Why TMM and Limma?

Garrett Dancik, PhD

(other than because Dr. Dancik said so)

Why TMM and Limma

- Many normalizations have been developed for RNA-seq data
- Many methods are also available for detecting differentially expressed genes
- TMM and Limma have been shown to be robust and accurate (along with a method called DESeq)

A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis

Marie-Agnès Dillies, Andrea Rau, Julie Aubert, Christelle Hennequet-Antier, Marine Jeanmougin, Nicolas Servant, Céline Keime, Guillemette Marot, David Castel, Jordi Estelle ... Show more

Author Notes

Briefings in Bioinformatics, Volume 14, Issue 6, November 2013, Pages 671–683, https://doi.org/10.1093/bib/bbs046

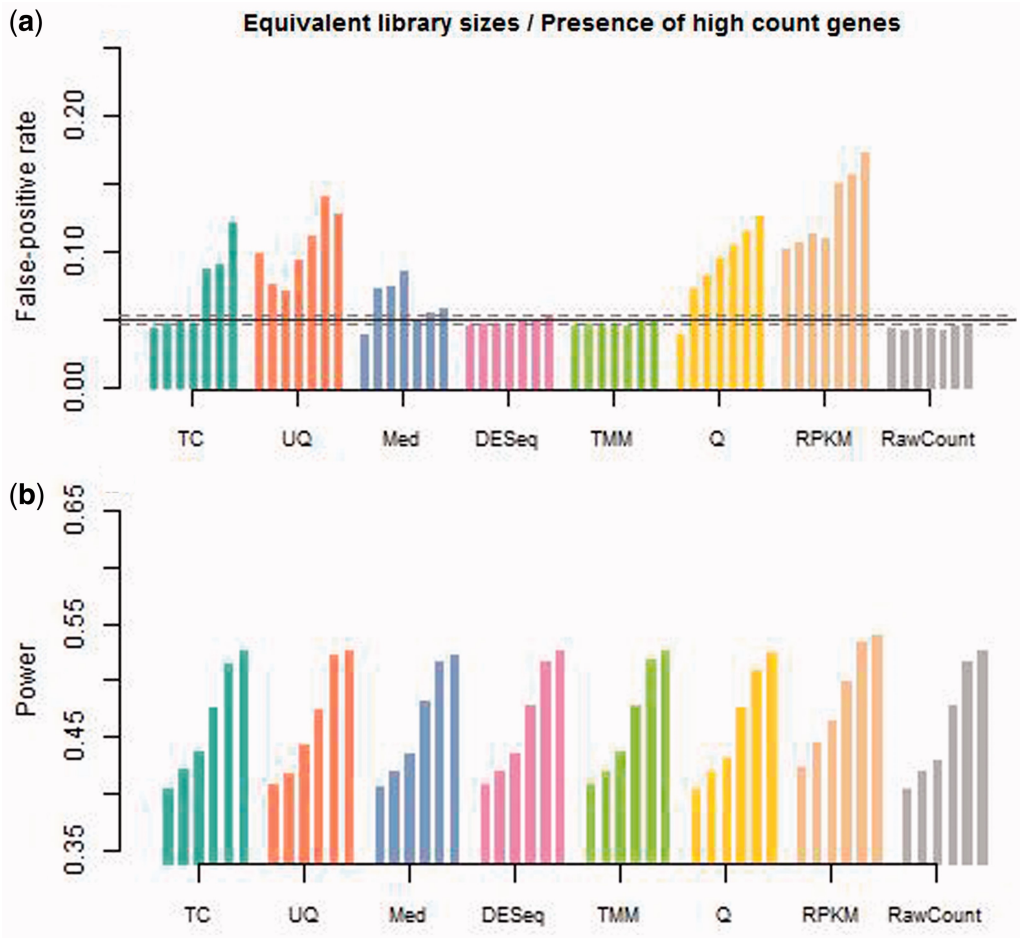
Published: 15 September 2012 **Article history** ▾

https://academic.oup.com/bib/article/14/6/671/189645

Figure 2: Comparison of normalization methods for simulated data with equal library sizes and the presence of ...

The Total Count and RPKM normalization methods, both of which are still widely in use, are ineffective and should be definitively abandoned in the context of differential analysis.

"Only DESeq and TMM are able to control the false-positive rate while also maintaining the power to detect differentially expressed genes."



Brief Bioinform, Volume 14, Issue 6, November 2013, Pages 671–683, https://doi.org/10.1093/bib/bbs046

Method | Open Access | Published: 03 February 2014

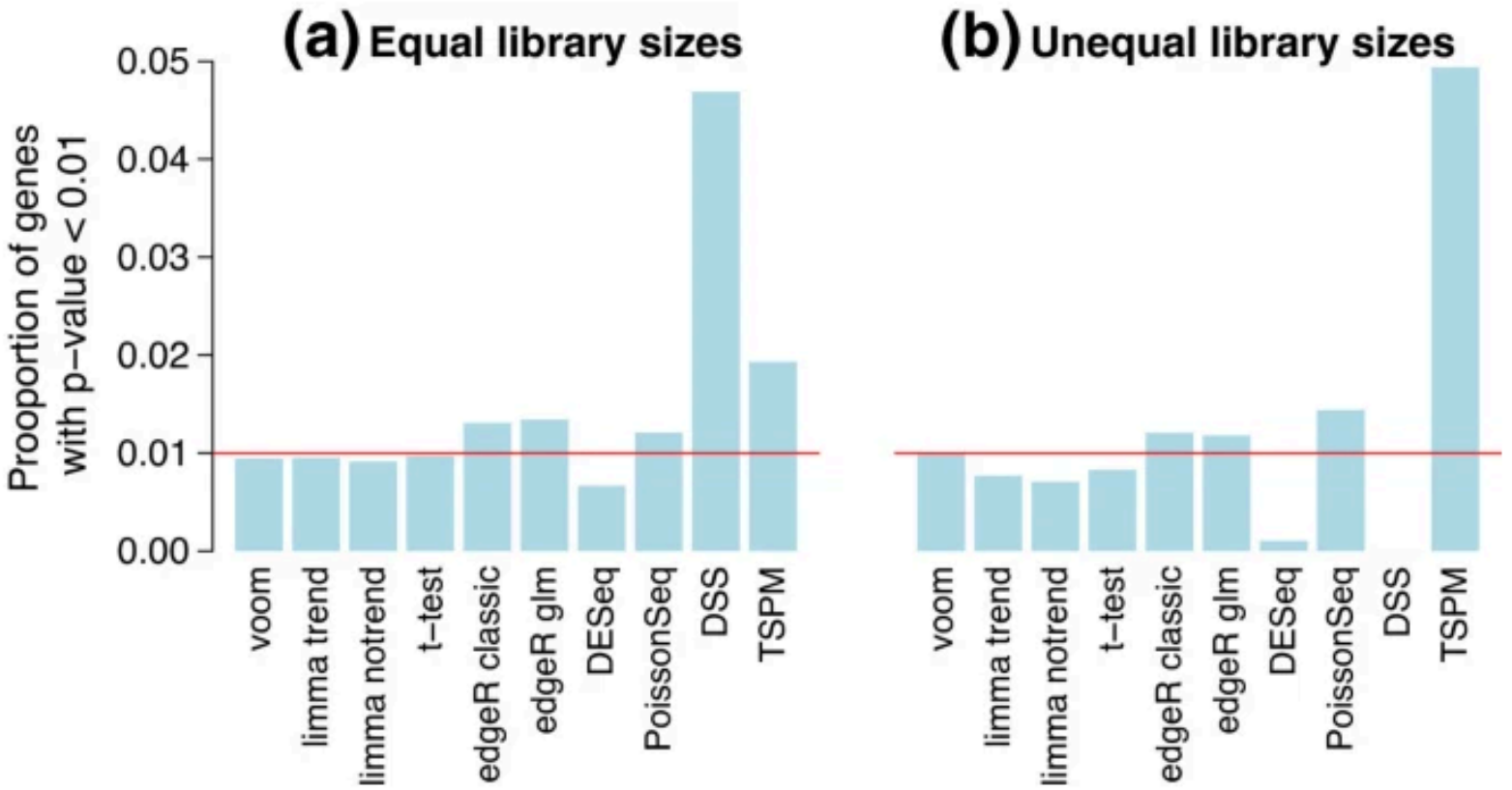
voom: precision weights unlock linear model analysis tools for RNA-seq read counts

Charity W Law, Yunshun Chen, Wei Shi & Gordon K Smyth

Genome Biology **15**, Article number: R29 (2014) | Cite this article

https://genomebiology.biomedcentral.com/articles/10.1186/gb-2014-15-2-r29

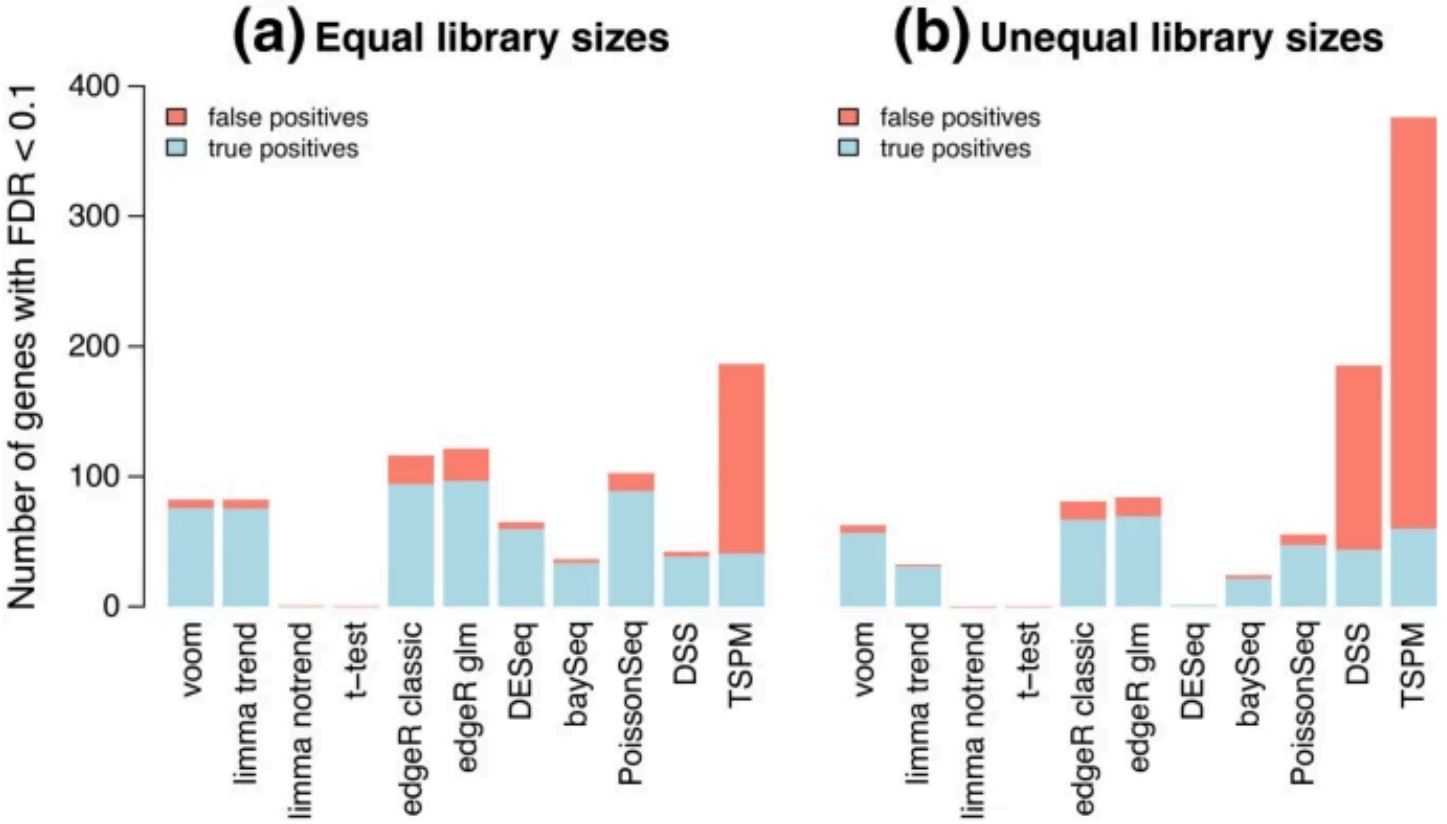
Figure 3



Type I error rates in the absence of true differential expression. The bar plots show the proportion of genes with $P<0.01$ for each method **(a)** when the library sizes are equal and **(b)** when the library sizes are unequal. The red line shows the nominal type I error rate of 0.01. Results are averaged over 100 simulations. Methods that control the type I error at or below the nominal level should lie below the red line.

Full size image ›

Figure 4



Power to detect true differential expression. Bars show the total number of genes that are detected as statistically significant (FDR < 0.1) **(a)** with equal library sizes and **(b)** with unequal library sizes. The blue segments show the number of true positives while the red segments show false positives. 200 genes are genuinely differentially expressed. Results are averaged over 100 simulations. Height of the blue bars shows empirical power. The ratio of the red to blue segments shows empirical FDR. FDR, false discovery rate.

Full size image ›

Comparison of software packages for detecting differential expression in RNA-seq studies

Fatemeh Seyednasrollah, Asta Laiho, Laura L. Elo

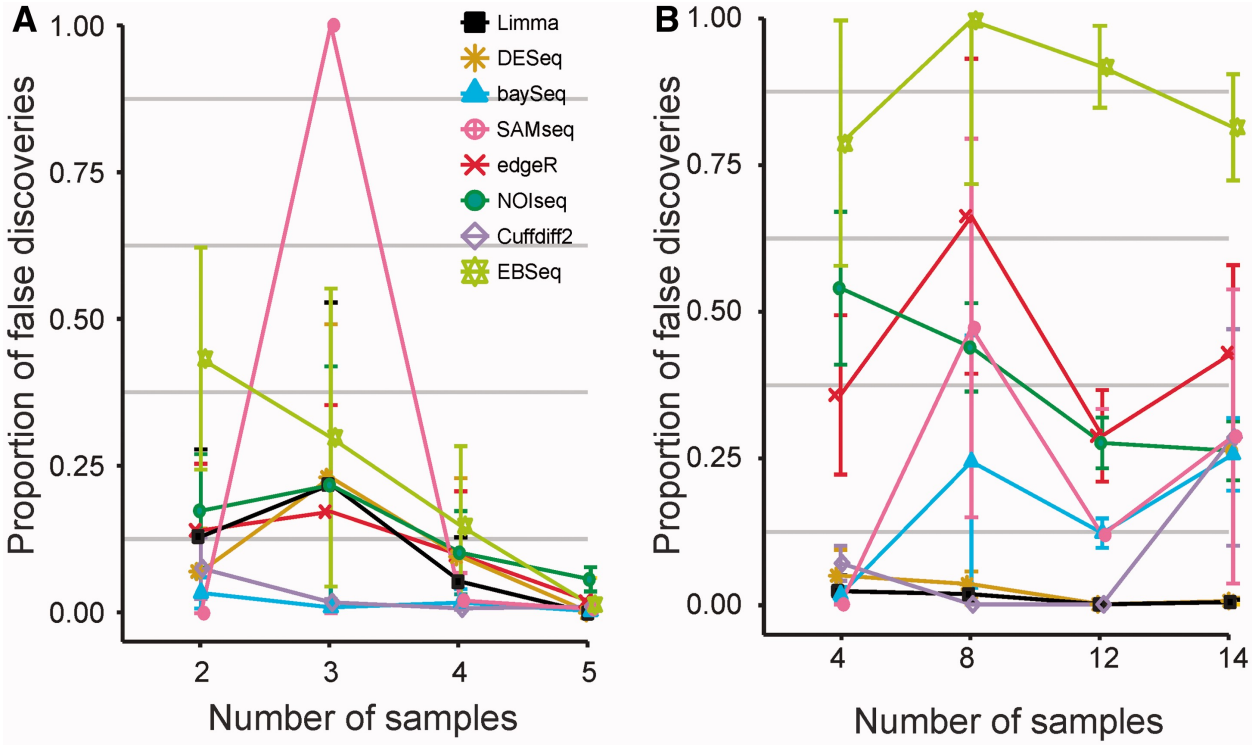
Briefings in Bioinformatics, Volume 16, Issue 1, January 2015, Pages 59–70,
https://doi.org/10.1093/bib/bbt086
Published: 02 December 2013 **Article history** ▾

https://academic.oup.com/bib/article/16/1/59/240754

Table 1: Software packages for detecting differential expression

Method	Version	Reference	Normalization[a]
edgeR	3.0.8	[4]	TMM/Upper quartile/RLE (DESeq-like
DESeq	1.10.1	[5]	DESeq sizeFactors
baySeq	1.12.0	[6]	Scaling factors (quantile/TMM/total)
NOIseq	1.1.4	[7]	RPKM/TMM/Upper quartile
SAMseq (samr)	2.0	[8]	SAMseq specialized method based on
Limma	3.14.4	[9]	TMM
Cuffdiff 2 (Cufflinks)	2.0.2-beta	[10]	Geometric (DESeq-like)/quartile/clas
EBSeq	1.1.7	[11]	DESeq median normalization

Figure 3: False discoveries on the basis of mock comparisons in the (A) mouse and (B) human data. In each mock ...



"For instance, in the inherently more heterogeneous human data, only DESeq and limma were able to produce low rates of false positives even if the number of samples was increased"

Brief Bioinform, Volume 16, Issue 1, January 2015, Pages 59–70, https://doi.org/10.1093/bib/bbt086