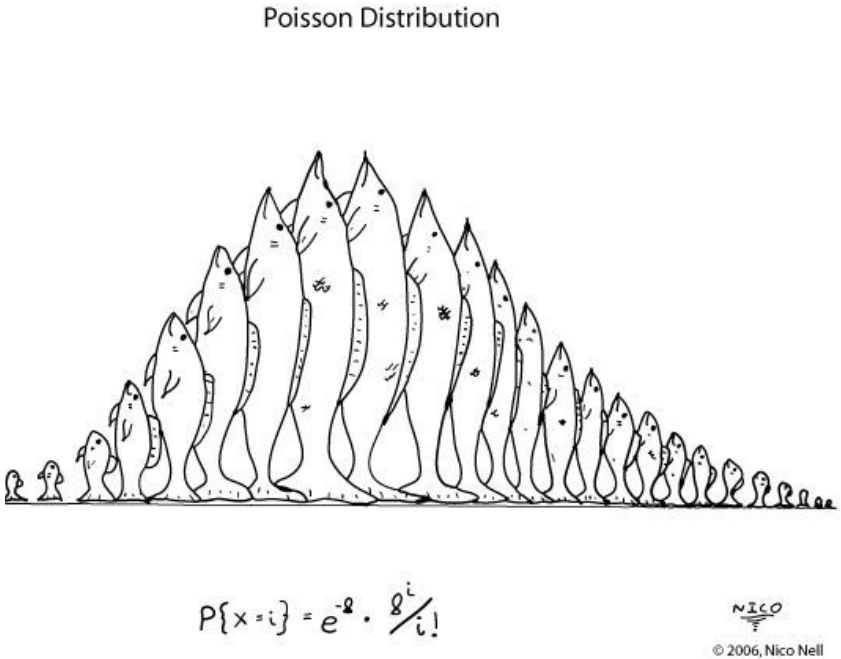
Differential Expression: Background

- Goal: determine if difference in read counts between samples is greater than natural random variation

- Assumption: if you randomly sample reads from a population of transcripts, the read counts should follow a Poisson distribution

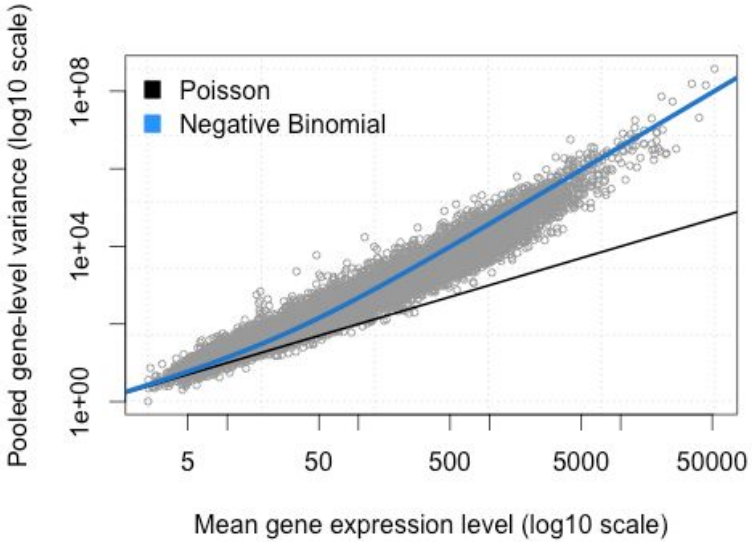


Differential Expression: Background

- Problem: Poisson distribution assumes that variance = mean

- Variance in read counts is actually very high, esp for highly expr genes = **overdispersion**

- **Negative binomial** approach allows for more variance (noise) in RNAseq data = includes *dispersion parameter*



Anders & Huber, 2010

Jesse Lipp, BIORAMBLE blog

Differential Expression Analysis

How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use?

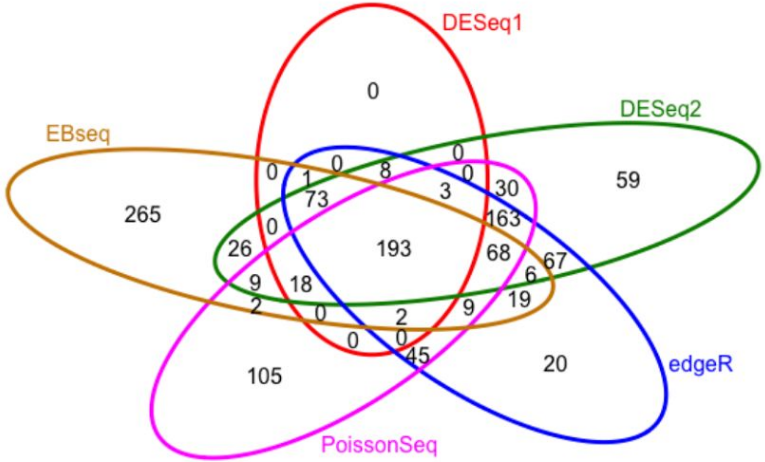
NICHOLAS J. SCHURCH,[1,6] PIETÅ SCHOFIELD,[1,2,6] MAREK GIERLIŃSKI,[1,2,6] CHRISTIAN COLE,[1,6] ALEXANDER SHERSTNEV,[1,6] VIJENDER SINGH,[2] NICOLA WROBEL,[3] KARIM GHARBI,[3] GORDON G. SIMPSON,[4] TOM OWEN-HUGHES,[2] MARK BLAXTER,[3] and GEOFFREY J. BARTON[1,2,5]

[1]Division of Computational Biology, College of Life Sciences, University of Dundee, Dundee DD1 5EH, United Kingdom
[2]Division of Gene Regulation and Expression, College of Life Sciences, University of Dundee, Dundee DD1 5EH, United Kingdom
[3]Edinburgh Genomics, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom
[4]Division of Plant Sciences, College of Life Sciences, University of Dundee, Dundee DD1 5EH, United Kingdom
[5]Division of Biological Chemistry and Drug Discovery, College of Life Sciences, University of Dundee, Dundee DD1 5EH, United Kingdom

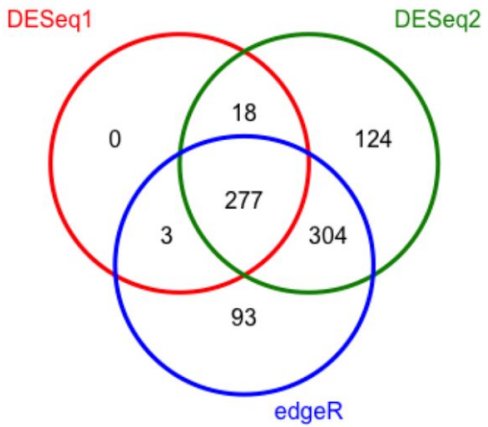
Schurch et al 2016

Differential Expression Analysis

- Different tools give different answers



Unique objects: All = 1191; S1 = 298; S2 = 723; S3 = 677; S4 = 647; S5 = 691



Unique objects: All = 819; S1 = 298; S2 = 723; S3 = 677

Cohen & Li
https://github.com/ljcohen/ECE221_final_project/blob/master/Cohen_Li_ECE221_review-differential-gene.pdf

Differential Expression Analysis: Many Tools, Many Answers

Trade-offs between:

- false positive rate
- true positive rate
- statistical power

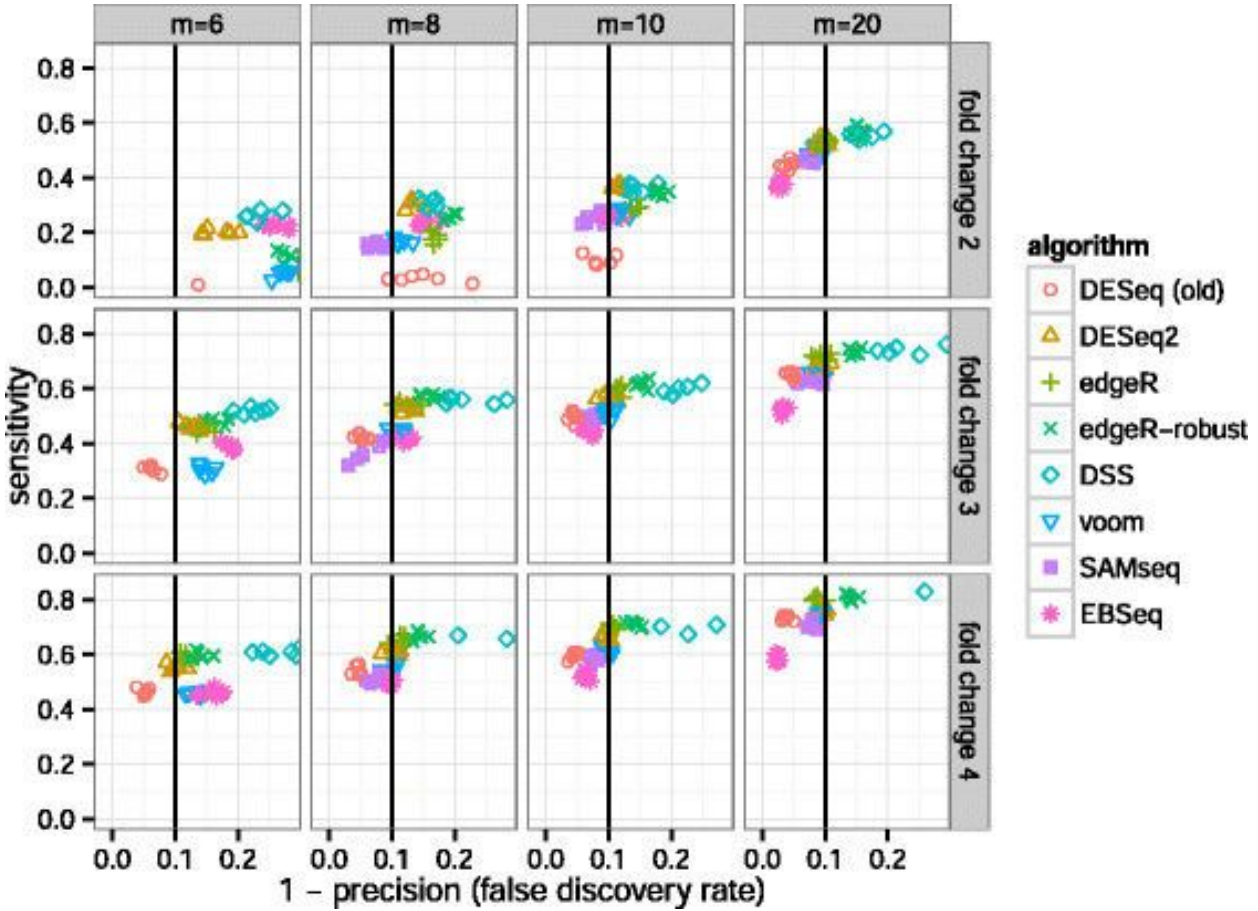
TABLE 1. RNA-seq differential gene expression tools and statistical tests

Name	Assumed distribution	Normalization	Description	Version	Citations[d]	Reference
t-test	Normal	DEseq[a]	Two-sample *t*-test for equal variances	–	–	–
log *t*-test	Log-normal	DEseq[a]	Log-ratio *t*-test	–	–	–
Mann-Whitney	None	DEseq[a]	Mann-Whitney test	–	–	Mann and Whitney (1947)
Permutation	None	DEseq[a]	Permutation test	–	–	Efron and Tibshirani (1993a)
Bootstrap	Normal	DEseq[a]	Bootstrap test	–	–	Efron and Tibshirani (1993a)
baySeq[c]	Negative binomial	Internal	Empirical Bayesian estimate of posterior likelihood	2.2.0	159	Hardcastle and Kelly (2010)
Cuffdiff	Negative binomial	Internal	Unknown	2.1.1	918	Trapnell et al. (2012)
DECseq[c]	Binomial	None	Random sampling model using Fisher's exact test and the likelihood ratio test	1.22.0	325	Wang et al. (2010)
DESeq[c]	Negative binomial	DEseq[a]	Shrinkage variance	1.20.0	1889	Anders and Huber (2010)
DESeq2[c]	Negative binomial	DEseq[a]	Shrinkage variance with variance based and Cook's distance pre-filtering	1.8.2	197	Love et al. (2014)
EBSeq[c]	Negative binomial	DEseq[a] (median)	Empirical Bayesian estimate of posterior likelihood	1.8.0	80	Leng et al. (2013)
edgeR[c]	Negative binomial	TMM[b]	Empirical Bayes estimation and either an exact test analogous to Fisher's exact test but adapted to over-dispersed data or a generalized linear model	3.10.5	1483	Robinson et al. (2010)
Limma[c]	Log-normal	TMM[b]	Generalized linear model	3.24.15	97	Law et al. (2014)
NOISeq[c]	None	RPKM	Nonparametric test based on signal-to-noise ratio	2.14.0	177	Tarazona et al. (2011)
PoissonSeq[c]	Poisson log-linear model	Internal	Score statistic	1.1.2	37	Li et al. (2012)
SAMSeq[c]	None	Internal	Mann-Whitney test with Poisson resampling	2.0	54	Li and Tibshirani (2013)

Schurch et al 2016

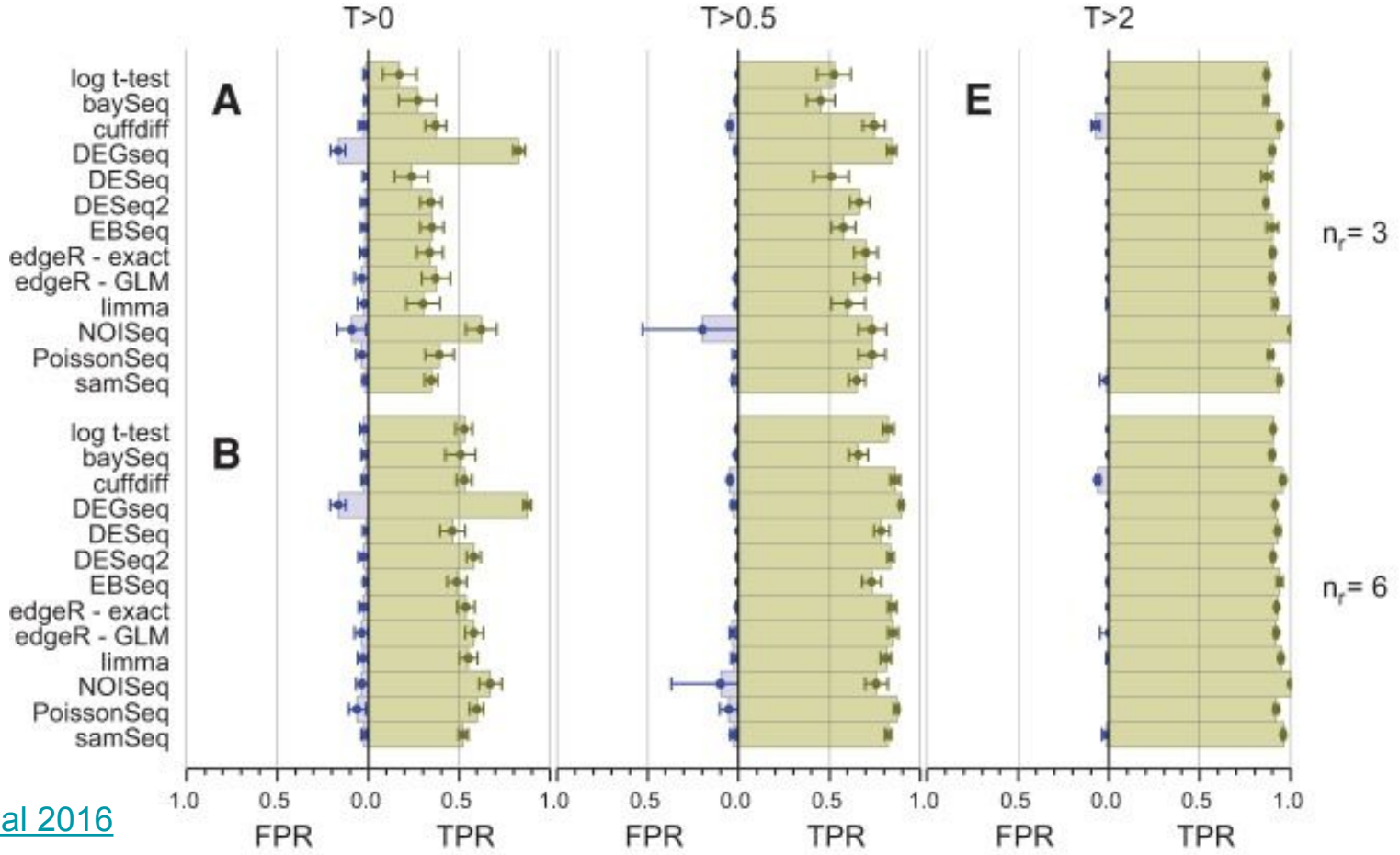
Differential Expression Analysis: edgeR and DESeq2

What about with even fewer replicates?



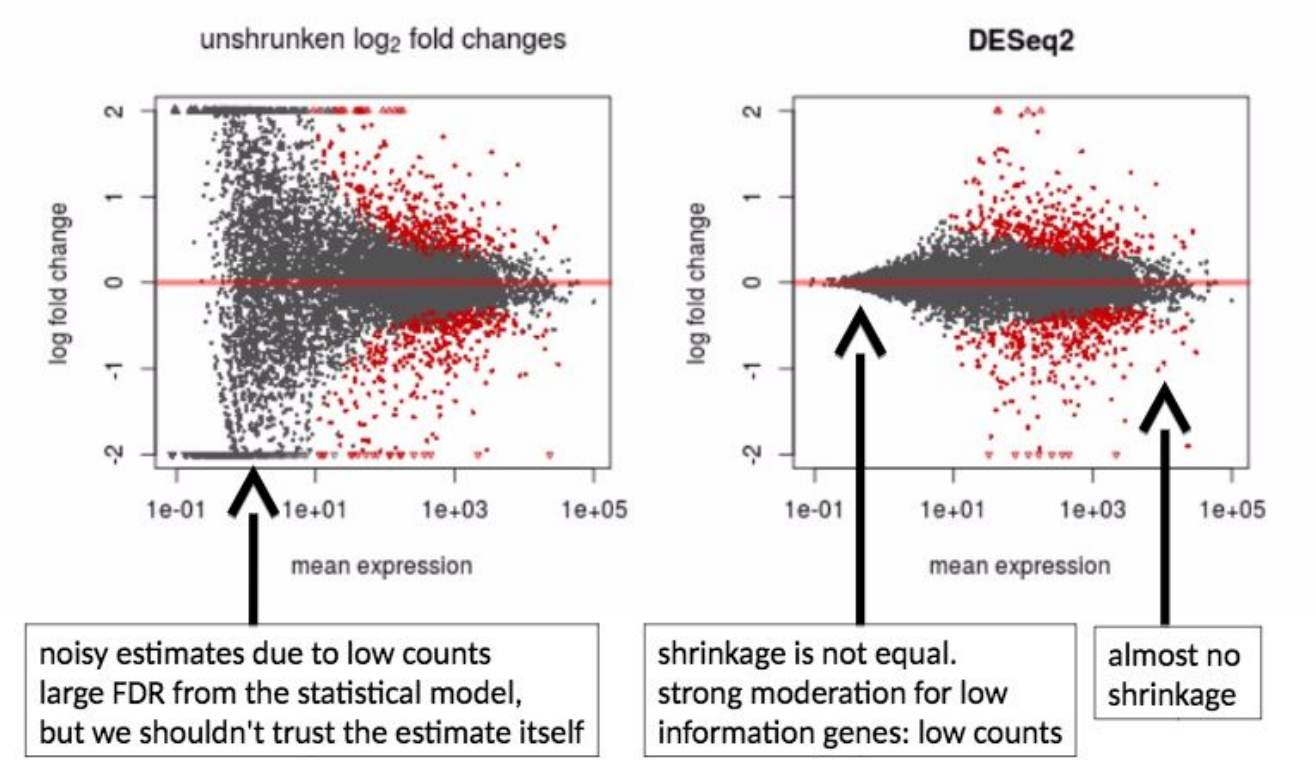
Love, Huber, & Anders 2014

Differential Expression Analysis: edgeR and DESeq2



Schurch et al 2016

DESeq2 and shrinkage of fold changes

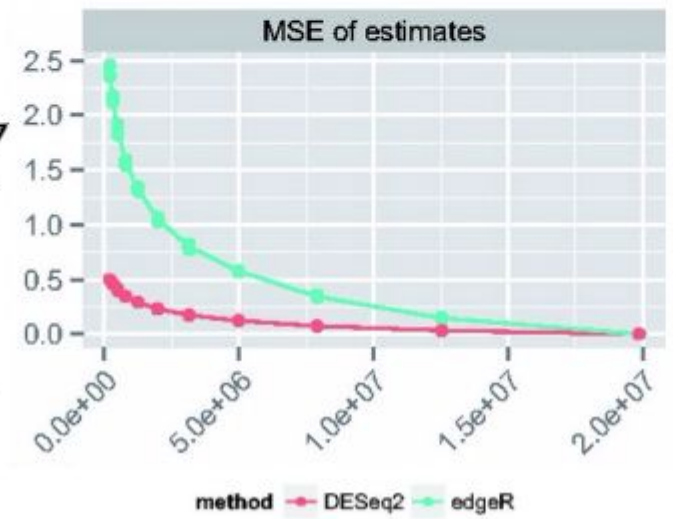


Mike Love, ANGUS 2017

Why shrink fold changes?

Estimates remain reliable even for small sample size and low counts:

Mean Squared Error when subsampling to lower seq. depth in silico



Mike Love, ANGUS 2017