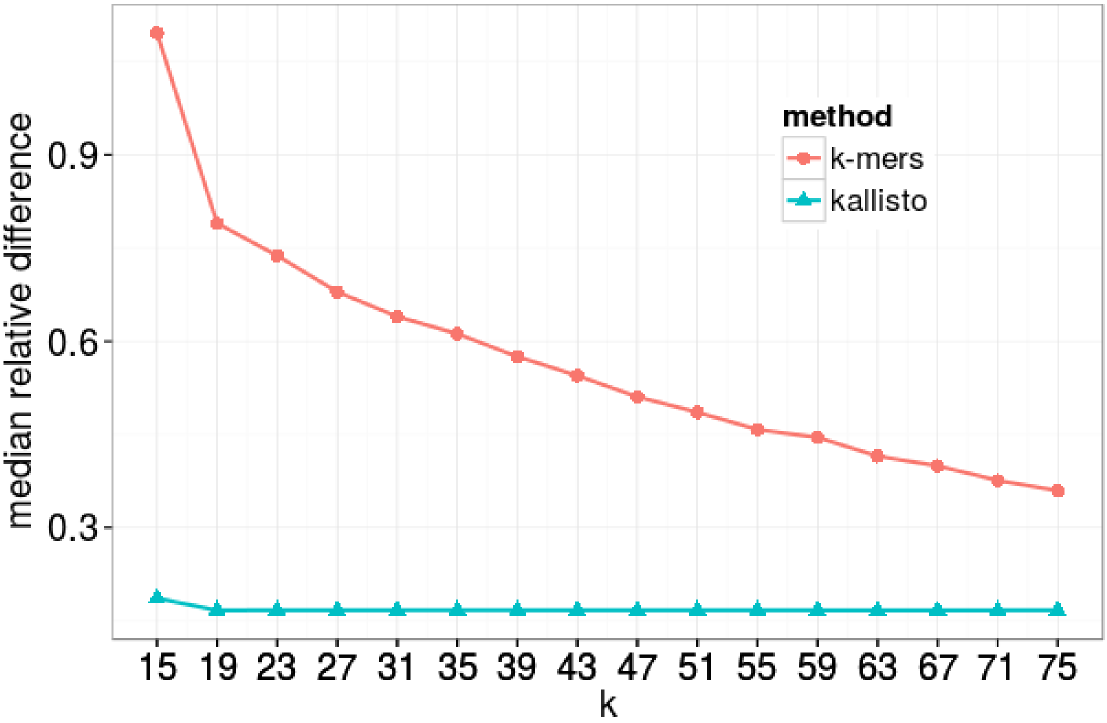
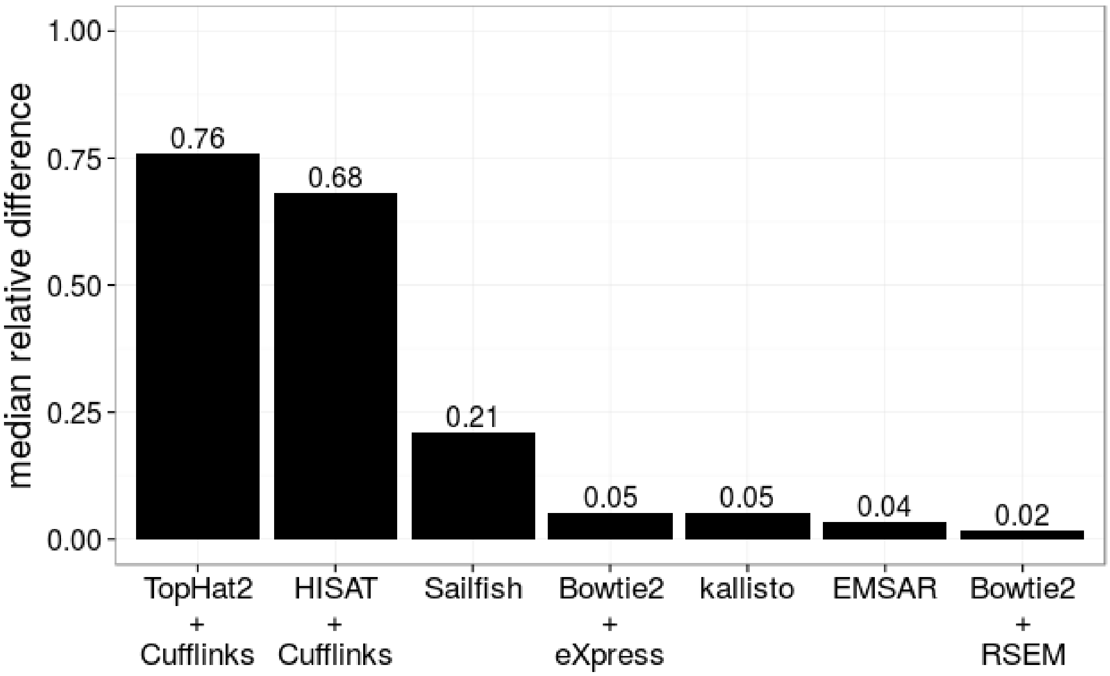


Supplementary Figure 1

Median relative difference for abundance estimates using varying values of *k*.

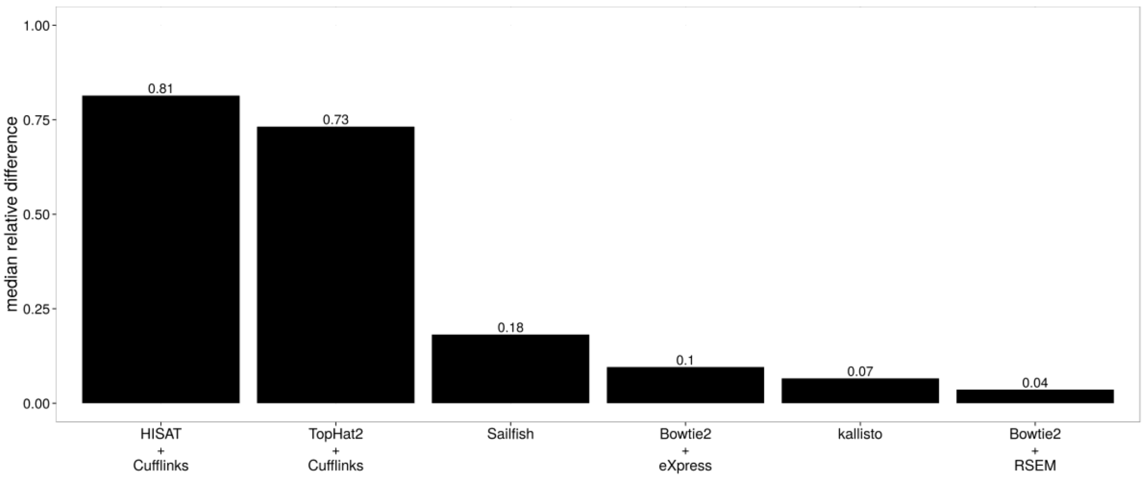
Median relative difference for abundance estimates using varying values of *k* on a dataset of 30 million 75bp paired-end reads that were simulated without errors. The "*k*-mers method" uses the *k*-compatibility of each *k*-mer independently and runs the EM algorithm on *k*-mers, whereas kallisto uses the intersection of *k*-compatibility classes across both ends of a read. Even for *k*=75, the full read length in the simulation, independent use of *k*-mers results in a significant drop in accuracy due to the loss of paired-end information.



Supplementary Figure 2

Accuracy of kallisto, Cufflinks, Sailfish, eXpress and RSEM.

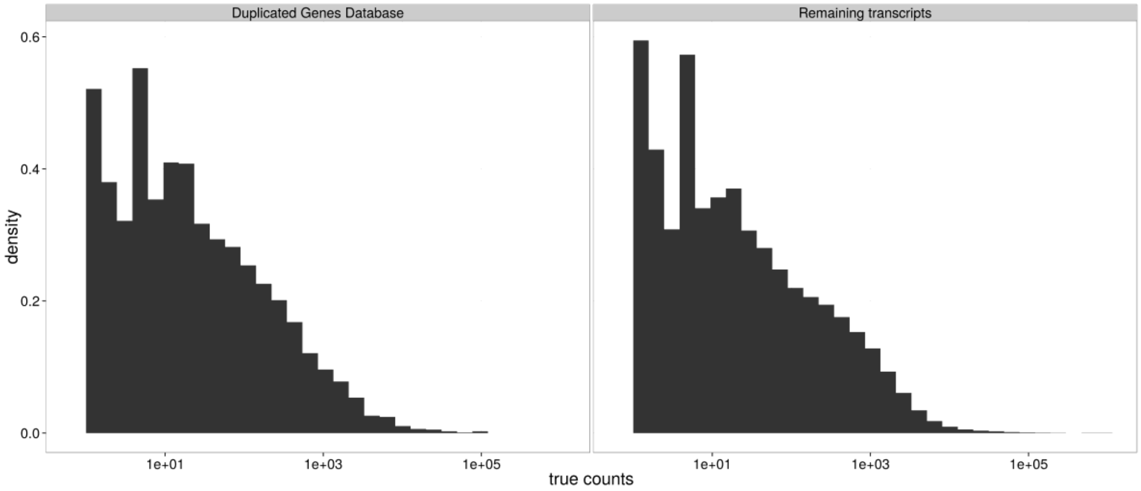
Accuracy of kallisto, Cufflinks, Sailfish, eXpress and RSEM on 20 RSEM simulations of 30 million 75bp paired-end reads based on the TPM estimates and error profile of Geuvadis sample NA12716 (selected for its depth of sequencing). For each simulation we report the accuracy as the median relative difference in the estimated TPM value of each transcript. The values reported are means across the 20 simulations (the variance was too small for this plot). Relative difference is defined as the absolute difference between the estimated TPM values and the ground truth divided by the average of the two.



Supplementary Figure 3

Performance of different quantification programs on the set of paralogs in the human genome.

Performance of different quantification programs on the set of paralogs in the human genome supplied by the Duplicated Genes Database (http://dgd.genouest.org). This set includes 8,636 transcripts in 3,163 genes.

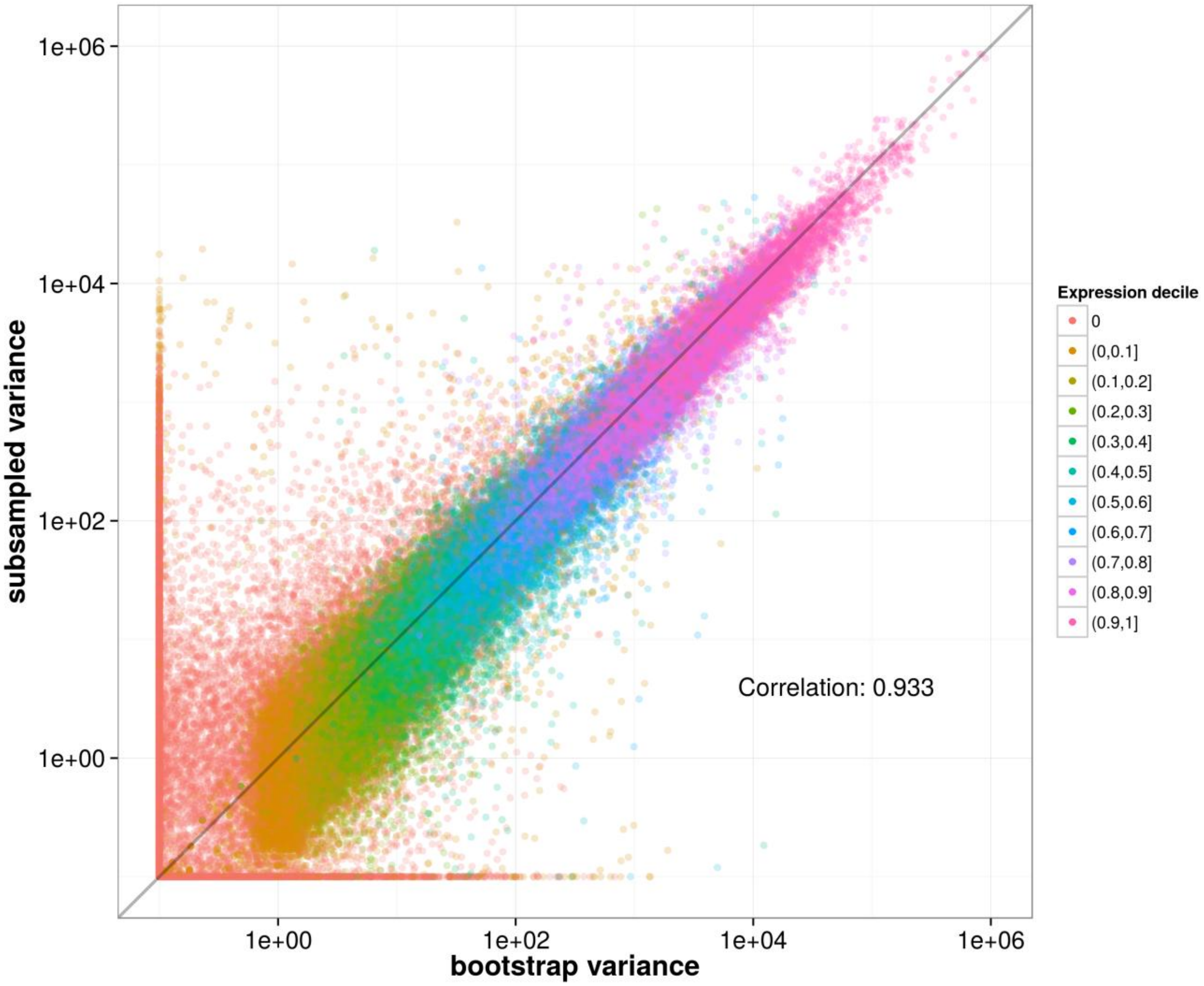


Supplementary Figure 4

Count distribution of one simulation.

Count distribution of one simulation. The left panel contains the transcripts used in Supplementary Figure 3. The right panel contains the remaining transcripts. The x-axis is on the log scale. Both distributions appear very similar, suggesting that the drop in performance in Supplementary Figure 3 is from sequence similarity and not oddities in the distribution such as very low counts.

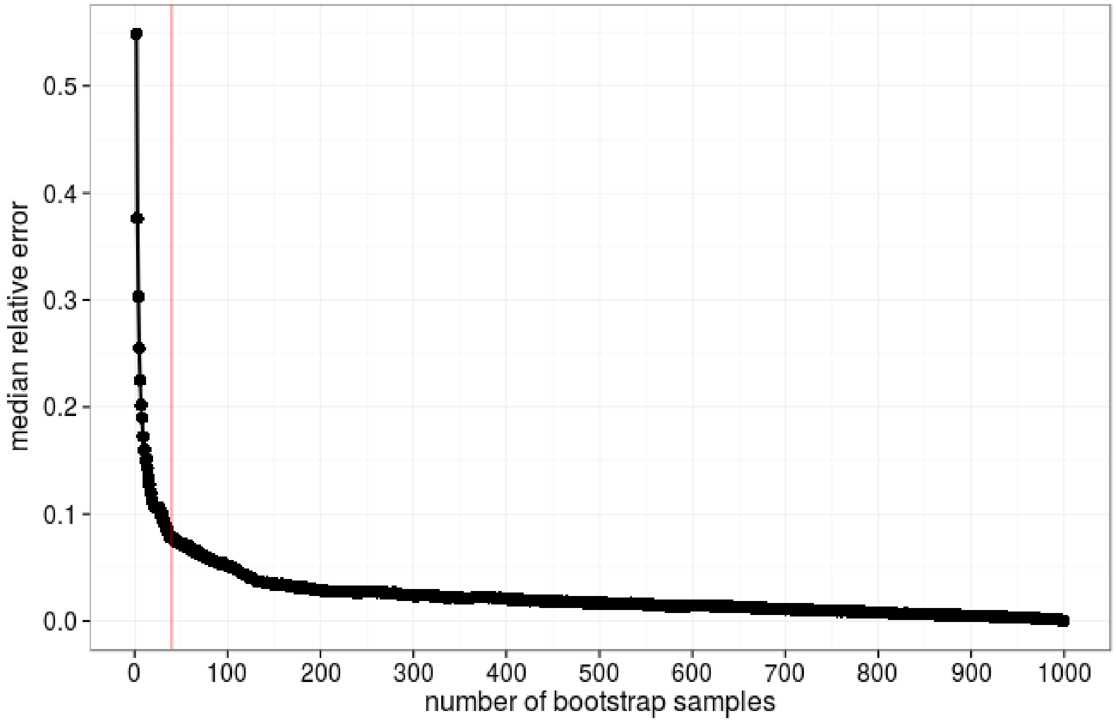
Nature Biotechnology: doi:10.1038/nbt.3519



Supplementary Figure 5

Comparison of technical variance in abundances.

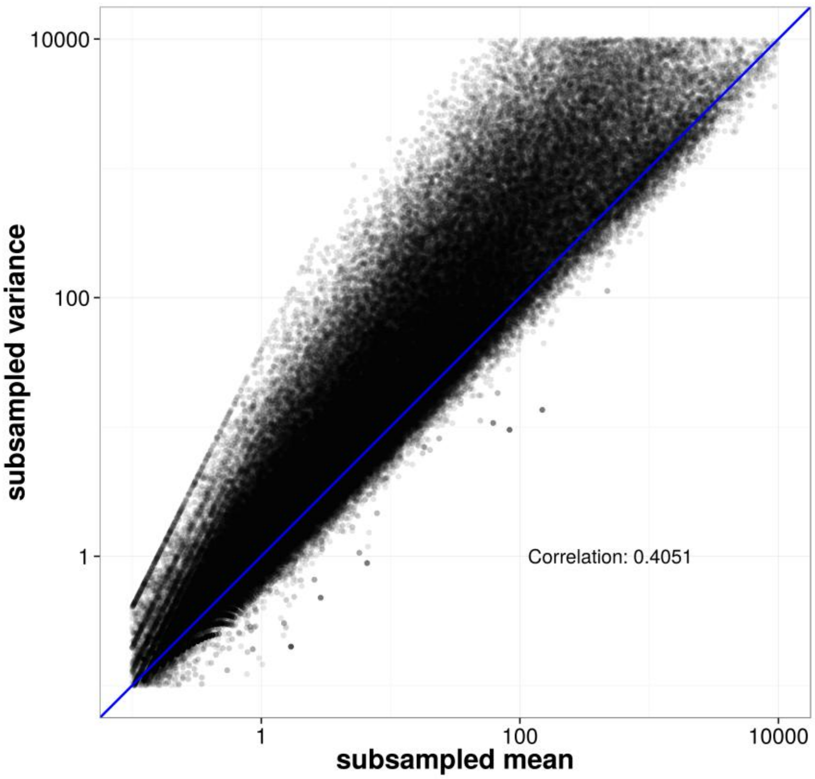
The data comes from a single library with 216M, 101bp paired-end reads sequenced. Each point corresponds to a transcript and is colored by the decile of its expression level in the single bootstrapped subsample. The Y-axis represents variance of abundance estimates across 40 subsamples, with 30M reads in each subsample. The X-axis represents variance as computed from 40 bootstraps of a single subsampled dataset of 30M reads. The red lines emanating from the lower left corner consist of transcripts that have an estimated abundance of zero in the single bootstrapped experiment, but show expression in some of the subsamples (12968 transcripts), and vice versa (720 transcripts).



Supplementary Figure 6

Median relative error (with respect to 1,000 bootstraps) of inferred transcript variances.

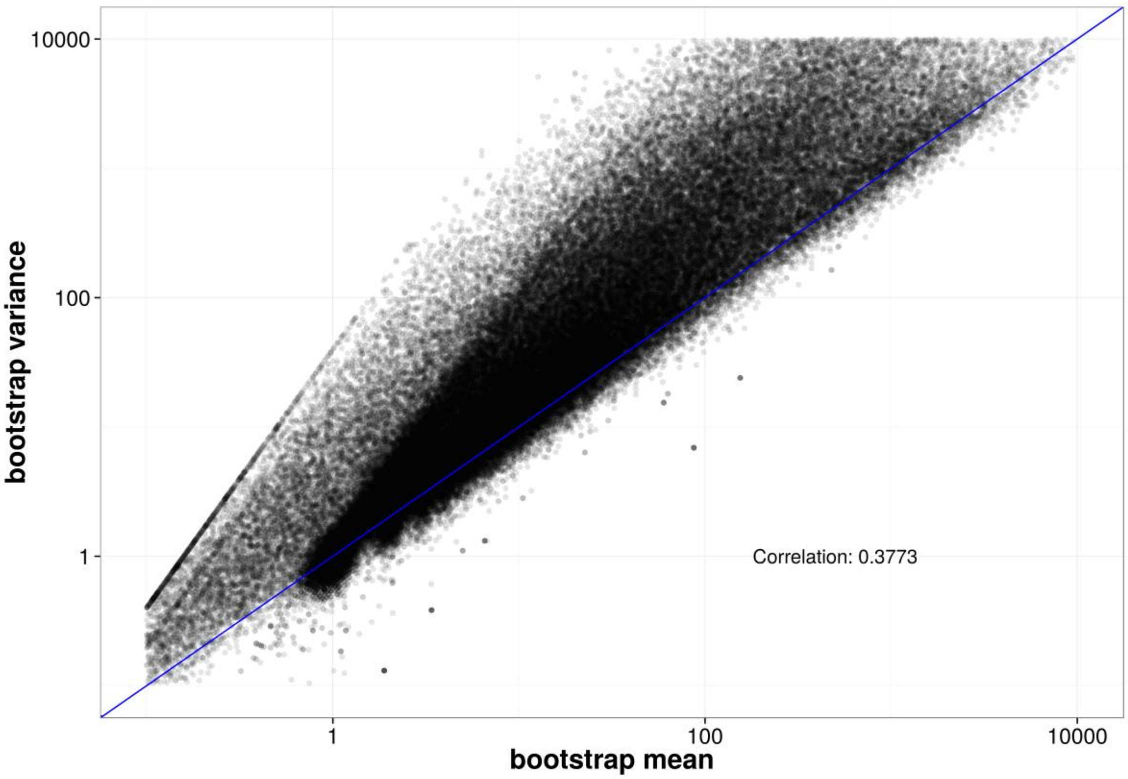
Median relative error (with respect to 1000 bootstraps) of inferred transcript variances as a function of number of bootstrap samples performed. The relative error with 40 bootstraps (red line) is 7.8%.



Supplementary Figure 7

Relationship between the mean and variance of estimated counts from subsamples.

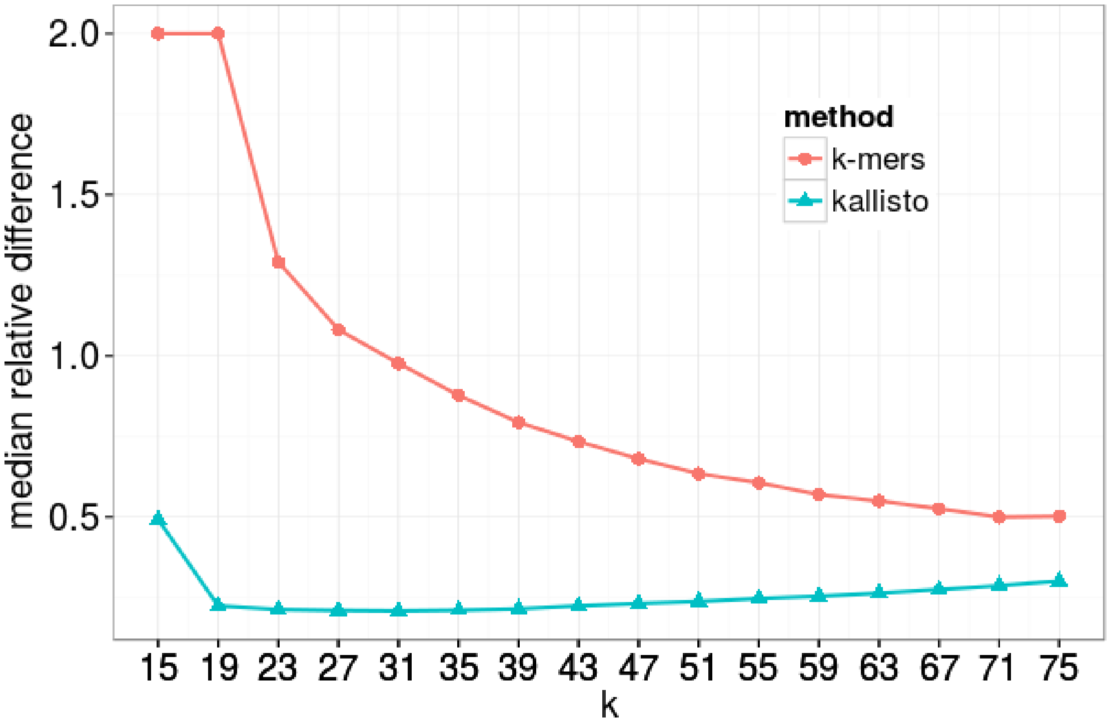
Relationship between the mean and variance of estimated counts for each transcript (x and y axes are on log scale) based on 40 subsamples of 30M reads from a dataset of 216M PE reads. The *x*-axis is the mean of each count estimate calculated across the subsamples. The *y*-axis is the variance of the count estimates calculated across subsamples.



Supplementary Figure 8

Relationship between the mean and variance of estimated counts from bootstraps.

Relationship between the mean and variance of estimated counts for each transcript (x and y axes are on log scale) based on 40 bootstraps of a single subsample of 30M reads from the same 216M PE read dataset. The *x*-axis is the mean of the count estimates calculated across the 40 bootstraps. The *y*-axis is the variance of the count estimates calculated across the 40 bootstraps

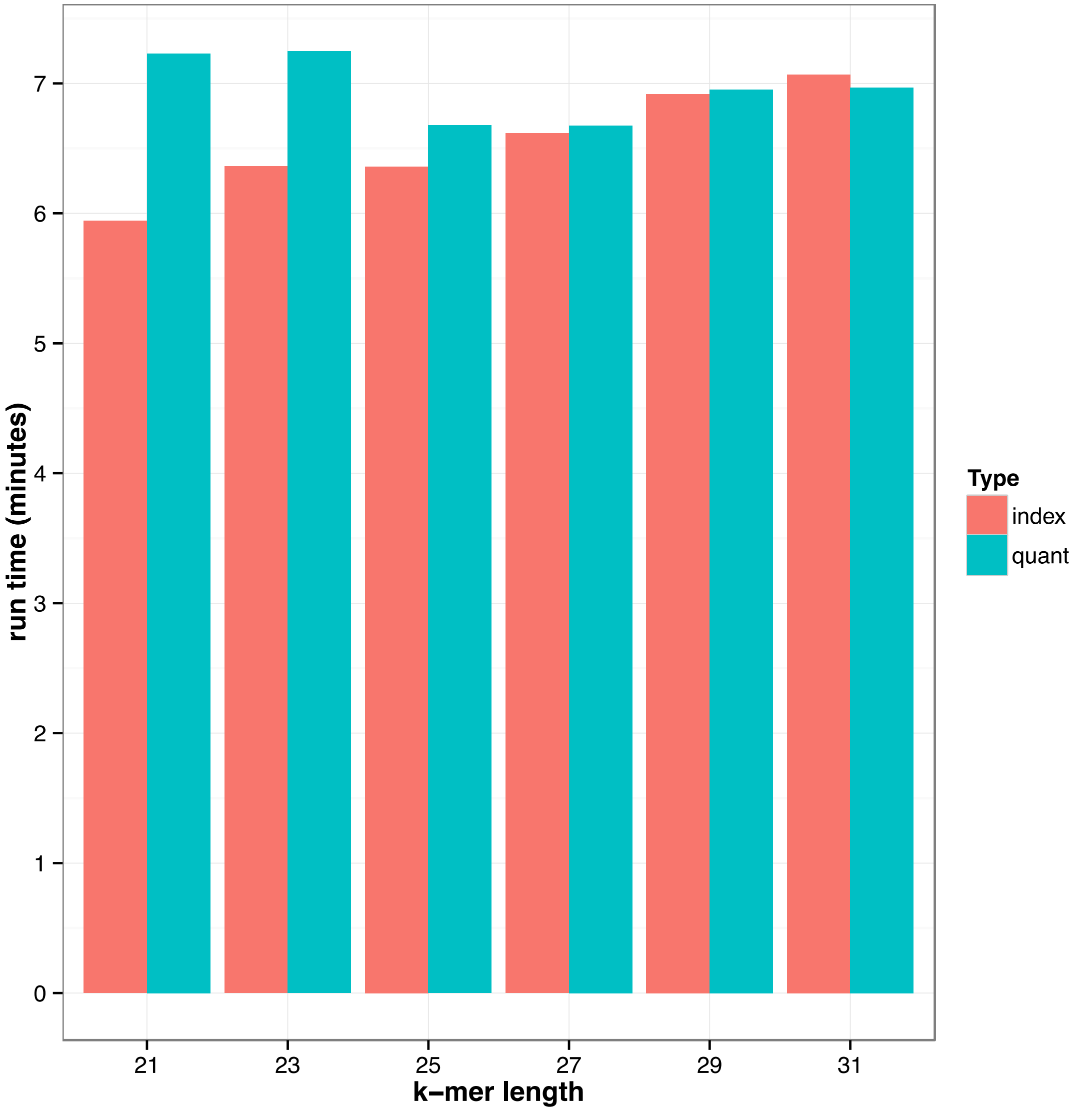


Supplementary Figure 9

Median relative difference from 30 million 75-bp PE reads simulated with error for different values of *k.*

Median relative difference from 30M 75bp PE reads simulated with error for different values of *k*. The "*k*-mers method" uses the *k*-compatibility of each *k*-mer independently and runs the EM algorithm on *k*-mers, whereas kallisto uses the intersection of *k*-compatibility classes across both ends of reads. When there are errors in the reads, kallisto requires smaller *k*-mer lengths for robustness in pseudoalignment.

Nature Biotechnology: doi:10.1038/nbt.3519

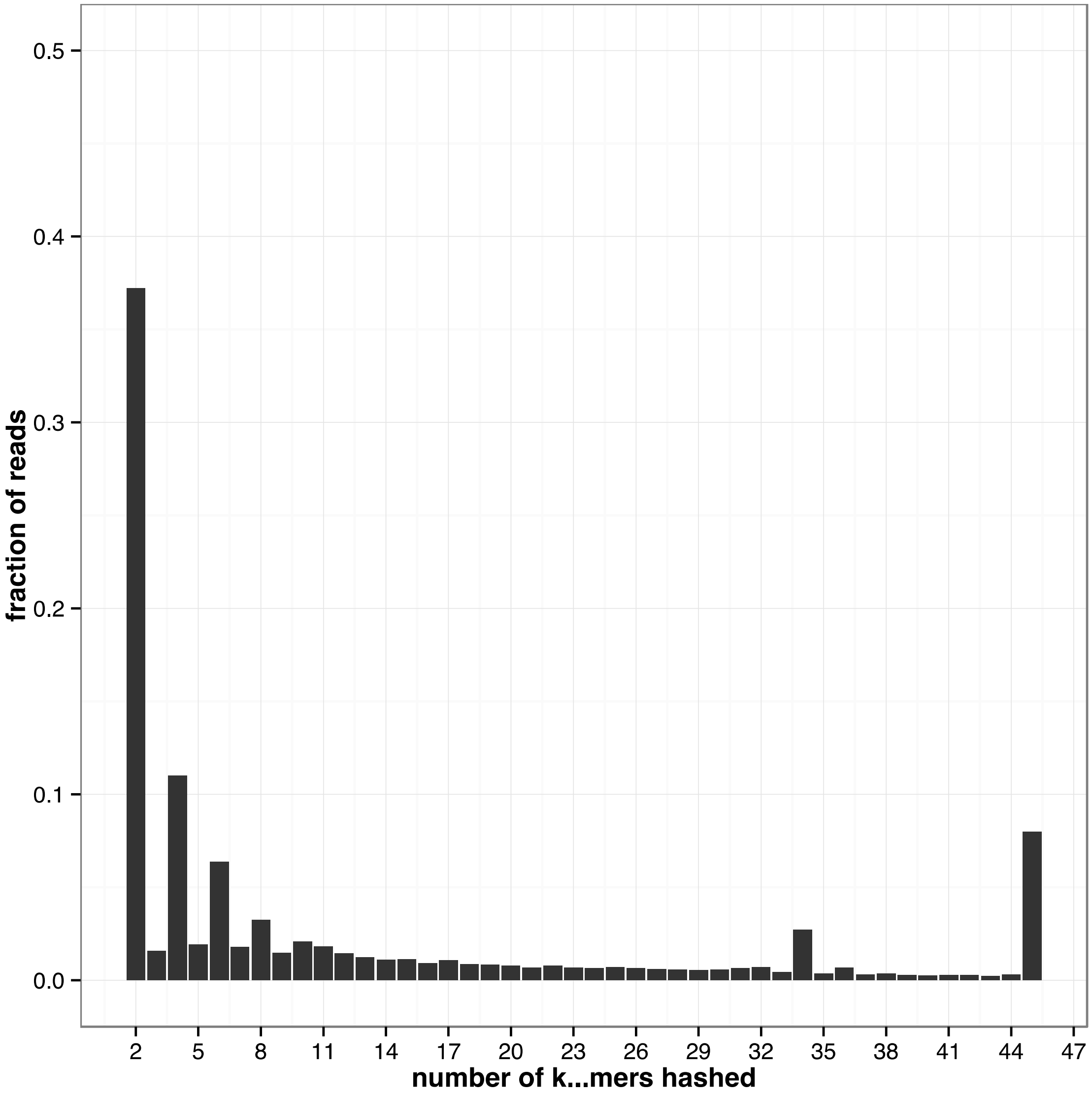


Supplementary Figure 10

Run time for index building and quantification

Run time for index building and quantification as a function of k-mer length for one of the simulated samples.

Nature Biotechnology: doi:10.1038/nbt.3519



Supplementary Figure 11

The distribution of the number of *k*-mers hashed per read.

The distribution of the number of *k*-mers hashed per read for k=31. Note that for the majority of reads (61.35%) only two *k*-mers are hashed. This happens when the entire read pseudoaligns to a single contig of the T-DBG and we can skip to the end of the read. Since we also check the last *k*-mer we can skip over, the most common cases are checking 2, 4, 6, and 8 *k*-mers. Only 1.6% of reads required hashing every *k*-mer of the read.