**Voom/Limma :**

Our preferred practice is to count the total number of reads overlapping annotated exons for each genes. While this approach does not allow for the possibility that different isoforms ofthe same gene may be differentially expressed in different directions, it does provide a statistically robust gene-level

analysis even when the sequencing depths are quite modest.

The voom analysis can also be conducted at the exon level instead of at the gene level as an aid to detecting alternative splicing between the treatment groups.

voom can work, however, just as easily with logged rpkm values in place of log-cpm, because the precision weights are the same for both measures.

If the genomic length of **each gene is known, then the log-cpm values output by voom can be converted to log-rpkm by subtracting**

**the log2 gene length in kilobases.**

From this point of view, the normal-based methods only need

to perform comparably to the count-based methods in

terms of power and FDR control in order to be a success.

Our comparisons suggest not only that this is so, but that

the normal-based methods actually have a performance

advantage.

We found voom to be the best performer

across our simulations and comparisons, and even the

simpler limma-trend method performed equal or better

than the count-based methods. voom and limma-trend

perform almost equally when the library sizes are equal,

but voom has the advantage when the library sizes are

unequal.

The best performing count-based methods were

edgeR and PoissonSeq, although neither of those meth-

ods controlled the type I error rate at the nominal level,

both being somewhat liberal

In our experience markedly unequal library sizes can arise in real

RNA-seq experiments for a variety of reasons. One scenario is when an experiment is conducted in stages and samples sequenced at a later time have a much higher sequencing depth. Other possible scenarios occur when technical replicates are combined for a subset of samples or when DNA samples are multiplexed onto a sequencing lane in unequal quantities. In contrast, voom shows consistent performance in all scenarios.

This shows that traditional statistical methods can not be reliably applied to genomic data without borrowing strength between genes.

The voom and limma-trend methods inherit good robustness properties

from the normal-based procedures in limma [28]. If necessary, they can be made extremely robust to outliers and hypervariable genes using the robust empirical Bayes options of the limma package.

voom has a number of qualitative advantages over the count-based methods.

It is fast and convenient. It allows RNA-seq and microarray data to be analyzed in closely comparable ways, which may be an attraction for analysts comparing results from the two technologies.

voom performs as well or better than existing RNA-seq methods, especially when the library sizes are unequal. It is moreover faster and more convenient, and converts RNA-seq data into a form that can be analyzed using similar tools as for microarrays.

**EDGE-R** :

edgeR is a Bioconductor software package for examining differential expression of replicated count data. An overdispersed Poisson model + empirical Bayes procedure is used to account for both biological and technical variability. Applications beyond sequencing data, such as proteome peptide count data. We model the data as negative binomial (NB) distributed. In this way, our model is able to separate biological from technical variation.

edgeRis the only software for SAGE orDGE data at this time which can account for biological variability when there are only one or two replicate samples.

DesEq2 :