (v2.1.1)

```
cufflinks -p 2 -o <output dir> -G <GTF file> -b <Genome file> -u <BAM file>
```

QuasR call (v.1.6.2)

```
proj <- qAlign(<BAM annotation file>, genome=<genome file>,
alignmentsDir=<align dir>, cacheDir=<cache dir>, splicedAlignment=TRUE,
paired="fr")
chrLen <- scanFaIndex(<genome file>)
chrominfo <- data.frame(chrom=as.character(seqnames(chrLen)),
                        length=width(chrLen),
                        is_circular=rep(FALSE, length(chrLen)))
txdb <- makeTranscriptDbFromGFF(file=<GTF file>, format="gtf",
                                exonRankAttributeName="exon_number",
                                gffGeneIdAttributeName="gene_name",
```

```
                                    chrominfo=chrominfo,
                                    dataSource="Ensembl",
                                    species="Homo sapiens")
geneLevels <- qCount(proj, txdb, reportLevel="gene")
```

Quantifier execution times

The following figure shows the run times of the different alignment tools across the 64 Fastq files of the SEQC data set.

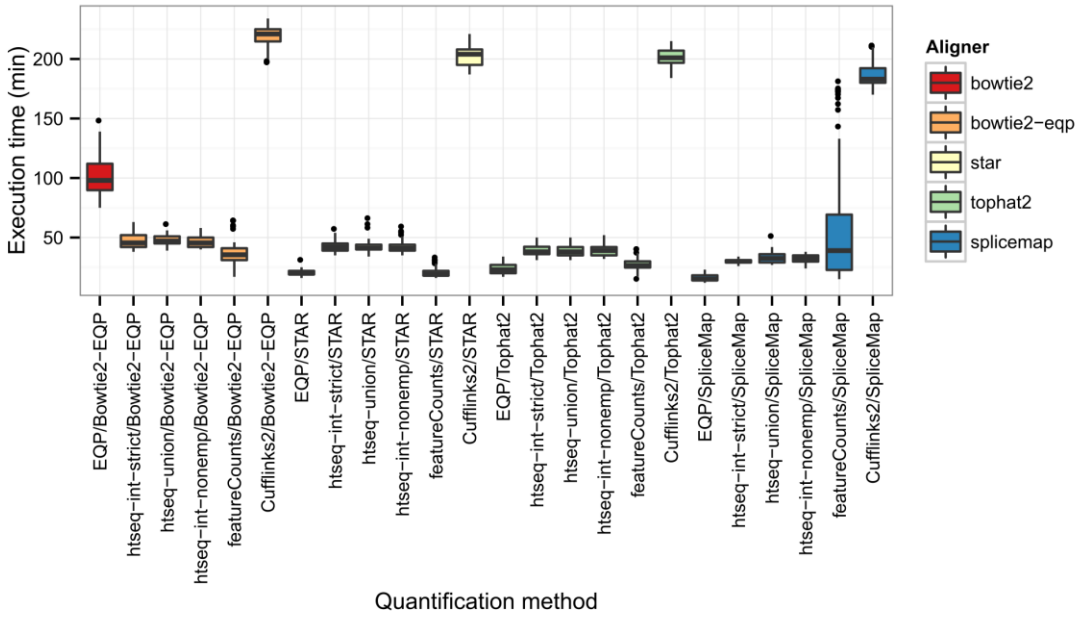


Figure 3 Boxplots of the run times of the quantifiers in minutes grouped by aligners.

In Table 1 the median run times for each combination of quantifier and aligner is listed. On average EQP and htSeq need about 40 minutes, featureCounts about 25 minutes – if the run times for the SpliceMap alignments are considered as outliers – and Cufflinks2 about 3h 20 min. If the genome alignment file is provided, then EQP is very efficient and comparable to featureCounts with about 20 minutes on average. The high execution time of EQP on its own alignment module is due to the fact that EQP needs to process and combine three alignment files (the genome, transcript, and junction alignment file) before the actual quantification can be started.

	Bowtie2-EQP	**STAR**	**Tophat2**	**SpliceMap**
EQP	100.55	20.48	23.78	16.30
htseq-int-strict	47.53	42.69	38.94	30.17
htseq-union	47.53	42.27	39.14	33.00
htseq-int-nonemp	46.30	42.31	39.39	32.14

featureCounts	36.67	20.80	27.30	59.14
Cufflinks2	218.75	202.59	200.69	187.77

Table 1 The median run times for quantifier/aligner combinations.

Computation of Taqman log fold change values

In order to be able to also assign a fold change value to sample pairs in which the count(s) in one (or both) sample(s) are zero we add a pseudo count of 0.1 to the RNA-seq derived gene counts (which corresponds to about one assigned fragment). In order to be able to use the same pseudo-count also for the FPKM values provided by Cufflinks, the Cufflinks gene abundance estimates are converted to CPMs by multiplying the values with the gene length in kilo bases for each gene.

Note that the value of 0.1 for the pseudo count is quite low and should not lead to artificially inflated correlation values as most variation in the data is maintained.

The 32 fold change values for the different RNA-seq quantification methods are chosen based on the lane pairs SEQC-A-BC<n>-s_<m> and SEQC-B-BC<n+1>-s_<m> where <n> is a barcode number (01, 02, 05, 06, etc.) and <m> is a lane number (1, 2, …, 8), for instance, the first pair used for the calculation of a fold change is SEQC-A-BC01-s_1/ SEQC-B-BC02-s_1 and the last pair is SEQC-A-BC13-s_8/ SEQC-B-BC14-s_8.

Comparison of Taqman log fold changes with the coefficient of determination

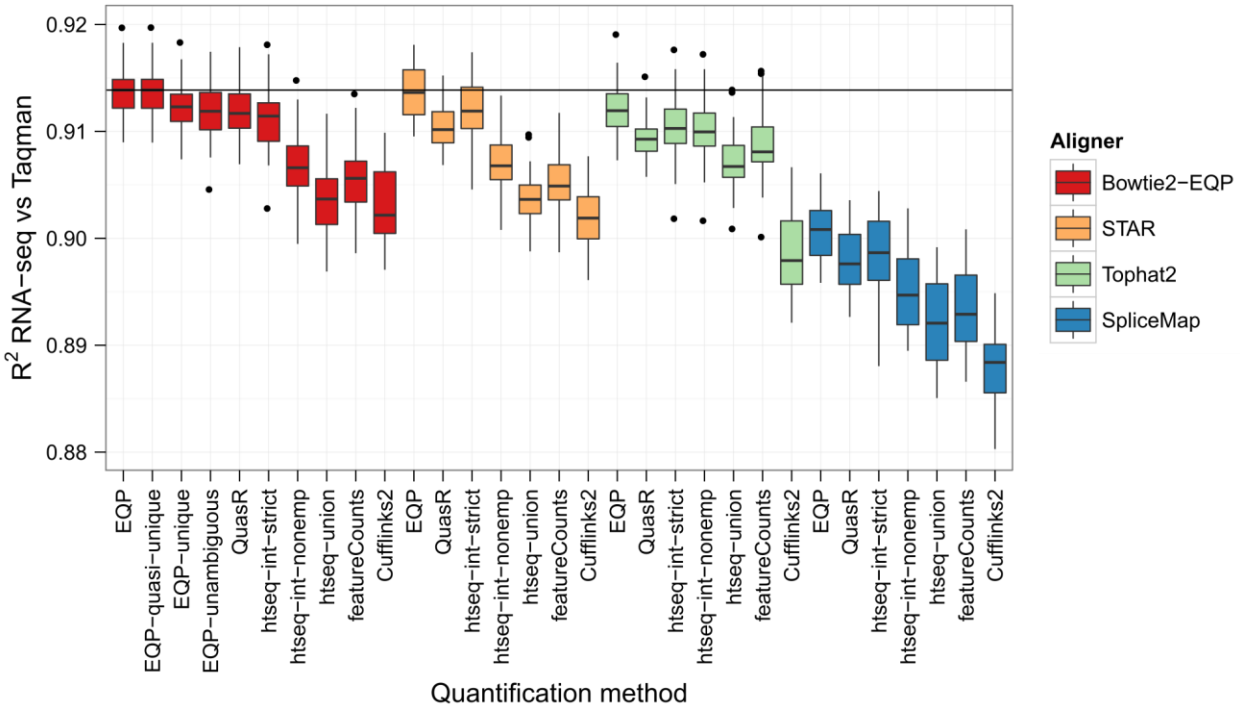
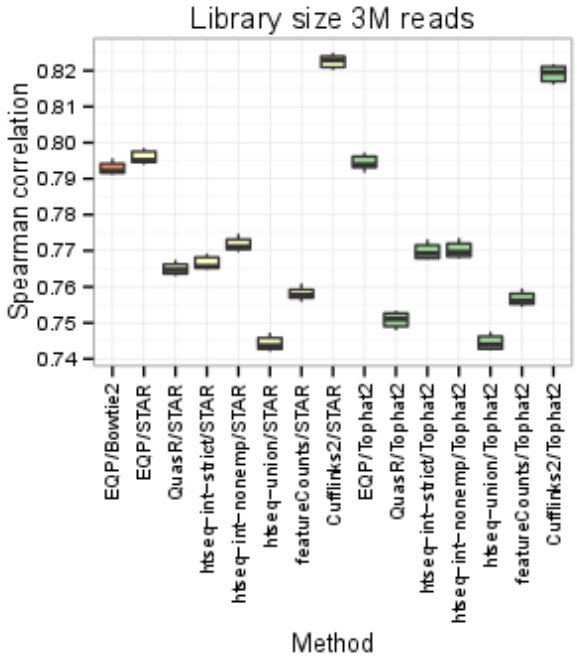
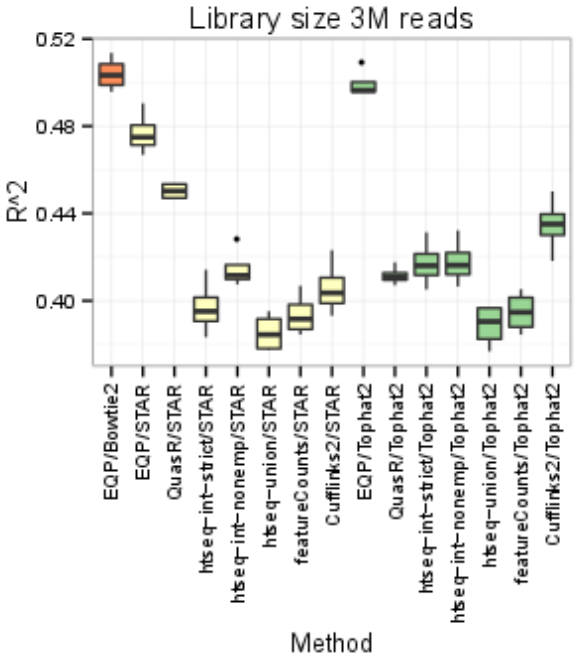


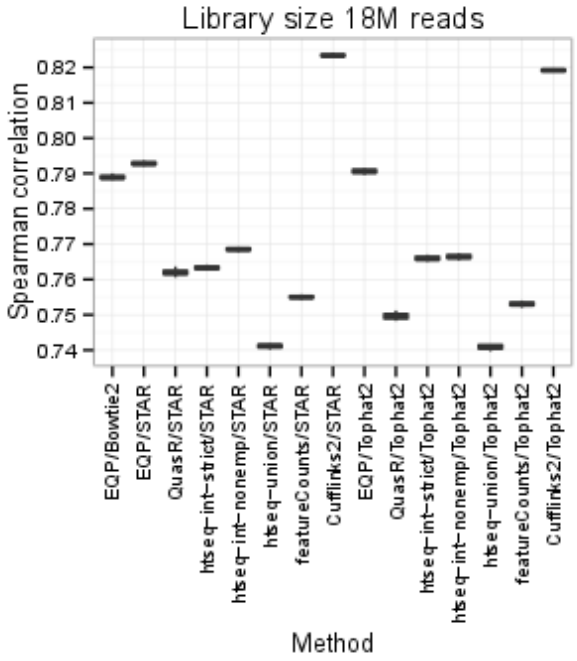
Figure D Comparison of log fold changes to Taqman log fold changes for different alignment and quantification methods using the coefficient of determination (R^2).

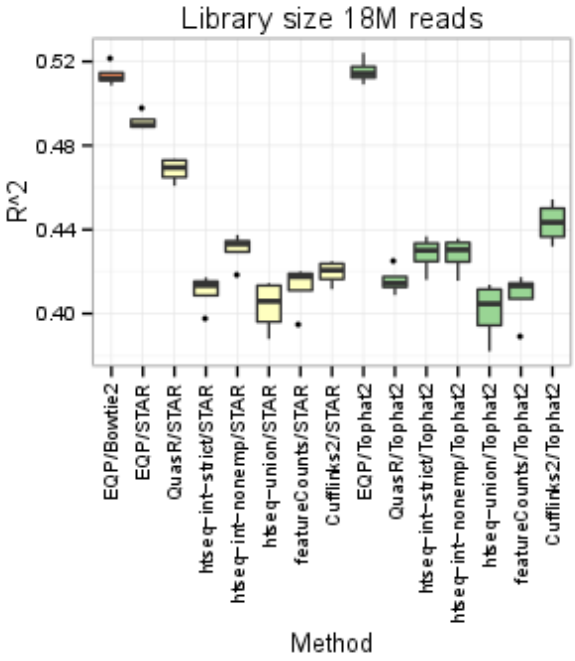
The general ranking of aligners and quantification methods with respect to the coefficient of determination is very similar to the ranking given by RMSD with the exception of Cufflinks2 which now performs slightly. This is in agreement with the results shown in (1) where also transcript abundance estimation methods fared considerably better with RMSD than with the correlation coefficient compared to htseq-count.

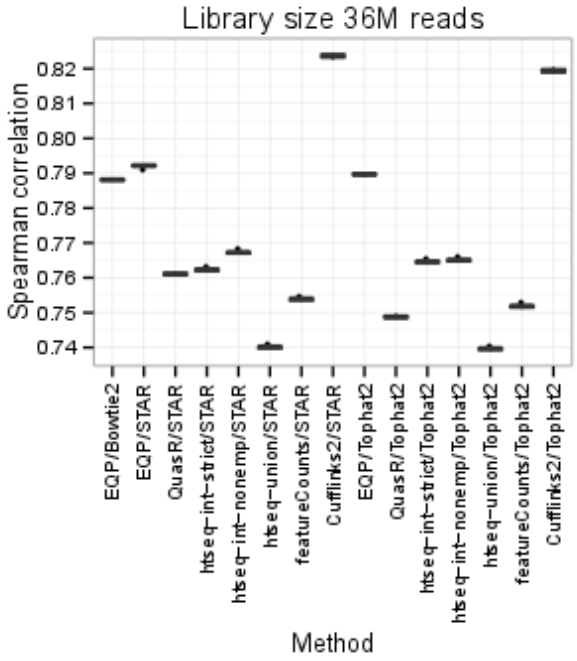
Simulated RNA-seq data: Spearman correlation and log FC R^2 values

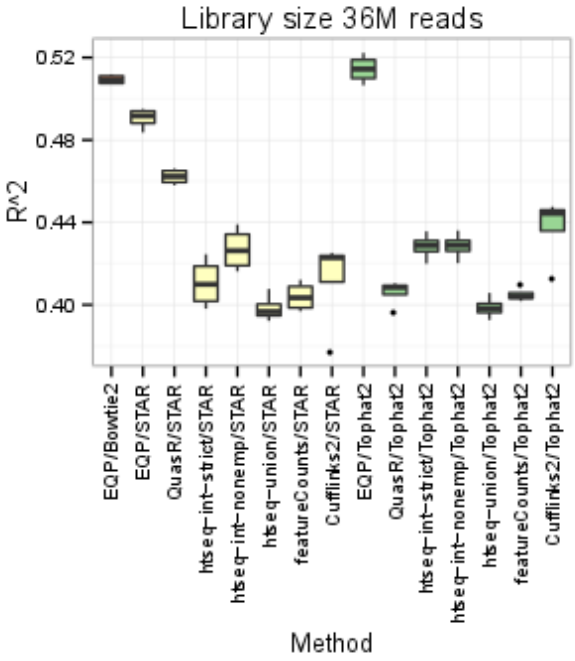




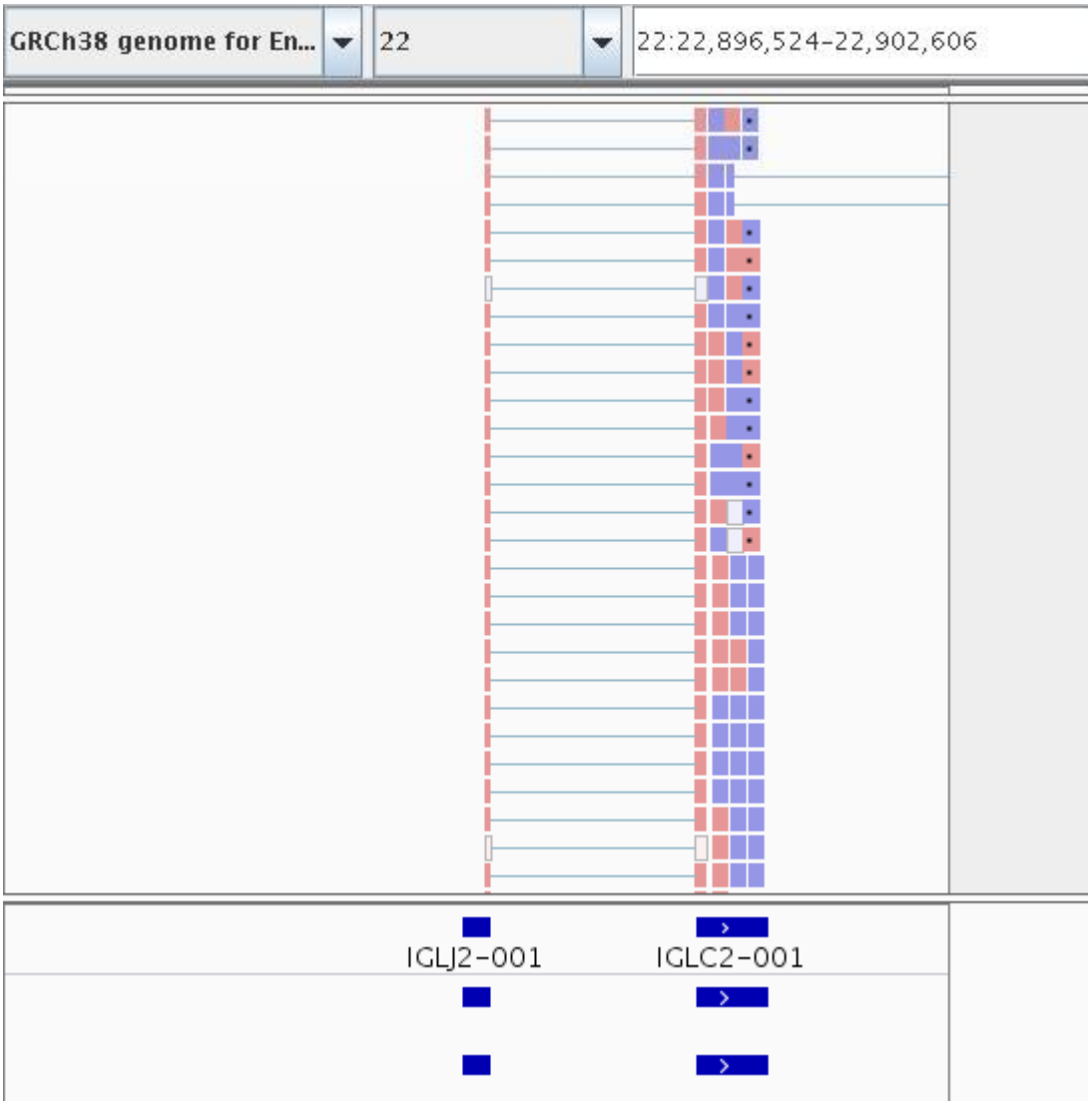





Library size 36M reads
Spearman correlation
0.82
0.81
0.80
0.79
0.78
0.77
0.76
0.75
0.74
EQP/Bowtie2
EQP/STAR
QuasR/STAR
htseq-int-strict/STAR
htseq-int-nonemp/STAR
htseq-union/STAR
featureCounts/STAR
Cufflinks2/STAR
EQP/Tophat2
QuasR/Tophat2
htseq-int-strict/Tophat2
htseq-int-nonemp/Tophat2
htseq-union/Tophat2
featureCounts/Tophat2
Cufflinks2/Tophat2
Method


Library size 36M reads
R^2
0.52
0.48
0.44
0.40
EQP/Bowtie2
EQP/STAR
QuasR/STAR
htseq-int-strict/STAR
htseq-int-nonemp/STAR
htseq-union/STAR
featureCounts/STAR
Cufflinks2/STAR
EQP/Tophat2
QuasR/Tophat2
htseq-int-strict/Tophat2
htseq-int-nonemp/Tophat2
htseq-union/Tophat2
featureCounts/Tophat2
Cufflinks2/Tophat2
Method

An example of a non-overlapping exon with differing exon counts between EQP and dexseq-count



The two genes IGLJ2 and IGC2-001 consist of single exons and are connected by a number of reads. However, since the reads align to two genes they are not counted by dexseq-count.

Junction counts for the exons E_1 and E_2

The Ensembl exon ids for the exons E_1 and E_2 of the main text are ENSE00002290307 and ENSE00003736927. We show here the non-zero counts of junctions for eight upstream exons for which the exons E_1 or E_2 being predicted to be the splice acceptor for the STAR alignments of sample SEQC-A-BC01 lane s_1.

Junction Id with E_1 (ENSE00002290307) as splice acceptor	Count
Ensembl|junction|ENSE00001816391|ENSE00002290307	4
Ensembl|junction|ENSE00001202935|ENSE00002290307	4

Ensembl|junction|ENSE00002247518|ENSE00002290307	4

Junction Id with E_2 (ENSE00003736927) as splice acceptor	Count
Ensembl|junction|ENSE00001202935|ENSE00003736927	289
Ensembl|junction|ENSE00001511597|ENSE00003736927	217
Ensembl|junction|ENSE00001675096|ENSE00003736927	285
Ensembl|junction|ENSE00001816391|ENSE00003736927	289
Ensembl|junction|ENSE00002203018|ENSE00003736927	266
Ensembl|junction|ENSE00002228967|ENSE00003736927	286
Ensembl|junction|ENSE00002237874|ENSE00003736927	282
Ensembl|junction|ENSE00002247518|ENSE00003736927	289

It can be clearly seen that EQP predicts almost no junctions for E_1 and a high number of junctions for E_2. Note that reads are counted for duplicate junctions if their splicing pattern is consistent with those multiple exons that share the same boundary.

In QuasR the junction upstream of E_1 and E_2 is represented by the intron 12:130872127-130872583 and the counts are reported separately for the plus and minus strand:

Junction Id	Count
12:130872127-130872583:+	315
12:130872127-130872583:-	64

The difference in the reported counts (315 + 64 vs 289) is explained by the following observations:

1. QuasR reports read counts whereas EQP reports fragment counts, that is, if both the forward and reverse read cross the junction, EQP only reports one count whereas QuasR reports a count for the plus and a count for the minus strand.
2. In QuasR an overlap of one base is considered sufficient to report a spanning read whereas EQP requires an overlap of eight bases (or the length of the exon for exons shorter than eight nucleotides). There are 26 reads which overlap the upstream exon of the junction with less than eight nucleotides.

GEO qRT-PCR Taqman samples

GEO series	GEO sample	GEO sample name	Sample name	Sample
GSE5350	GSM129638	MAQC_TAQ_1_A1	SEQC-A-BC01	UHRR
GSE5350	GSM129639	MAQC_TAQ_1_A2	SEQC-A-BC05	UHRR
GSE5350	GSM129640	MAQC_TAQ_1_A3	SEQC-A-BC09	UHRR
GSE5350	GSM129641	MAQC_TAQ_1_A4	SEQC-A-BC13	UHRR
GSE5350	GSM129642	MAQC_TAQ_1_B1	SEQC-B-BC02	HBR
GSE5350	GSM129643	MAQC_TAQ_1_B2	SEQC-B-BC06	HBR
GSE5350	GSM129644	MAQC_TAQ_1_B3	SEQC-B-BC10	HBR
GSE5350	GSM129645	MAQC_TAQ_1_B4	SEQC-B-BC14	HBR

GEO RNA-seq samples

GEO series	GEO sample	GEO sample name	Sample name	Sample
GSE47774	GSM1157925	SEQC_ILM_NVS_A_1_L01_ATCACG_AD0902ACXX	SEQC-A-BC01-s_1	UHRR
GSE47774	GSM1157927	SEQC_ILM_NVS_A_1_L02_ATCACG_AD0902ACXX	SEQC-A-BC01-s_2	UHRR
GSE47774	GSM1157929	SEQC_ILM_NVS_A_1_L03_ATCACG_AD0902ACXX	SEQC-A-BC01-s_3	UHRR
GSE47774	GSM1157931	SEQC_ILM_NVS_A_1_L04_ATCACG_AD0902ACXX	SEQC-A-BC01-s_4	UHRR
GSE47774	GSM1157933	SEQC_ILM_NVS_A_1_L05_ATCACG_AD0902ACXX	SEQC-A-BC01-s_5	UHRR
GSE47774	GSM1157935	SEQC_ILM_NVS_A_1_L06_ATCACG_AD0902ACXX	SEQC-A-BC01-s_6	UHRR
GSE47774	GSM1157937	SEQC_ILM_NVS_A_1_L07_ATCACG_AD0902ACXX	SEQC-A-BC01-s_7	UHRR
GSE47774	GSM1157939	SEQC_ILM_NVS_A_1_L08_ATCACG_AD0902ACXX	SEQC-A-BC01-s_8	UHRR
GSE47774	GSM1157941	SEQC_ILM_NVS_A_2_L01_ACAGTG_AD0902ACXX	SEQC-A-BC05-s_1	UHRR
GSE47774	GSM1157943	SEQC_ILM_NVS_A_2_L02_ACAGTG_AD0902ACXX	SEQC-A-BC05-s_2	UHRR
GSE47774	GSM1157945	SEQC_ILM_NVS_A_2_L03_ACAGTG_AD0902ACXX	SEQC-A-BC05-s_3	UHRR
GSE47774	GSM1157947	SEQC_ILM_NVS_A_2_L04_ACAGTG_AD0902ACXX	SEQC-A-BC05-s_4	UHRR
GSE47774	GSM1157949	SEQC_ILM_NVS_A_2_L05_ACAGTG_AD0902ACXX	SEQC-A-BC05-s_5	UHRR
GSE47774	GSM1157951	SEQC_ILM_NVS_A_2_L06_ACAGTG_AD0902ACXX	SEQC-A-BC05-s_6	UHRR
GSE47774	GSM1157953	SEQC_ILM_NVS_A_2_L07_ACAGTG_AD0902ACXX	SEQC-A-BC05-s_7	UHRR
GSE47774	GSM1157955	SEQC_ILM_NVS_A_2_L08_ACAGTG_AD0902ACXX	SEQC-A-BC05-s_8	UHRR
GSE47774	GSM1157957	SEQC_ILM_NVS_A_3_L01_GATCAG_AD0902ACXX	SEQC-A-BC09-s_1	UHRR
GSE47774	GSM1157959	SEQC_ILM_NVS_A_3_L02_GATCAG_AD0902ACXX	SEQC-A-BC09-s_2	UHRR
GSE47774	GSM1157961	SEQC_ILM_NVS_A_3_L03_GATCAG_AD0902ACXX	SEQC-A-BC09-s_3	UHRR
GSE47774	GSM1157963	SEQC_ILM_NVS_A_3_L04_GATCAG_AD0902ACXX	SEQC-A-BC09-s_4	UHRR
GSE47774	GSM1157965	SEQC_ILM_NVS_A_3_L05_GATCAG_AD0902ACXX	SEQC-A-BC09-s_5	UHRR
GSE47774	GSM1157967	SEQC_ILM_NVS_A_3_L06_GATCAG_AD0902ACXX	SEQC-A-BC09-s_6	UHRR
GSE47774	GSM1157969	SEQC_ILM_NVS_A_3_L07_GATCAG_AD0902ACXX	SEQC-A-BC09-s_7	UHRR
GSE47774	GSM1157971	SEQC_ILM_NVS_A_3_L08_GATCAG_AD0902ACXX	SEQC-A-BC09-s_8	UHRR
GSE47774	GSM1157973	SEQC_ILM_NVS_A_4_L01_AGTCAA_AD0902ACXX	SEQC-A-BC13-s_1	UHRR
GSE47774	GSM1157975	SEQC_ILM_NVS_A_4_L02_AGTCAA_AD0902ACXX	SEQC-A-BC13-s_2	UHRR
GSE47774	GSM1157977	SEQC_ILM_NVS_A_4_L03_AGTCAA_AD0902ACXX	SEQC-A-BC13-s_3	UHRR
GSE47774	GSM1157979	SEQC_ILM_NVS_A_4_L04_AGTCAA_AD0902ACXX	SEQC-A-BC13-s_4	UHRR
GSE47774	GSM1157981	SEQC_ILM_NVS_A_4_L05_AGTCAA_AD0902ACXX	SEQC-A-BC13-s_5	UHRR
GSE47774	GSM1157983	SEQC_ILM_NVS_A_4_L06_AGTCAA_AD0902ACXX	SEQC-A-BC13-s_6	UHRR
GSE47774	GSM1157985	SEQC_ILM_NVS_A_4_L07_AGTCAA_AD0902ACXX	SEQC-A-BC13-s_7	UHRR
GSE47774	GSM1157987	SEQC_ILM_NVS_A_4_L08_AGTCAA_AD0902ACXX	SEQC-A-BC13-s_8	UHRR
GSE47774	GSM1157989	SEQC_ILM_NVS_B_1_L01_CGATGT_AD0902ACXX	SEQC-B-BC02-s_1	HBR
GSE47774	GSM1157991	SEQC_ILM_NVS_B_1_L02_CGATGT_AD0902ACXX	SEQC-B-BC02-s_2	HBR
GSE47774	GSM1157993	SEQC_ILM_NVS_B_1_L03_CGATGT_AD0902ACXX	SEQC-B-BC02-s_3	HBR
GSE47774	GSM1157995	SEQC_ILM_NVS_B_1_L04_CGATGT_AD0902ACXX	SEQC-B-BC02-s_4	HBR
GSE47774	GSM1157997	SEQC_ILM_NVS_B_1_L05_CGATGT_AD0902ACXX	SEQC-B-BC02-s_5	HBR
GSE47774	GSM1157999	SEQC_ILM_NVS_B_1_L06_CGATGT_AD0902ACXX	SEQC-B-BC02-s_6	HBR
GSE47774	GSM1158001	SEQC_ILM_NVS_B_1_L07_CGATGT_AD0902ACXX	SEQC-B-BC02-s_7	HBR
GSE47774	GSM1158003	SEQC_ILM_NVS_B_1_L08_CGATGT_AD0902ACXX	SEQC-B-BC02-s_8	HBR

GSE47774	GSM1158005	SEQC_ILM_NVS_B_2_L01_GCCAAT_AD0902ACXX	SEQC-B-BC06-s_1	HBR
GSE47774	GSM1158007	SEQC_ILM_NVS_B_2_L02_GCCAAT_AD0902ACXX	SEQC-B-BC06-s_2	HBR
GSE47774	GSM1158009	SEQC_ILM_NVS_B_2_L03_GCCAAT_AD0902ACXX	SEQC-B-BC06-s_3	HBR
GSE47774	GSM1158011	SEQC_ILM_NVS_B_2_L04_GCCAAT_AD0902ACXX	SEQC-B-BC06-s_4	HBR
GSE47774	GSM1158013	SEQC_ILM_NVS_B_2_L05_GCCAAT_AD0902ACXX	SEQC-B-BC06-s_5	HBR
GSE47774	GSM1158015	SEQC_ILM_NVS_B_2_L06_GCCAAT_AD0902ACXX	SEQC-B-BC06-s_6	HBR
GSE47774	GSM1158017	SEQC_ILM_NVS_B_2_L07_GCCAAT_AD0902ACXX	SEQC-B-BC06-s_7	HBR
GSE47774	GSM1158019	SEQC_ILM_NVS_B_2_L08_GCCAAT_AD0902ACXX	SEQC-B-BC06-s_8	HBR
GSE47774	GSM1158021	SEQC_ILM_NVS_B_3_L01_TAGCTT_AD0902ACXX	SEQC-B-BC10-s_1	HBR
GSE47774	GSM1158023	SEQC_ILM_NVS_B_3_L02_TAGCTT_AD0902ACXX	SEQC-B-BC10-s_2	HBR
GSE47774	GSM1158025	SEQC_ILM_NVS_B_3_L03_TAGCTT_AD0902ACXX	SEQC-B-BC10-s_3	HBR
GSE47774	GSM1158027	SEQC_ILM_NVS_B_3_L04_TAGCTT_AD0902ACXX	SEQC-B-BC10-s_4	HBR
GSE47774	GSM1158029	SEQC_ILM_NVS_B_3_L05_TAGCTT_AD0902ACXX	SEQC-B-BC10-s_5	HBR
GSE47774	GSM1158031	SEQC_ILM_NVS_B_3_L06_TAGCTT_AD0902ACXX	SEQC-B-BC10-s_6	HBR
GSE47774	GSM1158033	SEQC_ILM_NVS_B_3_L07_TAGCTT_AD0902ACXX	SEQC-B-BC10-s_7	HBR
GSE47774	GSM1158035	SEQC_ILM_NVS_B_3_L08_TAGCTT_AD0902ACXX	SEQC-B-BC10-s_8	HBR
GSE47774	GSM1158037	SEQC_ILM_NVS_B_4_L01_AGTTCC_AD0902ACXX	SEQC-B-BC14-s_1	HBR
GSE47774	GSM1158039	SEQC_ILM_NVS_B_4_L02_AGTTCC_AD0902ACXX	SEQC-B-BC14-s_2	HBR
GSE47774	GSM1158041	SEQC_ILM_NVS_B_4_L03_AGTTCC_AD0902ACXX	SEQC-B-BC14-s_3	HBR
GSE47774	GSM1158043	SEQC_ILM_NVS_B_4_L04_AGTTCC_AD0902ACXX	SEQC-B-BC14-s_4	HBR
GSE47774	GSM1158045	SEQC_ILM_NVS_B_4_L05_AGTTCC_AD0902ACXX	SEQC-B-BC14-s_5	HBR
GSE47774	GSM1158047	SEQC_ILM_NVS_B_4_L06_AGTTCC_AD0902ACXX	SEQC-B-BC14-s_6	HBR
GSE47774	GSM1158049	SEQC_ILM_NVS_B_4_L07_AGTTCC_AD0902ACXX	SEQC-B-BC14-s_7	HBR
GSE47774	GSM1158051	SEQC_ILM_NVS_B_4_L08_AGTTCC_AD0902ACXX	SEQC-B-BC14-s_8	HBR

1. Chandramohan, R., Po-Yen, W., Phan, J.H. and Wang, M.D. (2013), *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pp. 647-650.