Supplementary Text S1

1 EVIDENCE OF BOTH INTER- AND INTRA- GENE BIAS

We observed a large bias parameter $\hat{\lambda}$ difference among exons even within the single isoform genes, which proved the necessity of deliberating intra-gene bias for expression quantification (Section 3.1 of the main manuscript). In order to check how extreme the observed intra-gene bias is, the null distribution of the $\hat{\lambda}$ differences under the assumption that the bias factor is a constant within genes was obtained through the following simulation. Based on the HBM liver RNA-seq data, the median $\hat{\lambda}$ across all the exons was 0.660 (inferred from single-isoform genes). The median number of exons per single-isoform gene was 9. Accordingly, in our simulation we considered a single-isoform gene containing 9 exons with length of 250. The average number of reads starting from a position E[X] was set to 2 and $\hat{\lambda}$ was set to 0.660. We then simulated read counts for each exon according to GP ($E[X]\times\left(1-\hat{\lambda}\right), \hat{\lambda}$). We repeated the simulation 1000 times, and the maximum $\hat{\lambda}$ difference among exons ($\Delta\hat{\lambda}_{\max}$) was recorded each time. The histogram of $\Delta\hat{\lambda}_{\max}$ is plotted in **Figure. S2**. The probability of observing $\Delta\hat{\lambda}_{\max} > 0.2$ is less than 0.001 based on this empirical distribution, providing strong evidence for the existence of intra-gene bias difference since we observed a median of 0.20-0.32 $\Delta\hat{\lambda}_{\max}$ in the real RNA-seq data.

Similarly, we estimated the gene-level bias over the same simulated reads under the assumption that $\hat{\lambda}$ is the same across different genes. For each round, reads for 1000 genes were generated and the maximum gene-level $\hat{\lambda}$ difference was calculated. We repeated this procedure for 100 times. The largest $\Delta\hat{\lambda}_{\max}$ was 0.120 for the 100 simulations. Therefore the range of 0.21-0.95 as mentioned in Section 3.1 of the main manuscript strongly supports the existence of inter-gene bias heterogeneity.

2 POSITIONAL AND SEQUENCE-SPECIFIC BIAS CORRECTION IN SINGLE-ISOFORM GENES

For the comparison purpose, we considered the positional and sequence-specific bias removal. For the positional bias removal, we divided positions of each single-isoform gene into 20 bins from 5' to 3' of the gene. The average number of read counts $\bar{x}_k$ for the k^{th} bin across all the single-isoform genes was calculated. Then the positional bias $w_{p,k}$ for positions in the k^{th} bin was

$$w_{p,k} = \frac{\frac{1}{20}\sum_{k=1}^{20} \bar{x}_k}{\bar{x}_k}. \quad (1)$$

For the sequence-specific bias removal, we treated the preceding 6 nt of the first read end and the following 6 nt of the second read end as the primers. Let $c(p_i), i = 1, \cdots, n_p$ denote the occurrence of primer p_i in the reads mapped. Then the sequence based bias for read i was

$$w_{s,i} = \frac{\frac{1}{n_p}\sum_{i'=1}^{n_p} c(p_{i'})}{c(p_i)}. \quad (2)$$

Hence, read count x_i was adjusted according to the weights: $x_i' = x_i \times w_{p,bin(i)} \times w_{S,i}$ where $bin(i)$ denotes the relative position bin that the position i belongs to. For our model, we estimated the over-dispersion parameter for the virtual exon s as λ_s from the GP model. The corrected read count was $x_i'' = x_i \times (1 - \lambda_{s_i})$, where S_i denotes the virtual exon that the position i belongs to. For reads that span multiple virtual exons, the first virtual exon was used for the bias assignment.

Besides the RNA-seq data considered in **Figure 1**, we also applied our method on three tier-1 cell lines in the ENCODE data (K562, GM12878, and H1 embryonic stem cells). WemIQ still demonstrated improved read uniformity in the K562 and GM12878 cell lines. In the H1 embryonic stem cells, its result was comparable with that without the bias correction. The GP bias removal in WemIQ was still better than the positional and sequence-specific bias removal (**Fig. S4**).

3 SIMULATION OF RNA-SEQ DATA BASED ON BASIC GENE STRUCTURES

A variety of gene structures were used to compare the gene and isoform level expression quantification. We first simulated a two-isoform gene model, where a longer transcript with five 250-nt exons was generated and the second exon was skipped as a cassette exon to form the shorter isoform. We then considered a three-isoform gene model, where two shorter isoforms skipping cassette exon 2 (“cas2”) or 4 (“cas4”) were generated. To show the robustness of model, we also generated gene structures complicated by both alternative splice sites (“ASS”) and exon skipping events, or gene structures with transcripts missing from the annotation. These gene structures were shown in **Figure S5**.

Then two steps were used to simulate the pair-end RNA-seq reads: (1) to generate the read counts starting at each position, (2) to assign fragment length for each read pair. When generating the read counts, generative models, either with uniform read sampling assumption (Li and Dewey, 2011) or with only a few known bias factors (Griebel, et al., 2012), could be problematic because the true bias might arise from every step of the RNA-seq experiment and could tangle with each other in a dynamic way. Instead, we directly estimated the bias level in the real data, and simulated the reads with similar over-dispersion properties through the negative binomial distribution to better mimic the real situation. Although GP distribution has been proved to significantly improve the fitting of RNA-seq reads (Srivastava and Chen, 2010), we used the negative binomial distribution $NB(r, p)$ to simulate the reads to avoid the possible advantage to our model in the following analyses. The negative binomial distribution $NB(r, p)$ can be treated as a gamma-Poisson mixture: a $Poisson(\theta)$ distribution whose θ itself is a random variable and distributed as $Gammar\left(r, p/(1-p)\right)$. The mean of $NB(r, p)$ is $rp/(1-p)$ and the variance is $rp/(1-p)^2$. Therefore, we can vary p to simulate different degrees of overdispersion between exons. For exons shared between isoforms, we kept p_i constant and varied the ratio of their isoform-specific expression by r.

For the two isoform gene models, p_i was sampled uniformly from 0.75 to 0.95 for exon i and kept constant for exons shared between the two isoforms. Then we set $r = 1$ for the minor isoform, and sampled a transcript expression ratio f with equal probabilities from $\{2,3,4\}$ with replacement. Therefore, in the major isoform, reads for exon i follows the distribution $NB(f, p_i)$. To simulate the low-coverage cases, the r value was set to be 0.05. For the high coverage scenarios ($r = 1$ for minor isoforms), the average position-level read count was 19.35 to 22.28, equivalent to the 99.3 percentile in the ENCODE RNA-seq data (see details in the section of "real datasets"). For the relatively low coverage scenarios ($r = 0.05$ for minor isoforms), the average position-level read count was 0.96 to 1.11. For the three-isoform models, we first selected one isoform as the minor one and set its r equal to one, and then randomly selected two expression ratios f_1 and f_2 for the remaining two isoforms. The relative errors of isoform percentage estimates of the same gene were summed up to evaluate the overall performance. A total of 1,000 cases were simulated in the two-isoform gene model with both the high and the low coverage, as well as the model with incomplete annotation. For other three-isoform gene models, 500 cases were simulated.

After the read counts were generated for each exonic position, the fragment length was assigned for each 50-bp paired-end read using the Gaussian distribution with the mean of 236.49 and the standard deviation of 23.55. These parameters were inferred from the single-isoform genes in the human tissue RNA-seq datasets (Illumina BodyMap2 transcriptome, ERP000546). For a read pair, if the ending position was beyond the corresponding transcript, this fragment length was defined as invalid. Up to 5 trials were performed to find a valid fragment length. Otherwise, this read was discarded.

4 SIMULATION OF RNA-SEQ DATA BASED ON REAL ENSEMBL GENE STRUCTURES

To further demonstrate the performance of WemIQ, we selected a variety of Ensembl gene structures to simulate RNA-seq reads. We set up three criteria for the real gene structure selection: 1) this gene contains at least one protein coding transcript; 2) it contains at least one cassette exon with at least 50 bp in length to avoid extremely challenging gene structures (e.g., all the transcripts are non-identifiable, or they only contain several nucleotide differences); 3) the gene is on chromosome 1. The cassette exon list was downloaded from the Ensembl website (http://www.ensembl.org/index.html). As a result, 78 genes with 2~10 transcripts on chromosome 1 were selected based on the Ensembl Annotation (http://www.ensembl.org/index.html).

I. Assignment of over-dispersion parameters

Similarly we adopted the negative binomial distribution $NB(r, p)$ to simulate the reads. The reason that we chose p_i uniformly from 0.75 to 0.95 for exon i is the following. The expectation and variance of a NB distributed variable X can be written as

$$\begin{aligned} E[X] &= \frac{rp}{1-p} \\ \mathrm{var}[X] &= \frac{rp}{(1-p)^2} \end{aligned} \tag{3}$$

Due to the imperfectness of the experiment, variance larger than expectation is often observed. We could represent the over-dispersion from equation (3) by

$$over-dispersion = \frac{\mathrm{var}[x]}{E[x]} = \frac{1}{1-p} \tag{4}$$

The mean and variance of a GP distributed variable X can be represented as

$$E[X]=\frac{\theta}{1-\lambda}$$
$$\mathrm{var}[X]=\frac{\theta}{(1-\lambda)^3} \quad (5)$$

As a result the over-dispersion can be approximated as

$$over-dispersion=\frac{\mathrm{var}[X]}{E[X]}=\frac{1}{(1-\lambda)^2} \quad (6)$$

To match the two over dispersion rates, we require

$$\frac{1}{(1-\lambda)^2}=\frac{1}{1-p} \quad (7)$$

so that

$$p \doteq \left(1-\hat{\lambda}\right)^2 \quad (8)$$

We therefore utilized the estimated overdispersion rate from the real data to set the range of the parameter p. To exclude very extreme cases, we excluded the upper and bottom 10 percent quantiles from the estimated $\hat{\lambda}$ in the real data. As a result, $unif[0.75,0.95]$ was selected for p.

II. Generate the relative expression for each transcript

For the each transcript for a gene, the relative expression was generated from a uniform distribution. Specifically, we first sampled f_1 from $unif(0,1)$ for transcript 1, and then $f_2 \sim unif[0,1-f_1]$ and $f_3 \sim unif[0,1-f_1-f_2]$ for transcripts 2 and 3, and so on. To exclude the situation that we might be biased to give larger relative expression values to transcript 1, we then shuffled $\{f_1,f_2\cdots f_n\}$ to reassign them randomly to different isoforms.

It is reported that at a certain time point for a specific tissue, usually there is only one dominant transcript for the multiple-transcripts genes (Gonzalez-Porta, et al., 2013). Our simulation above incorporated such cases by giving very large values to a certain transcript. At the same time, we considered the effect of balanced expressions from similar transcripts to the final expression analysis.

The total r value was sampled uniformly from 0.5 to 2.0. Correspondingly, the average number of reads per position was from 4 to 60 if there were four transcripts, representing a variety of highly and moderately expressed genes.

III. Generate the number of reads at each position

For each position k of the transcript j, the number of fragments $m_{k,j}$ is generated by $NB\left(rf_i, p_{n_k}\right)$, where n_k indicates which the virtual exon this position belongs to.

IV. Generate the fragment length for each fragment

The fragment length was assigned randomly following a Gaussian distribution with the mean of 236.49 and the standard deviation of 23.55. Similarly, for a read pair, if the ending position was beyond the corresponding transcript, this fragment length was defined as invalid. Up to 5 trials were performed to find a valid fragment length. Otherwise, this read was discarded. We generated 75bp pair-end reads for these set of simulations.

Due to the fast speed of STAR (Dobin, et al., 2013), we utilized it instead of TopHat for mapping during the manuscript revision process. First, Ensembl V75 annotation was given to STAR to build the genome, and then the simulated reads were mapped to the genome with up to 2 mismatches. The resultant sam files were givens to WemIQ for expression quantification. Cufflinks requires that the input sam or bam files should be sorted, so we used the samtools to sort the mapped files before estimating expressions. Again, RSEM would call bowtie internally for its own quantification analysis. Additionally, as suggested during the manuscript revision process, we also added MMSEQ (Turro, et al., 2011) for the performance comparison. It calls Bowtie internally before expression quantification. Let f_i represent the relative expression of transcript j among the m isoforms of a gene, then the estimation error for isoforms of the same gene was calculated as

$$\left|f_j - \hat{f}_j\right| \qquad (9)$$

It is worth mentioning that in simulations based on basic gene structures (i.e. the synthetic ones), since the minor isoform percentage was at least 1/5, we utilized the relative error ($\sum_{j=1}^{m}\left|f_j - \hat{f}_j\right|/f_j$) for performance evaluation as mentioned in Section 2.2 of the main manuscript. However, in the real gene structure, due to the simulation scheme, the smallest isoform percentage could be very small and thus the relative error might be dominated by these isoforms. Hence we adopted equation (9) for simulations based on real gene structures.

For the above simulations, the boxplots of the errors are given in **Figure S6**. It is shown that WemIQ demonstrates lower estimation errors than those for Cufflinks, RSEM, and MMSEQ (Ps < 0.0007, Wilcoxon tests).

For a better virtualization, the scatter plots of the estimated isoform percentage vs. the true percentage was given in **Figure 5**, where the red dots represent the isoforms with errors larger than 0.1. As it is shown, WemIQ apparently provides smaller number of such red dots. The R-squared values using linear regressions from WemIQ was as high as 0.9273, larger than 0.8828 for MMSEQ, 0.879 for cufflinks, and 0.8748 for RSEM.

In order to further test the robustness of WemIQ, we selected all the multi-isoform autosomal genes from the Ensembl annotation (version 75). Similar strategy was performed to simulate the reads except that r was uniformly selected from 0.25 to 0.5 to control the whole coverage of the full dataset. For the analysis of MMSEQ, we included the isoform percentage estimation from its Gibbs sampling results (referred as MMSEQ_GS) in addition to the direct ratio between isoform expression and gene expression which were also reported from MMSEQ (referred as MMSEQ_ratio). Note that the direct ratio was used in other places with "MMSEQ" labels. The isoform-level estimation errors were summarized in **Figure S7**. WemIQ still provides more accurate estimations. Here, we turned on (w/) and turned off (w/o) the bias correction of Cufflinks and RSEM for the comparison.

5 MORE DETAILS OF REAL DATASETS AND READ MAPPING

For the bias removal comparison, Ensembl and Refseq gene annotations (version GRCh37/hg19) were downloaded from the UCSC genome browser http://genome.ucsc.edu/. To show the non-uniformity of read sampling, we analyzed the single-isoform genes with 50-bp paired-end reads from the human brain, liver, kidney, and muscle data sets (Illumina BodyMap2 transcriptome, ERP000546, **Fig. 1**). These genes have only one isoform according to both the Ensembl and Refseq annotations. Additionally we applied our method on three tier-1 cell lines in the ENCODE data (K562, GM12878, and H1 embryonic stem cells) as shown in **Figure S4**.

TaqMan qRT-PCR data in a human brain sample from the MAQC project (Shi, et al., 2006) was used for the evaluation of gene expression estimation. Specifically, we applied WemIQ, Cufflinks, RSEM on a set of 50-bp pair-end reads from the same human brain sample (SRP002274 (Robinson and Smyth, 2007)). We required that for each gene, at least 75% of the qRT-PCR replicates had a detectable expression and WemIQ, Cufflinks, and RSEM all provided a reliable estimate. Finally, 526 genes were compared between the qRT-PCR estimates and the RNA-seq estimates from each method. The scatter plots of the estimates were given in **Figure S8.** We also turned off

the bias correction option for Cufflinks and RSEM. Without the bias correction of Cufflinks, the correlation value was 0.580 which was lower than the Cufflinks result with the correction (0.681). For RSEM, the correlation without the bias correction was 0.697, very similar to the result with the bias correction (0.700). Both of them were lower than WemIQ (0.734).

We downloaded four RNA-seq datasets on the human GM12878 cells from the ENCODE project to compare the consistency of estimates across technical replicates and laboratories (GEO Accession code: GSM958728 and GSM758559) (2004). The two datasets (technical replicates) from the California Institute of Technology laboratory ("Caltech") were 75-bp pair-end reads and the other two (technical replicates) from the Cold Spring Harbour laboratory ("Cshl") were 76-bp pair-end reads preserving the strand information. The gene and isoform expressions were estimated according to the Refseq annotation by WemIQ, Cufflinks, or RSEM. Before the comparison among multiple methods, we found that several highly expressed genes took up to 40% of the total expression in all four datasets based on Cufflinks, but were barely expressed according to RSEM and WemIQ estimation. After careful deliberation, these genes were short microRNAs or small nuclear RNAs that are usually less than 100bp in length. And only one end of the pair-end reads were mapped to several beginning positions of these transcripts. But in WemIQ only concordant pair-end reads were considered, so these genes were recognized as lowly expressed. RSEM also assigned very low expression levels to these genes. Therefore, we removed transcripts less than 150-nt from the following studies. In the analyses, we assembled a group of highly expressed genes as a high-confident set to test the methods. They were the top 4 percent genes (628 genes) according to the gene length normalized read counts summed from all the four data sets. On the isoform level, we had to rely on the isoform expression estimates to define highly expressed isoforms, but different methods provided different isoform estimates. Hence, we selected the top 10 percent expressed isoforms from each method and the union of these isoforms (1,938 transcripts) was defined as highly expressed isoforms. For the comparison of two datasets, the fold change was calculated as the ratio of the larger value to the smaller value. For a better virtualization, the scatter plots of cross-lab estimates are given in **Figures S9** and **S10**. Cufflinks without the bias correction provided slightly worse results than the version with the correction. The median expression fold change across difference labs was 1.52 VS. 1.50 and 1.71 VS. 1.68 for the gene and isoform expression levels. RSEM also demonstrated similar or slightly worse performance without the bias correction

compared with that with the bias correction (1.98 VS. 1.80 at the gene level and 1.66 VS.1.66 at the isoform level). WemIQ still outperformed both methods no matter whether they turned on the bias correction or not.

Twenty-one mouse RNA-seq samples were downloaded from SRA (SRP015959 (Shalek, et al., 2013)), with libraries constructed from 18 single cells, and 3 populations of 10,000 cells. Due to the high contamination of the adaptor sequences (Shalek, et al., 2013), we first trimmed the reads into 50 bp before mapping them to the transcriptome. NCBI annotation build37.1 for mice was used and genes that are less than 150bp in length were removed. For each gene, we summed the read counts from the 18 single-cell RNA-seq datasets and selected the top 4 percent of genes as the highly expressed gene group (991 genes). Isoform expression estimates were averaged across the 18 single-cell data for each of the three methods, and the union of the top 4 percent average isoforms from each method was defined as the highly expressed isoforms (907 isoforms). Cufflinks with or without the bias correction performed similarly. At the gene level, the median coefficient of variation across different cells for Cufflinks with and without the bias correction was 0.87 and 0.88 respectively. For the highly expressed isoforms, the coefficient of variation did not show significant differences (0.84 VS. 0.82 for with and without the bias correction). Besides, RSEM also provided similar results with or without the bias correction. For example, the CV values for the highly expressed genes and isoforms were 0.89 and 0.84 respectively without the bias correction (compared to 0.89 and 0.85 for RSEM with the bias correction).

In terms of read mapping, both WemIQ and Cufflinks rely on additional software for read mapping before quantification, while RSEM and MMSEQ called Bowtie internally. In the main manuscript, Sections 3.1-3.4, Sections 3.6-3.8 all used TopHat for the pair-end reads mapping. During the manuscript revision process, STAR was used as the alternative mapper due to its faster speed for all the remaining analysis (including Section 3.5 of the main manuscript, Sections 1, 2 (only for the ENCODE data sets), 4-7 of **Text S1**).

6 ACCURACY OF WEMIQ ON GENE-LEVEL ESTIMATION WITH HEAVILY OVERLAPPED TRANSCRIPTS

To test the accuracy of WemIQ gene-level estimation under heavily overlapped transcripts, we selected the gene ENSG00000162598 from the Ensembl annotation to simulate the reads. The gene structure is shown in **Figure S11**. In total there are 6

transcript isoforms in ENSG00000162598, of which ENST00000488027 and ENST00000395552 overlap significantly. Then we repeated 50 simulations to evaluate the gene-level and isoform-level estimation error. For each repeat, we have

$$e_{i,j,iso} = | \hat{\theta}_{i,j} - \theta_{i,j,true} |$$
$$e_{i,gene} = \left| \frac{\hat{g}_i}{\sum_{i'} \hat{g}_{i'}} - \frac{g_{i,true}}{\sum_{i'} g_{i',true}} \right| \quad (10)$$

where, $\hat{\theta}_{i,j}$ represents the percentage of isoform j in gene i, and $\hat{g}_i$ represents the expression of gene i. Then the boxplot of gene and isoform level errors are given in **Figure S11**. The isoform-level errors were generally high but the gene-level errors were still low. WemIQ provided more accurate gene-level expression estimates than other methods (the Wilcoxon test P values = 8.8×10^{-09}, 3.1×10^{-12}, and 8.1×10^{-06} as compared with RSEM, Cufflinks, and MMSEQ). In addition, WemIQ also provided more accurate isoform-level estimation for this highly overlapped gene structure (P < 0.00015 in the Wilcoxon tests with the other three methods).

Based on the simulations, we did observe the anti-correlated errors for genes and isoforms ($e_{i,gene}$ vs. $e_{i,iso} = \sum_j e_{i,j,iso}$) in MMSEQ estimates (–0.17). However, RSEM, Cufflinks, and WemIQ, the correlation between gene and isoform estimation errors was 0.13, 0.20, and –0.04 respectively. None of these correlations were significant by the correlation test in R.

7 ROBUSTNESS OF WEMIQ ESTIMATION BASED ON READ RESAMPLING

A good expression quantification method requires highly consistent estimation over replicates. To test the robustness, we compared the isoform quantification results across different sub-sampled reads from the same RNA-seq experiment. Specifically, we selected all the genes on chromosome 1 according to the Gencode annotation version 19 (http://www.gencodegenes.org/). RNA-seq data were downloaded from ENCODE on the GM12878 cell line (cytosol long RNA, polyA+). For all the reads that were uniquely mapped to chromsome1, we sub-sampled 75% of the reads and repeated it for 10 times. By requiring non-zero expression in at least one of the sampled datasets, 9281 isoforms were included in the following analysis. The boxplots of the maximum isoform percentage difference across the subsamples are plotted in **Figure S12** (outliers not shown for visual convenience). The median of the maximum isoform

percentage difference for these 9281 isoforms was only 0.007 by WemIQ, less 0.01 from Cufflinks (the Wilcoxon test $P < 2.2\times10^{-16}$).

In order to evaluate the stability of WemIQ under different read coverage, we selected one gene ENSG00000127184 from the Ensembl annotation (ENCODE GM12878 cell line), which has 7 isoforms and some of the isoforms heavily overlap with each other, leaving the isoform quantification extremely challenging. Ten sub-sampling rates (10%, 20%, …, 90%) were used and for each case 100 samplings were repeated. The boxplots of isoform percentage of ENST00000247655 (the major isoform) are given in **Figure S13**. It is shown that the variance of isoform percentage estimates is coverage dependent and higher coverage would lead to more robust estimates. However, even when the sub-sampling rate was only 10%, the interquartile range of the estimates was still less than 0.05, which again suggests very robust isoform-level quantification for our WemIQ.

8 REFERENCES

(2004) The ENCODE (ENCyclopedia Of DNA Elements) Project, *Science*, **306**, 636-640.

Dobin, A.*, et al.* (2013) STAR: ultrafast universal RNA-seq aligner, *Bioinformatics*, **29**, 15-21.

Gonzalez-Porta, M.*, et al.* (2013) Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene, *Genome biology*, **14**, R70.

Griebel, T.*, et al.* (2012) Modelling and simulating generic RNA-Seq experiments with the flux simulator, *Nucleic acids research*, **40**, 10073-10083.

Hansen, K.D., Brenner, S.E. and Dudoit, S. (2010) Biases in Illumina transcriptome sequencing caused by random hexamer priming, *Nucleic acids research*, **38**, e131.

Li, B. and Dewey, C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, *BMC Bioinformatics*, **12**, 323.

Robinson, M.D. and Smyth, G.K. (2007) Moderated statistical tests for assessing differences in tag abundance, *Bioinformatics*, **23**, 2881-2887.

Shalek, A.K.*, et al.* (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells, *Nature*, **498**, 236-240.

Shi, L.*, et al.* (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements, *Nat Biotechnol*, **24**, 1151-1161.

Srivastava, S. and Chen, L. (2010) A two-parameter generalized Poisson model to improve the analysis of RNA-seq data, *Nucleic acids research*, **38**, e170.

Turro, E.*, et al.* (2011) Haplotype and isoform specific expression estimation using multi-mapping RNA-seq reads, *Genome biology*, **12**, R13.

Supplementary Figures

Figure S1. Flowchart of WemIQ.

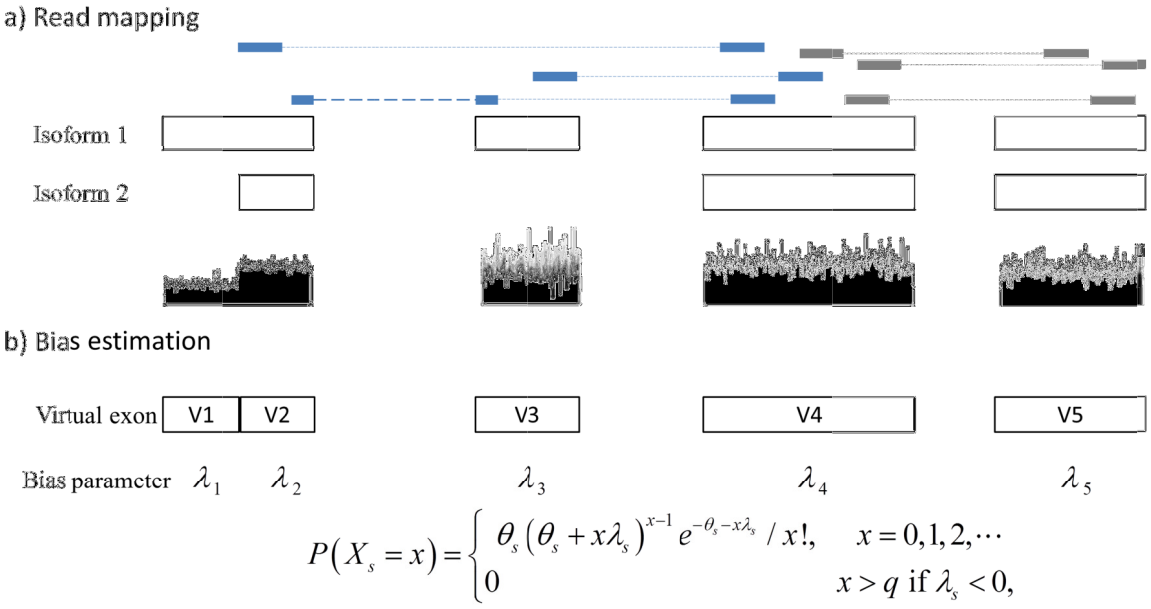

Figure S2. Histogram of $\Delta\hat{\lambda}_{\max}$ within a gene under the null assumption of the same bias within a gene.

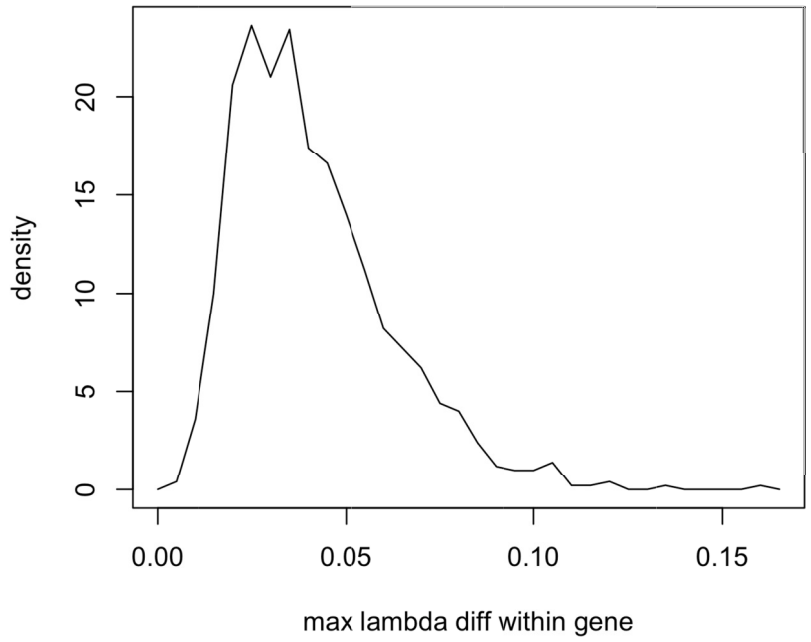

Figure S3. Boxplots of the KS statistics by different methods in four HBM RNA-seq data sets.

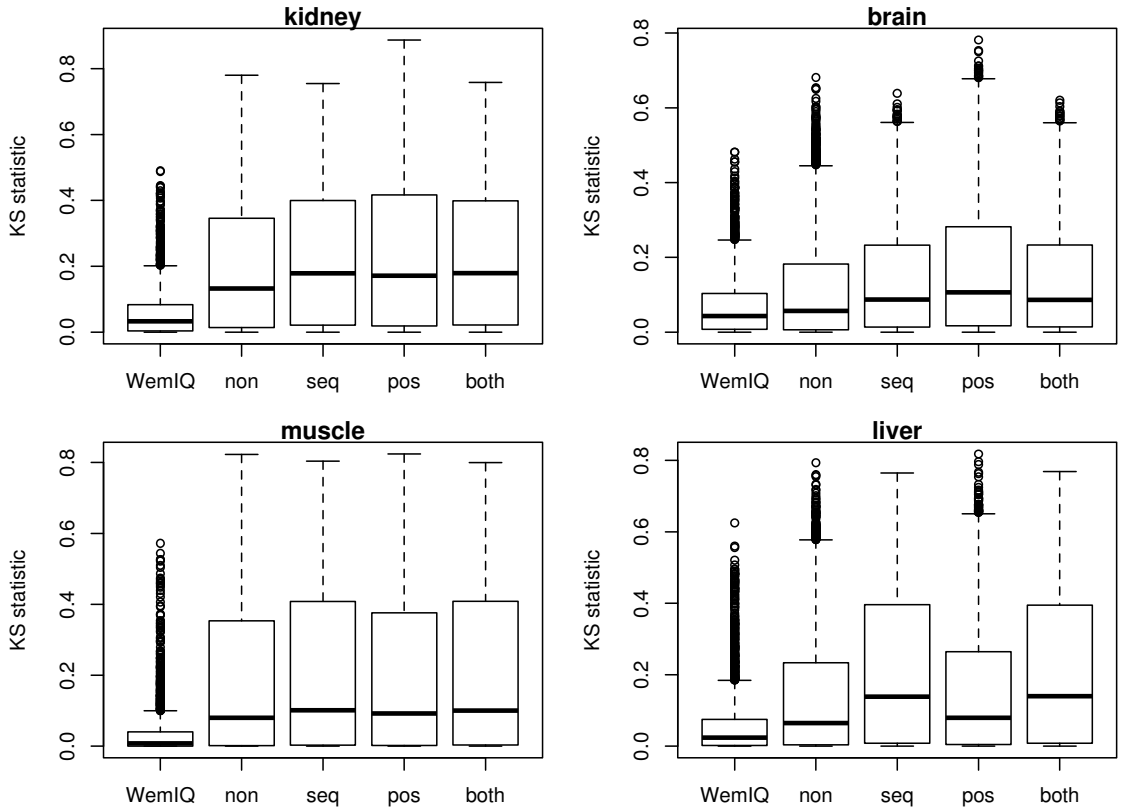

Figure S4. Bias correction on 3 tier-1 ENCODE RNA-seq data sets.

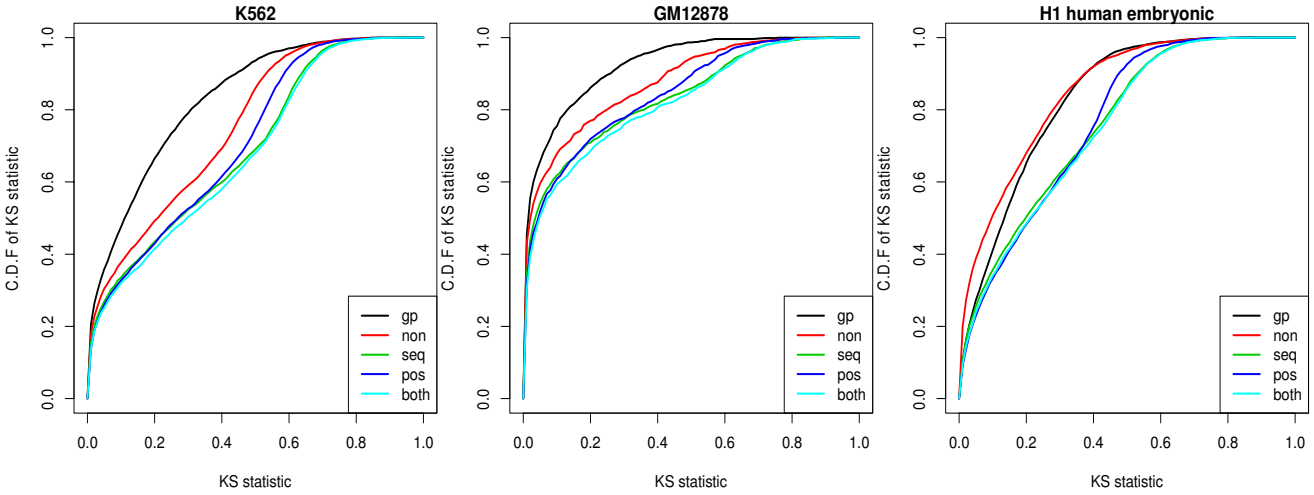

Figure S5. Basic gene structures used for simulations.

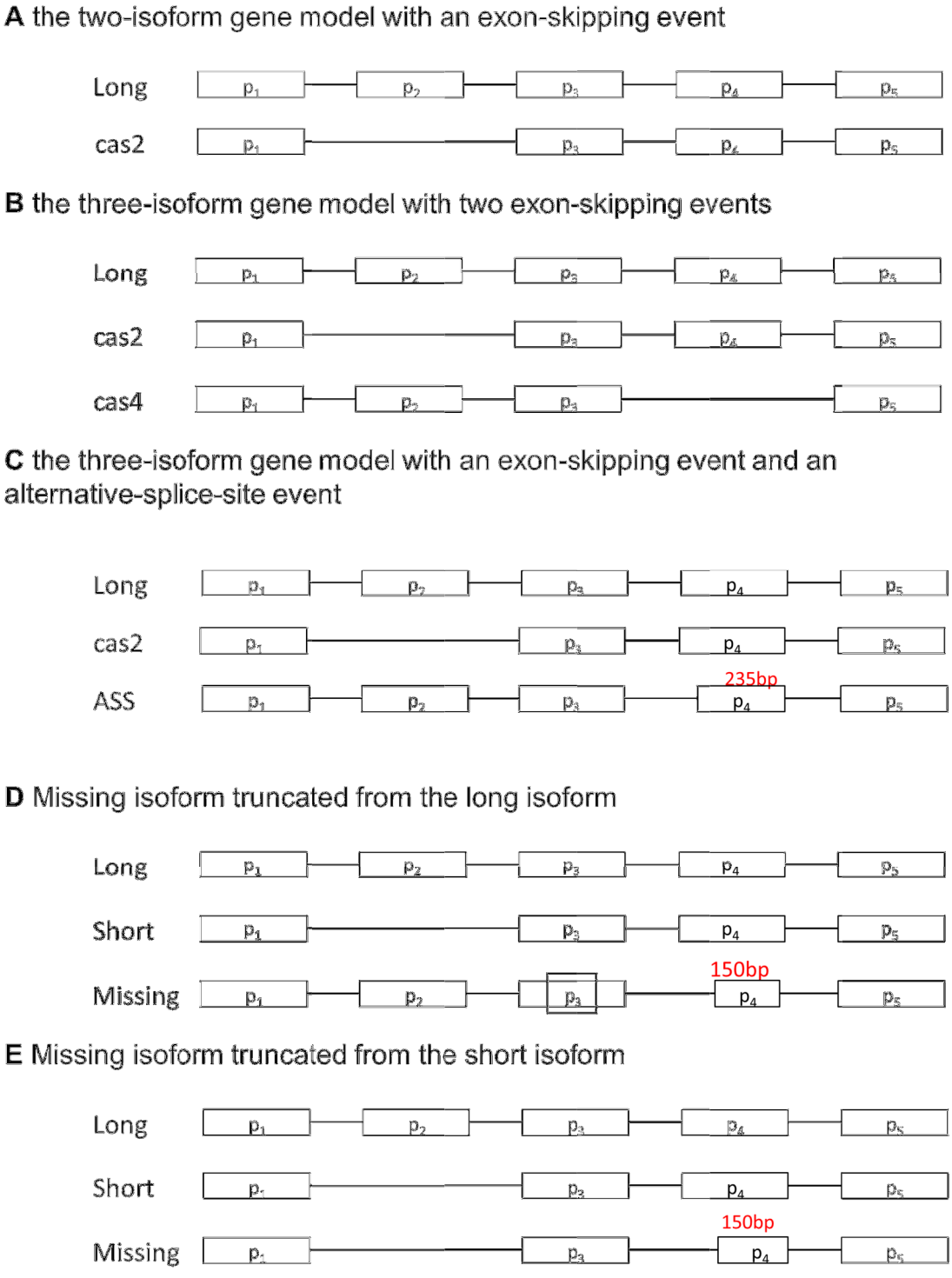

Figure S6. Isoform percentage errors of four methods under 78 real gene structures.

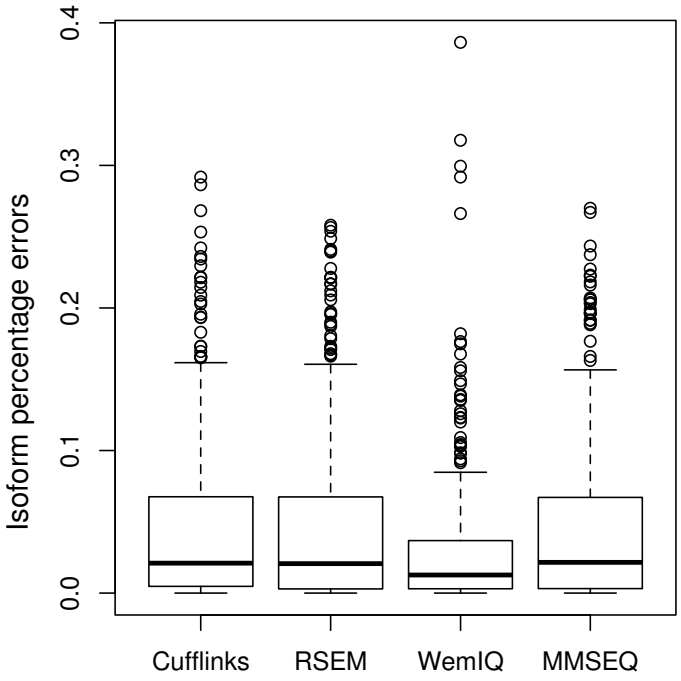

Figure S7. Isoform percentage errors of four methods under the genome-wide real gene structures.

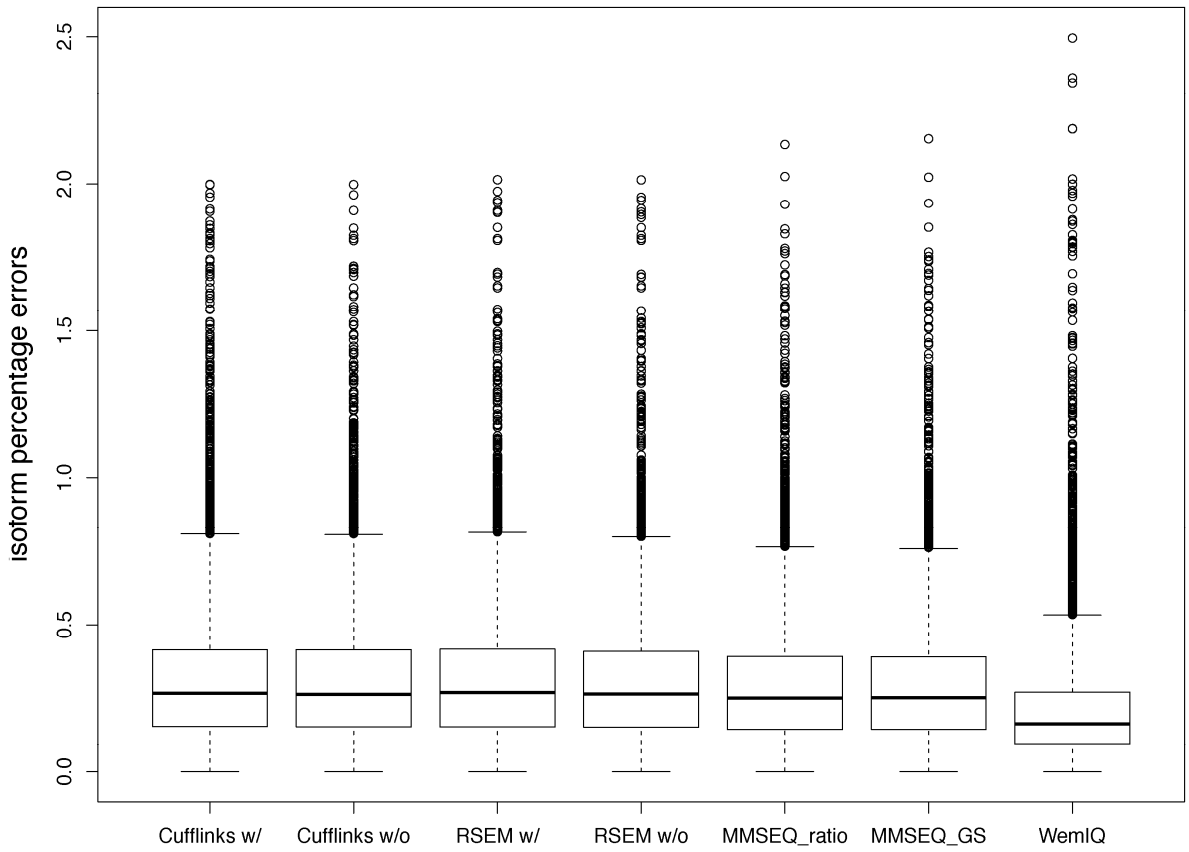

Figure S8. Scatter plots of RNA-seq expression estimates VS. qRT-PCR values for the MAQC data.

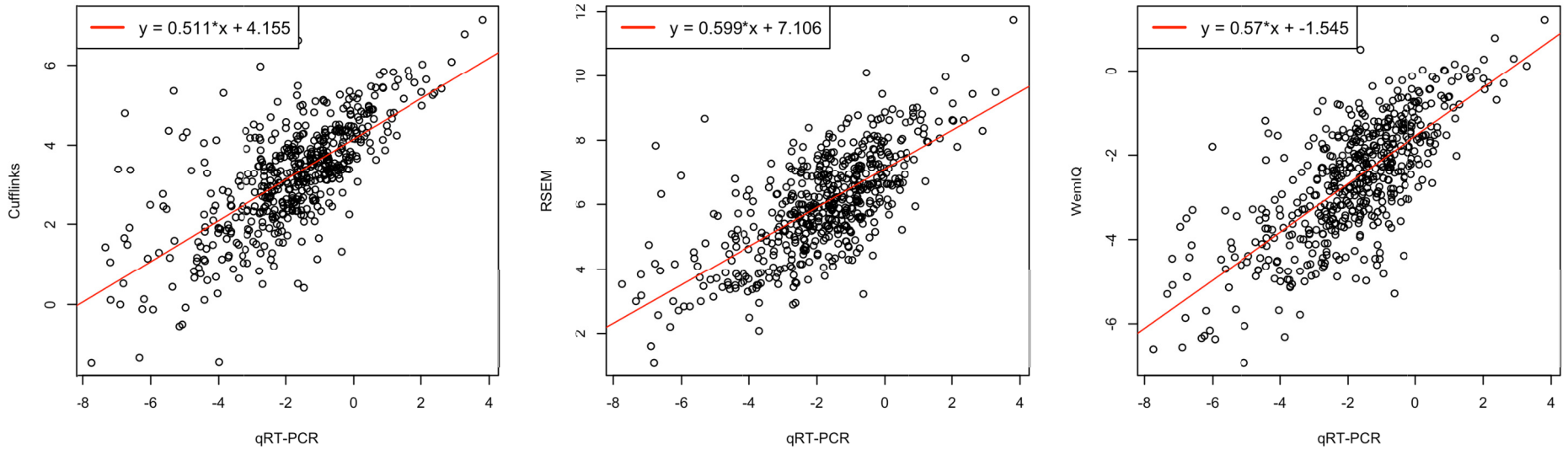

Figure S9. Scatter plots of the gene-level quantification across different labs. Red dots are those with gene-level quantification with larger than 2 fold change.

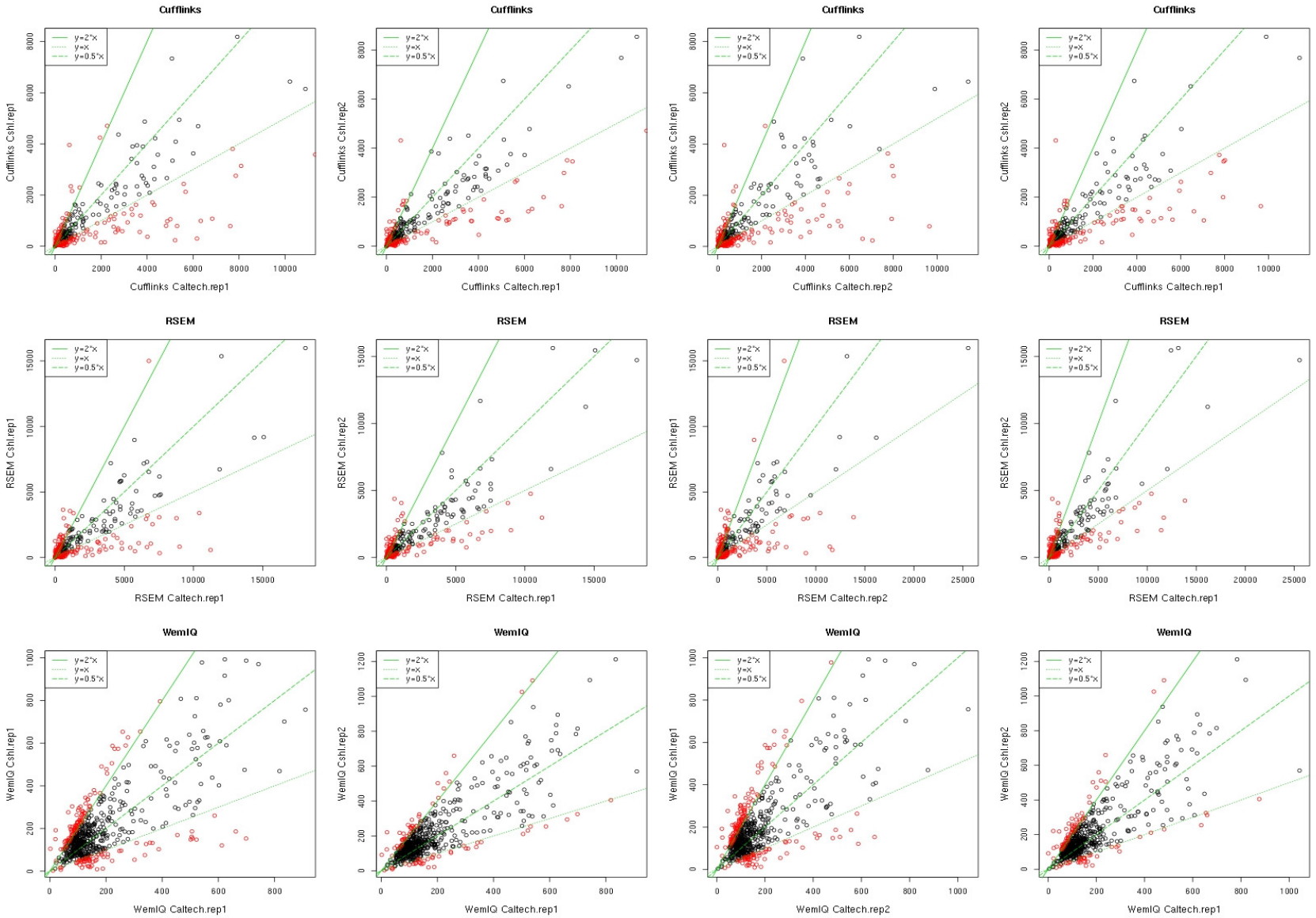

Figure S10. Scatter plots of the isoform-level quantification across different labs. Red dots are those with isoform-level quantification with larger than 2 fold change.

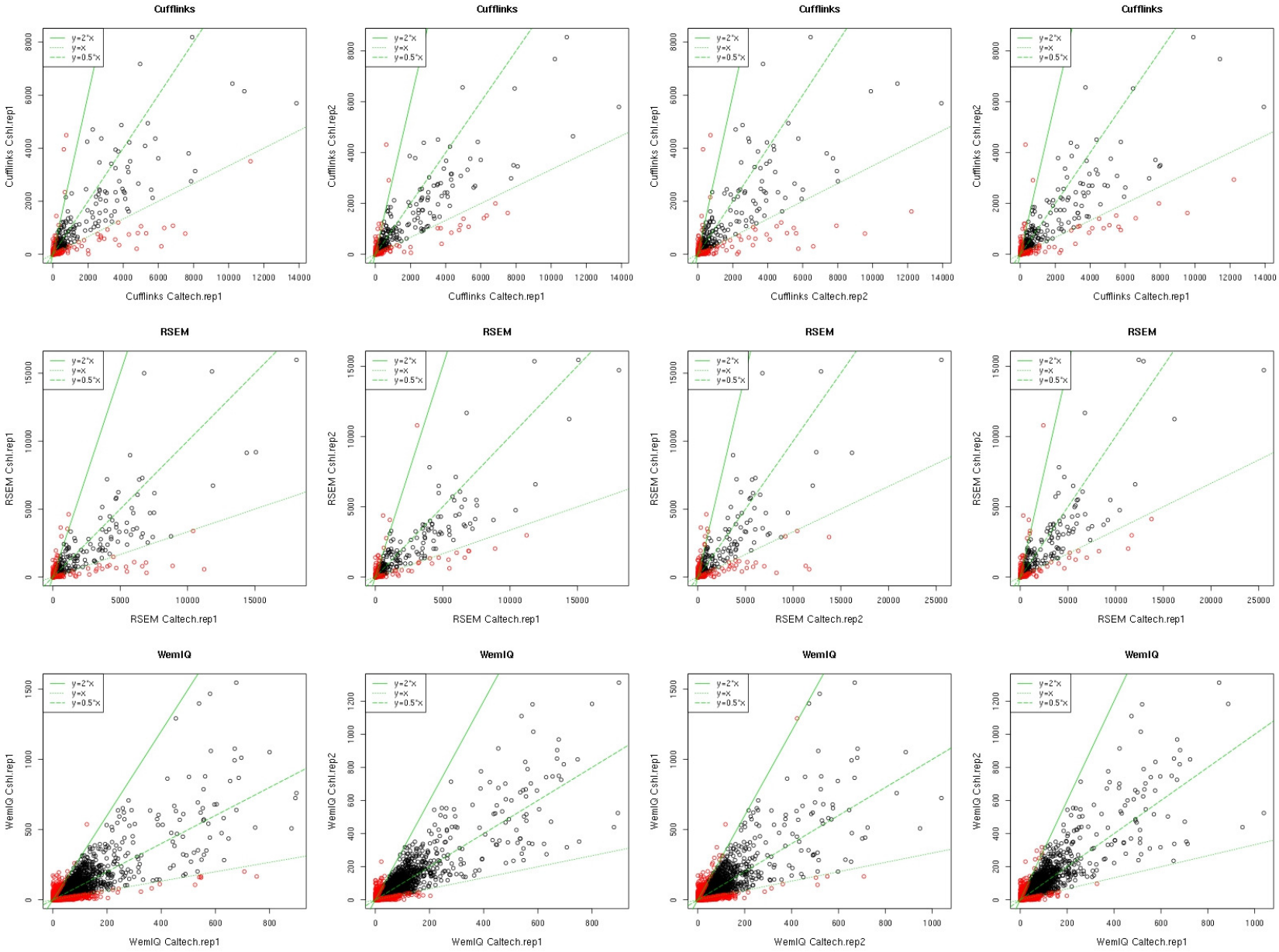

Figure S11. Gene and isoform level estimation errors of simulated reads under the structure of ENSG00000162598.

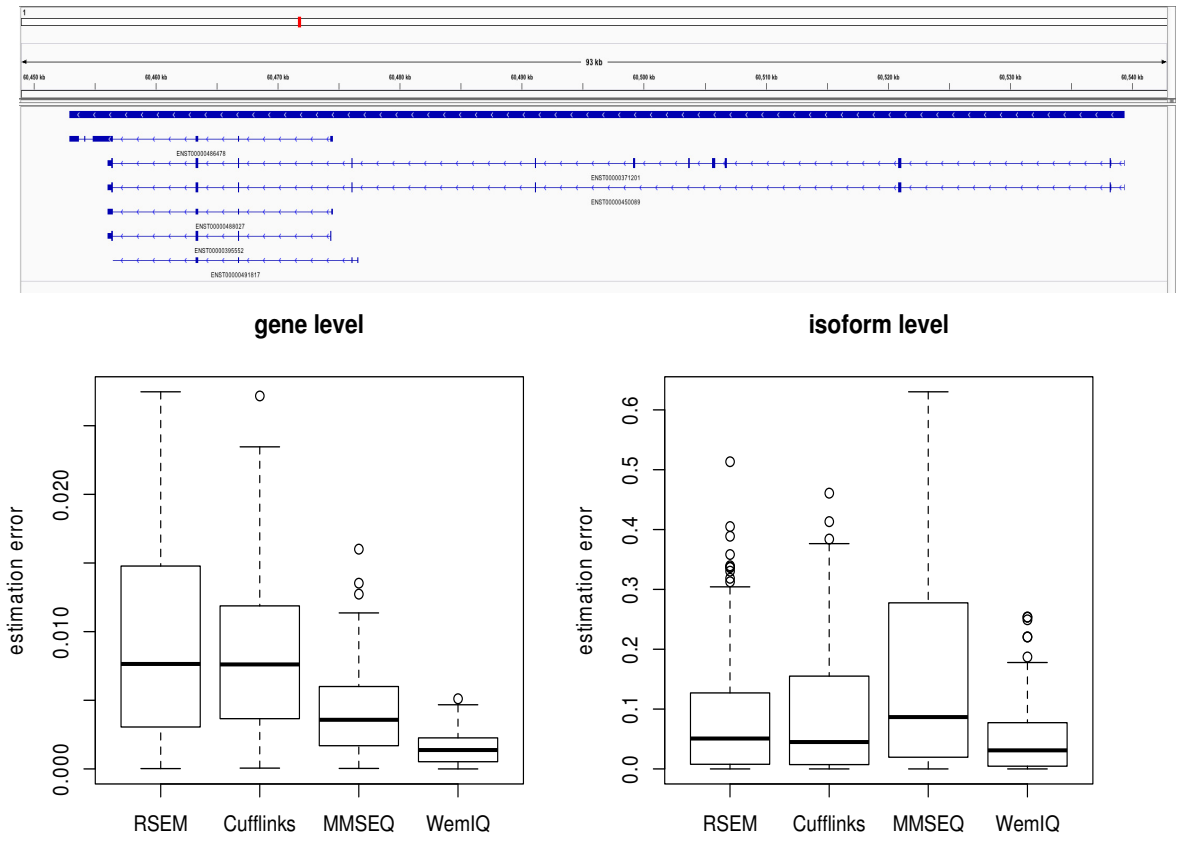

Figure S12. Boxplots of the maximum isoform percentage estimation difference across 10 sub-sampled reads from a high coverage ENCODE RNA-seq experiment. 9,281 isoforms were included.

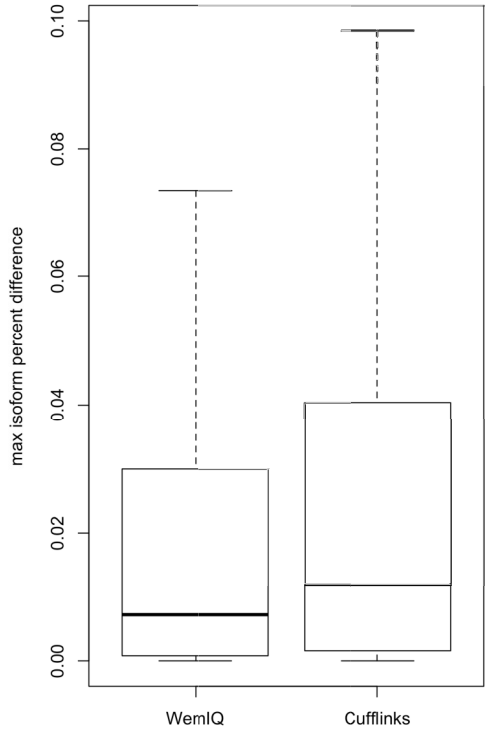

Figure S13. Boxplots of the isoform percentage estimates of the major isoform (ENST00000247655) in gene ENSG00000127184 with different sub-sampling rates from the real RNA-seq data.