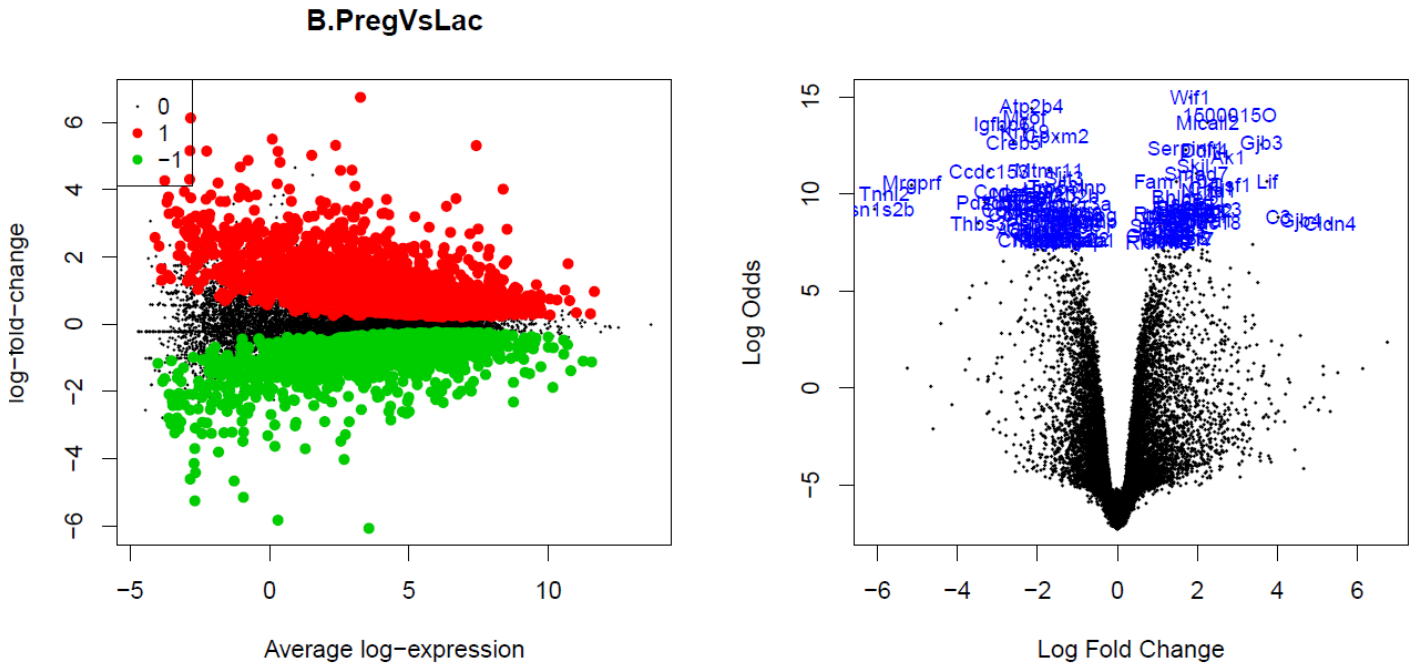
Gene set testing

RNASeq Workshop

September 2016

Why?

- Sometimes after differential expression testing, we have a long list of 1000's of genes
- Too difficult to go through one by one
- Want to understand pathways involved in the biological system being studied

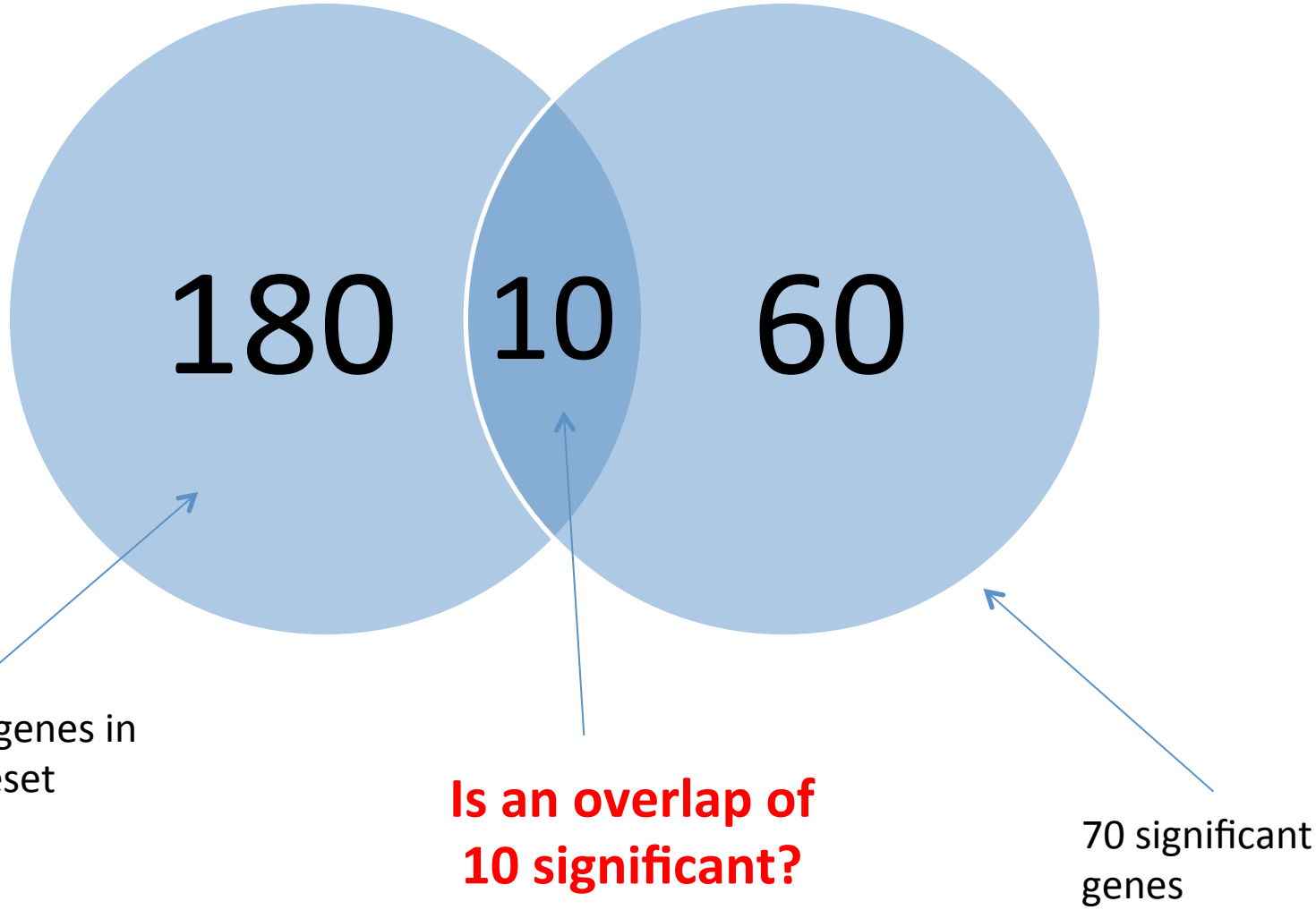


Gene set tests available in limma

- Want to test LOTS of gene sets?
  - goana() function
    - Gene Ontology (GO) analysis
  - camera() function
    - User specified gene sets
- Want to test just a few gene sets?
  - roast() function

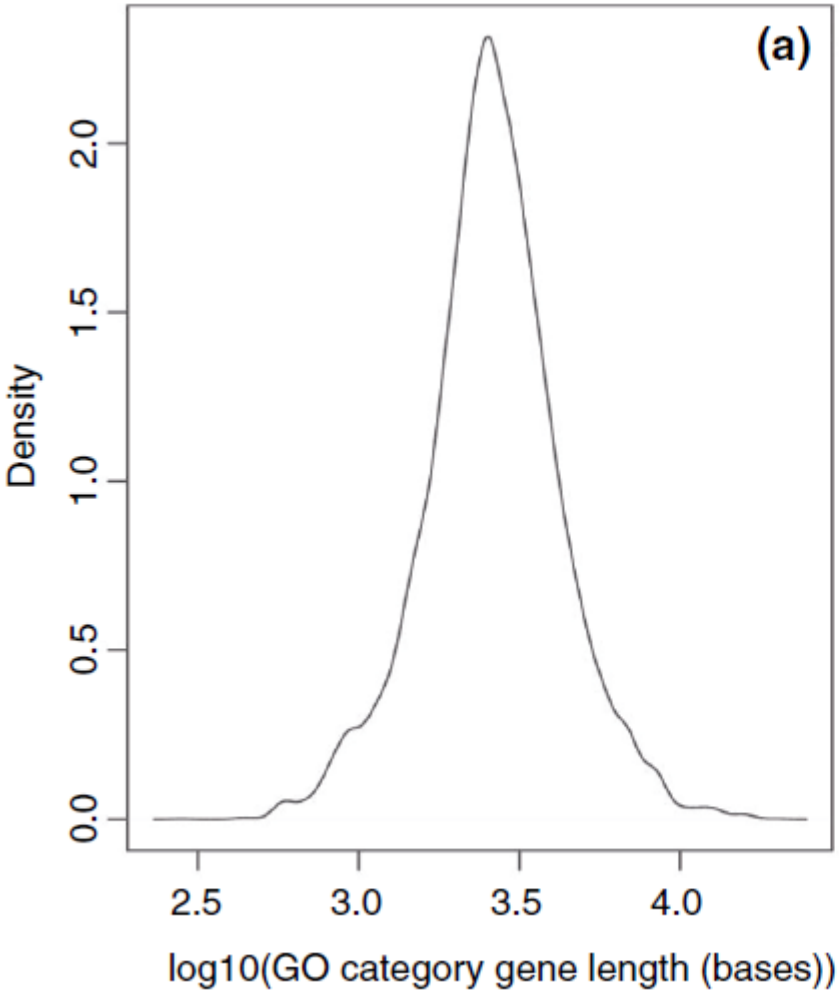
Basic principles behind gene set testing

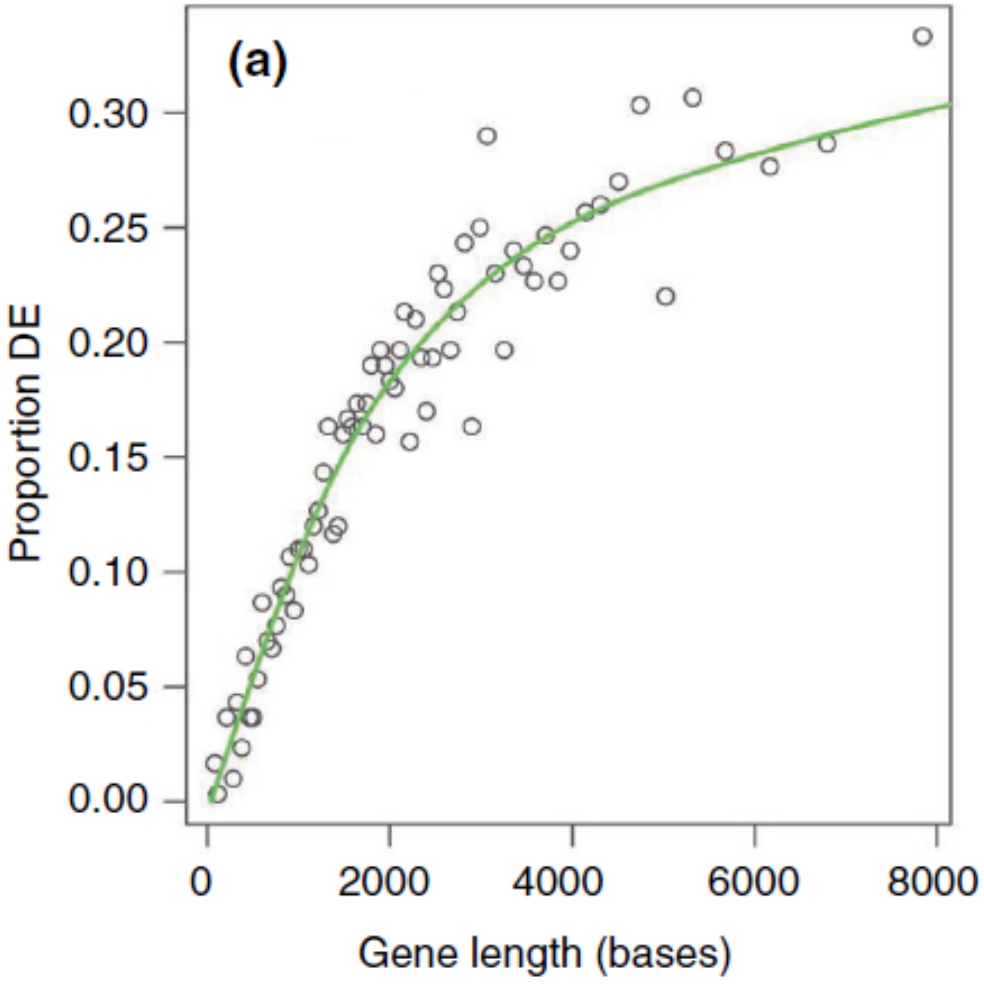
“Overlap” analysis: goana, DAVID, toppfun, GOstats (& most web-based tools)



Problem: this test is biased due to the fact that longer genes tend to have more reads assigned to them

GO categories have different avg gene lengths





GOseq, Young et al, 2010

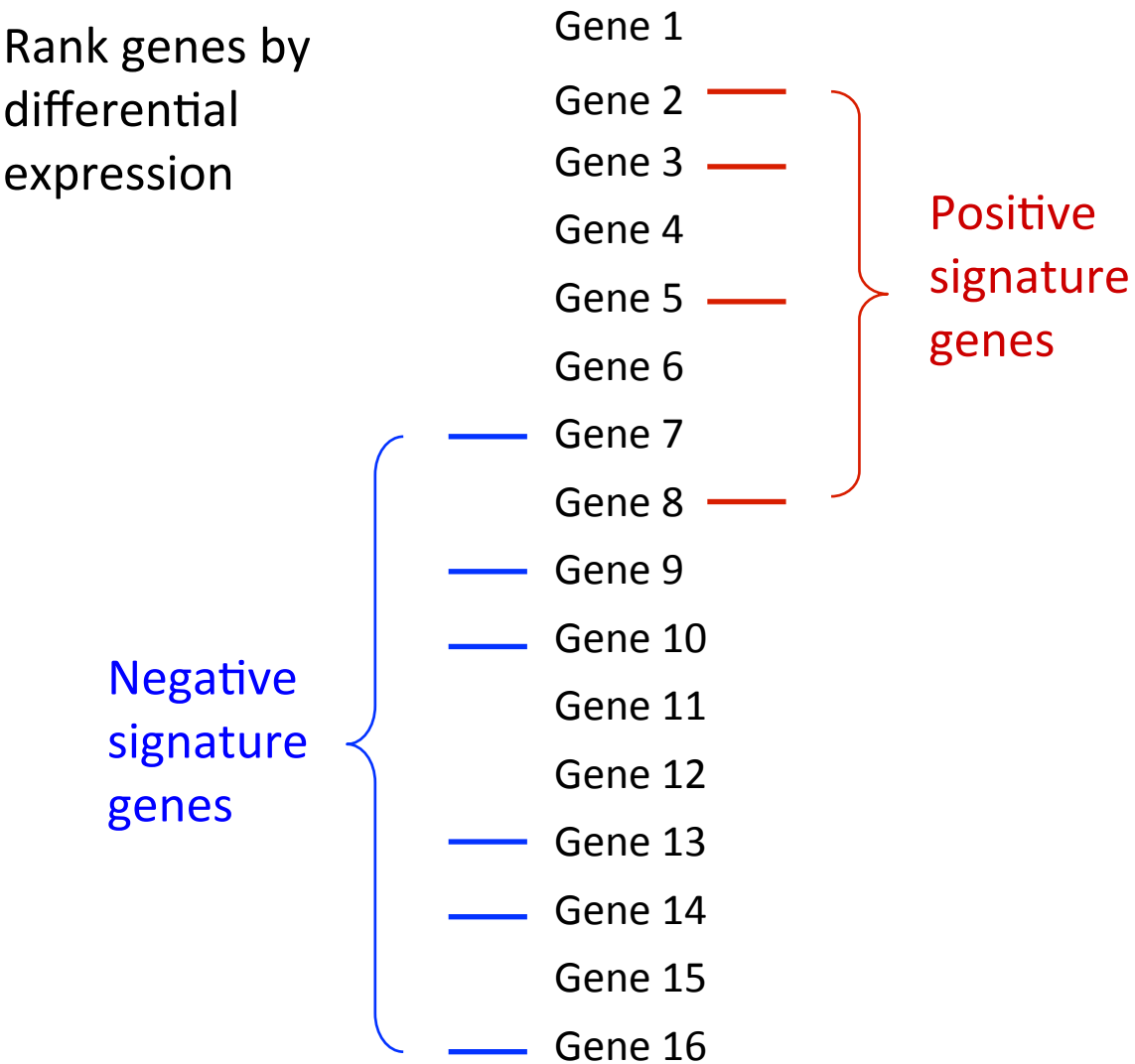
Solution: take into account gene length in your GO analysis

- goana() has the ability to take into account gene length using the "covariate" argument
- The GOseq bioconductor package contains the original method

CAMERA

- An "overlap" analysis assumes the genes are independent
- CAMERA tests the ranking of the gene set relative to the other genes in the experiment, while taking into account inter-gene correlations
- It also takes into account strength of evidence of DE by using the moderated t-statistics

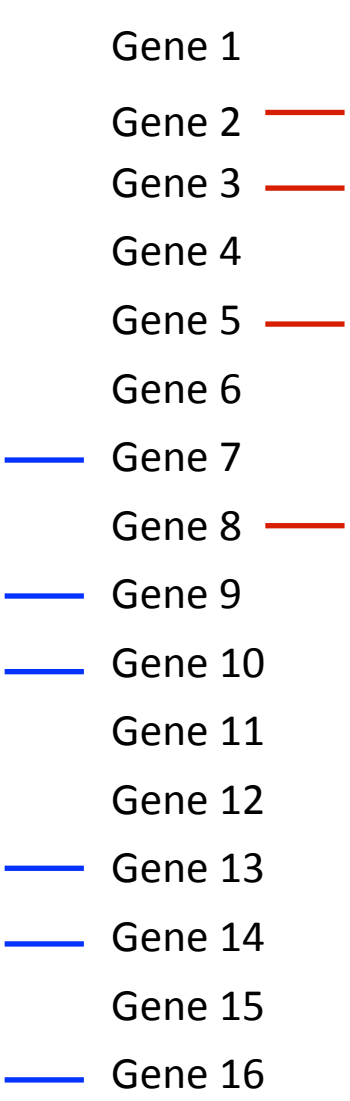
Rank genes and mark signature

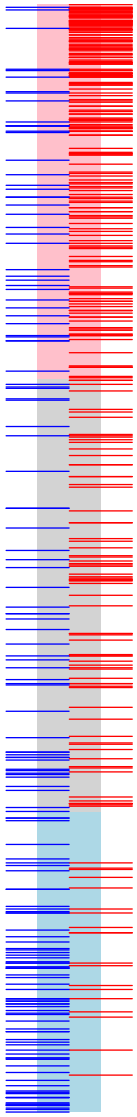


Slide courtesy of Gordon Smyth

Rank genes and mark signature

Rank genes by differential expression

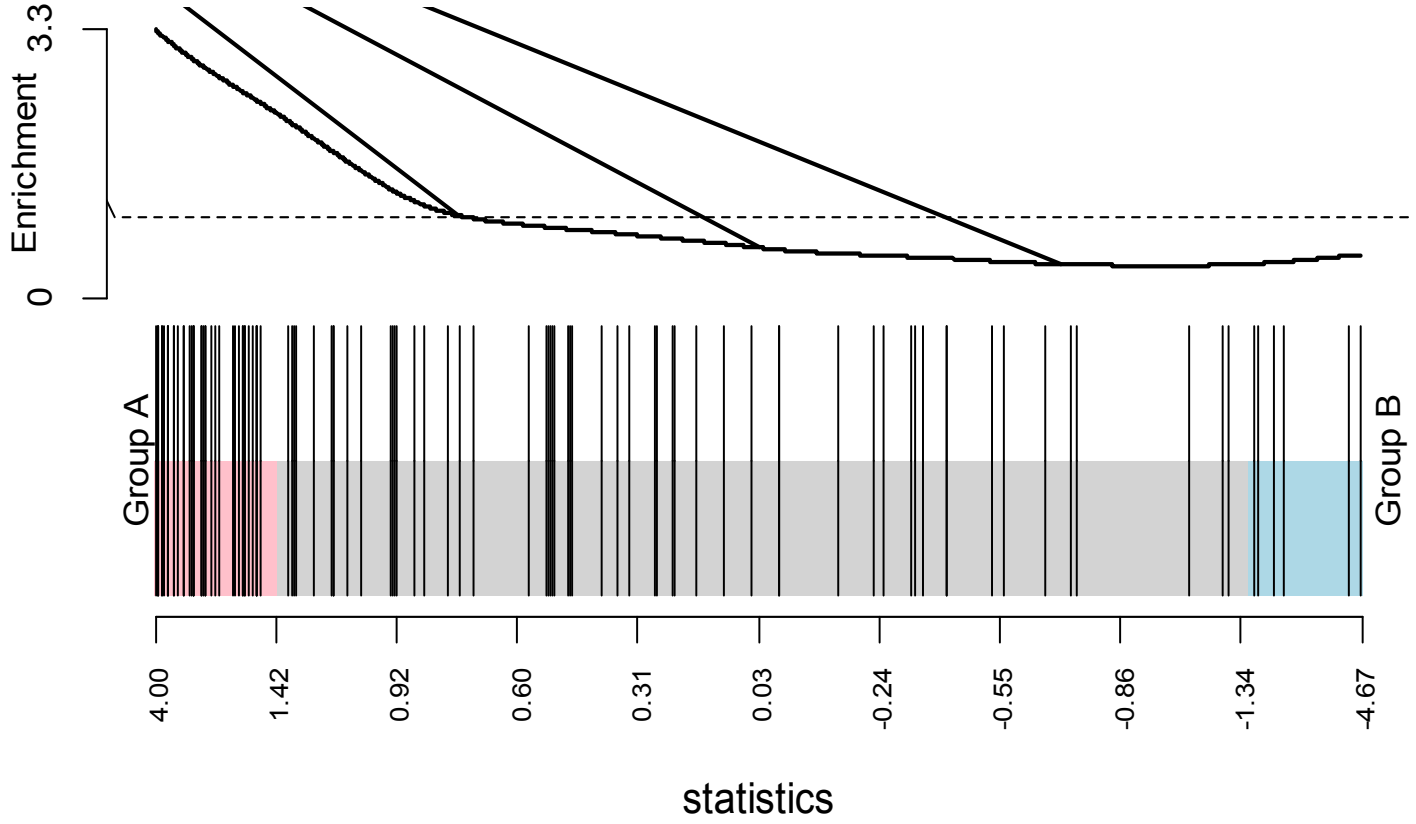




Genome-wide barcode plot

Slide courtesy of Gordon Smyth

Visualisation: Barcodeplot + enrichment worm



Slide courtesy of Gordon Smyth

ROAST gene set test

- The question asked is "Do the genes in this gene set tend to be differentially expressed?"
- It is NOT compared relative to other genes
- It is designed such that if > 25-50% of genes in the gene set are differentially expressed it will be significant
- It uses sophisticated techniques (rotation) to preserve gene-gene dependence in the data.

Summary

- Gene set testing techniques range from simple (overlap analysis) to quite complex (CAMERA and ROAST)
- Which test you choose depends on what your hypothesis is
- Sometimes we just do them all…

Acknowledgements

- Gordon Smyth