IPRG 2015

(PROTEOME INFORMATICS RESEARCH GROUP)

DIFFERENTIAL ABUNDANCE
IN
LABEL-FREE PROTEOMICS

Olga Vitek

Northeastern University

IPRG 2015

iPRG committee

Henry Lam - Hong Kong University of Science and Technology (Co-chair)
Eugene Kapp - Walter and Eliza Hall Institute of Medical Research (Co-chair)
Brett Phinney - University of California at Davis (ABRF EB Liaison)
John Cottrell - Matrix Science Ltd
Michael Hoopmann - Institute for Systems Biology
Sangtae Kim - Pacific Northwest National Laboratory
Thomas Neubert - New York University School of Medicine
Magnus Palmblad - Leiden University Medical Center
Olga Vitek - Northeastern University
Sue Weintraub - University of Texas Health Science Center at San Antonio

Special thanks to

Jingjing Deng - New York University School of Medicine (generating data)
Justin Locke, University of California at Davis (anonymizer)

STUDY GOALS

- Evaluate the performance of data analysis approaches for label-free quantitative proteomics

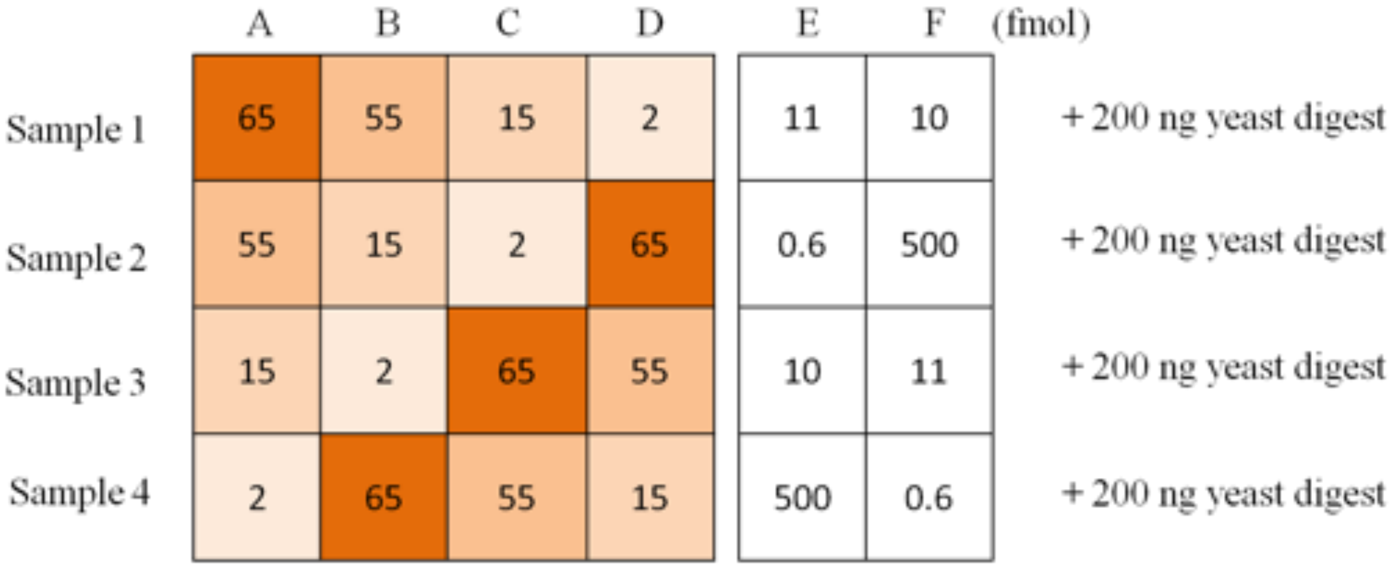
- Provide a well-designed dataset for assessing label-free quantitative proteomics software tools.

- Raise the awareness of the importance of statistical methods and provide an educational opportunity.

OUTLINE

- Study design
    - Experimental procedures
    - Data analysis
- Summary of the submissions
    - Participants
    - Methods
- Summary of the results
    - Comparative performance
    - Method characteristics

STUDY DESIGN



	A	B	C	D	E	F	(fmol)
Sample 1	65	55	15	2	11	10	+ 200 ng yeast digest
Sample 2	55	15	2	65	0.6	500	+ 200 ng yeast digest
Sample 3	15	2	65	55	10	11	+ 200 ng yeast digest
Sample 4	2	65	55	15	500	0.6	+ 200 ng yeast digest

	Name	Origin	MW
A	Ovalbumin	Chicken Egg White	45KD
B	Myoglobin	Equine Heart	17KD
C	Phosphorylase b	Rabbit Muscle	97KD
D	Beta-Galactosidase	Escherichia Coli	116KD
E	Bovine Serum Albumin	Bovine Serum	66KD
F	Carbonic Anhydrase	Bovine Erythrocytes	29KD

- Whole yeast cell lysate
- 6 spiked proteins
- Shotgun proteomics sample prep
- Randomized order
- 2-hour runs on Thermo Q-Exactive
  - DDA
  - HCD fragmentation
  - High mass accuracy for MS1 & MS2
  - 3 replicate runs

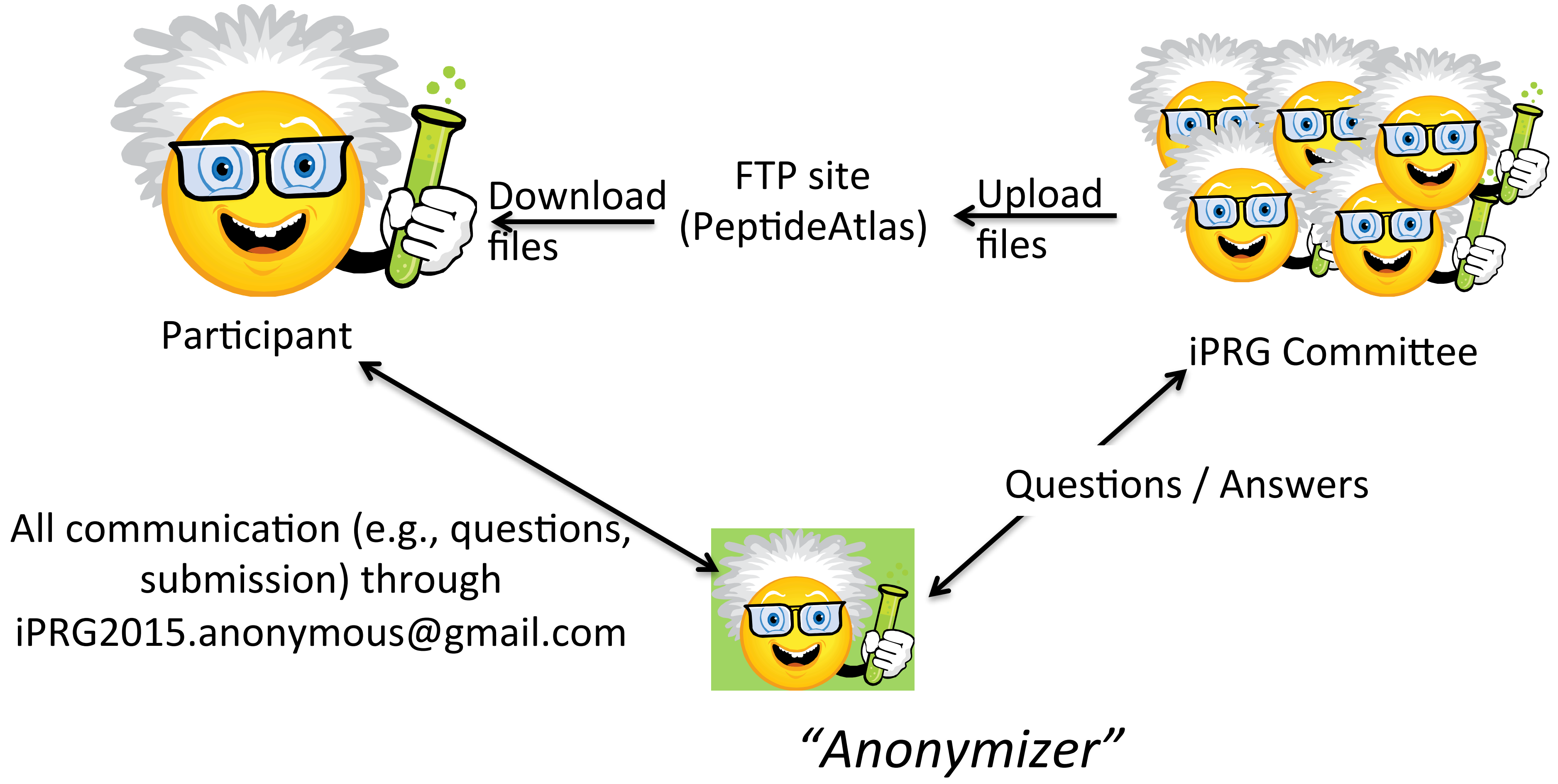
STUDY DESIGN

- Questions of interest
  - Estimate (log-) fold changes
  - Detect differentially abundant proteins
  - Characterize the associated uncertainty
- Approaches
  - Intensity or spectral counts
  - Any computational/statistical method of choice
- Input data
  - Raw, and/or peak ids by iPRG, and/or peak intensities by iPRG

IPRG DATA PROCESSING

- Search database
  - FASTA database containing UniProt yeast proteins
  - 6 spiked proteins disguised as yeast proteins
  - 1:1 appended decoys (shuffling aa between tryptic sites)
- Identification of MS/MS spectra
  - .RAW, .mzML (converted by msconvert - ProteoWizard)
  - Combining 3 search engines (Comet, MSGF+, OMSSA), validated by PeptideProphet/iProphet
- Quantification of MS1 peaks
  - Peptides mapping to multiple proteins are removed
  - Extracted ion intensities extracted using Skyline (default parameters, report all isotopes separately) based on search ids

IPRG PARTICIPATION



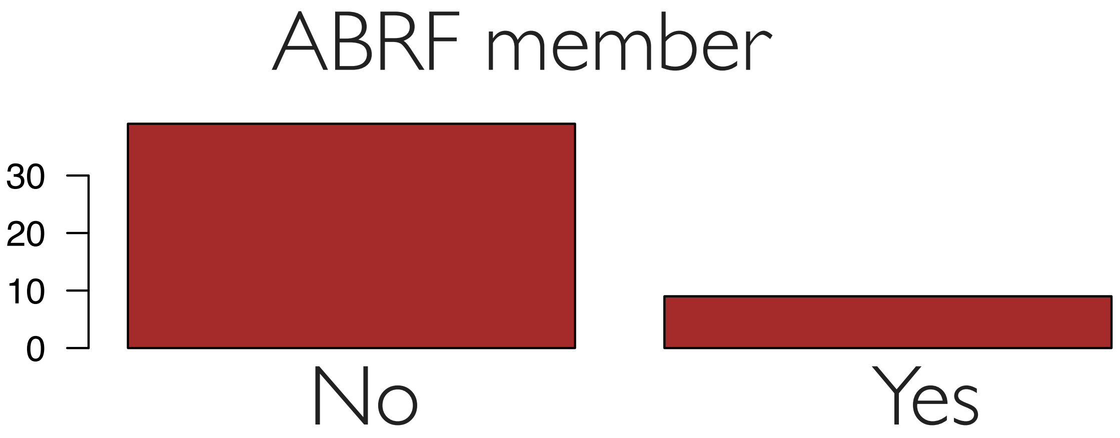
- Recruitment
  - advertised on mailing lists, ASMS 2014, direct invitations
- Required submission items
  - results template
  - online survey (survey monkey)

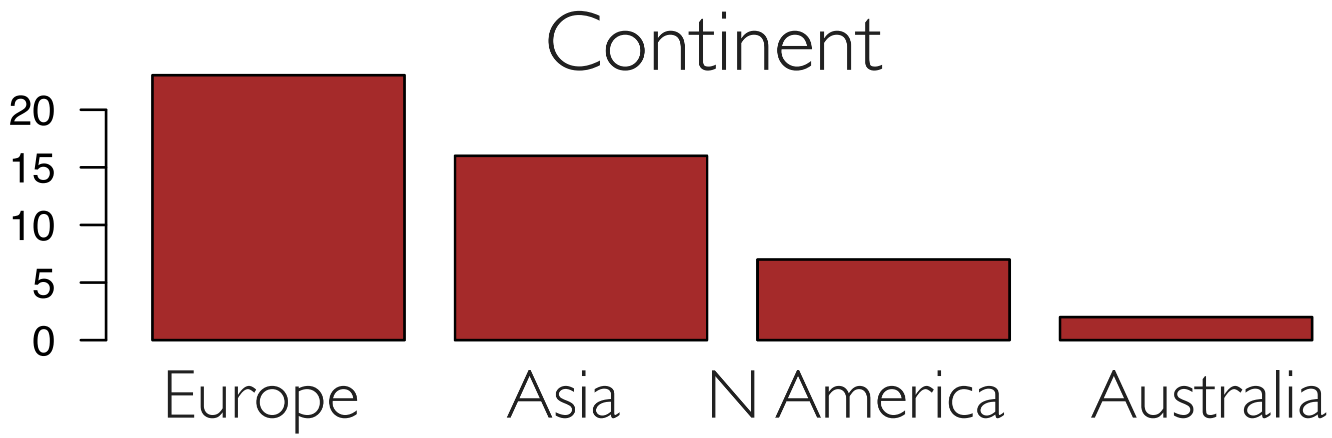
OUTLINE

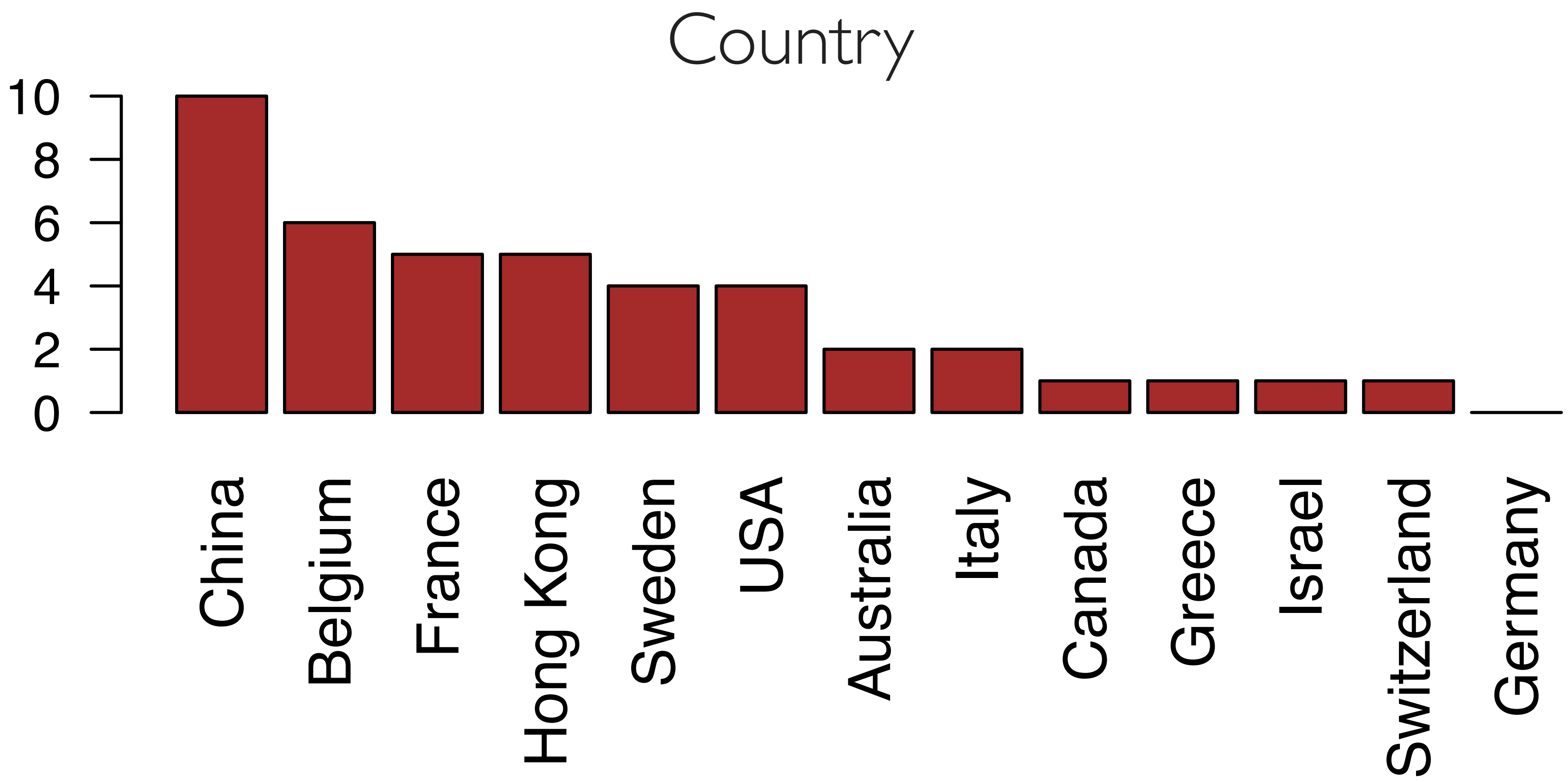
- Study design
  - Experimental procedures
  - Data analysis
- Summary of the submissions
  - Participants
  - Methods
- Summary of the results
  - Comparative performance
  - Method characteristics

51 SUBMISSIONS

49 readable submissions, 45 readable surveys

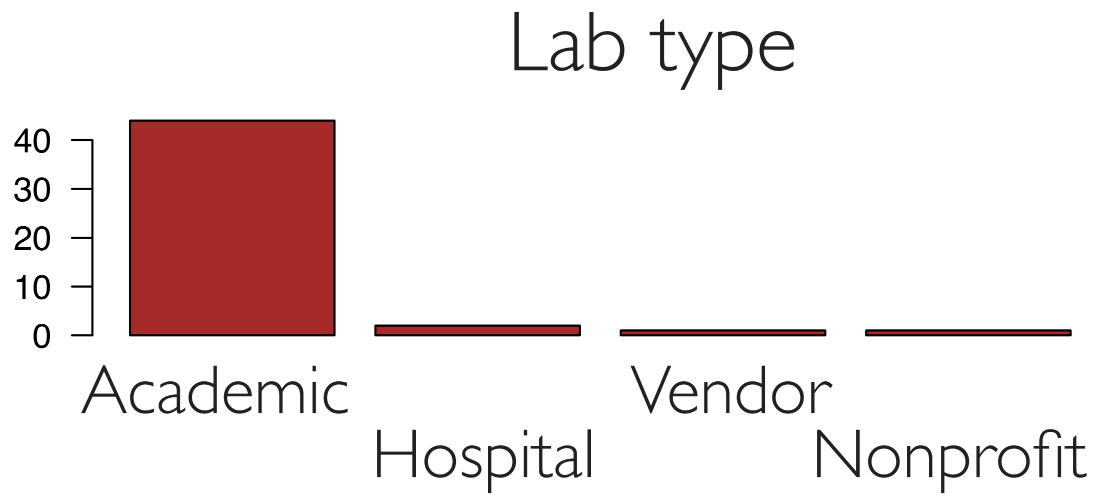

ABRF member

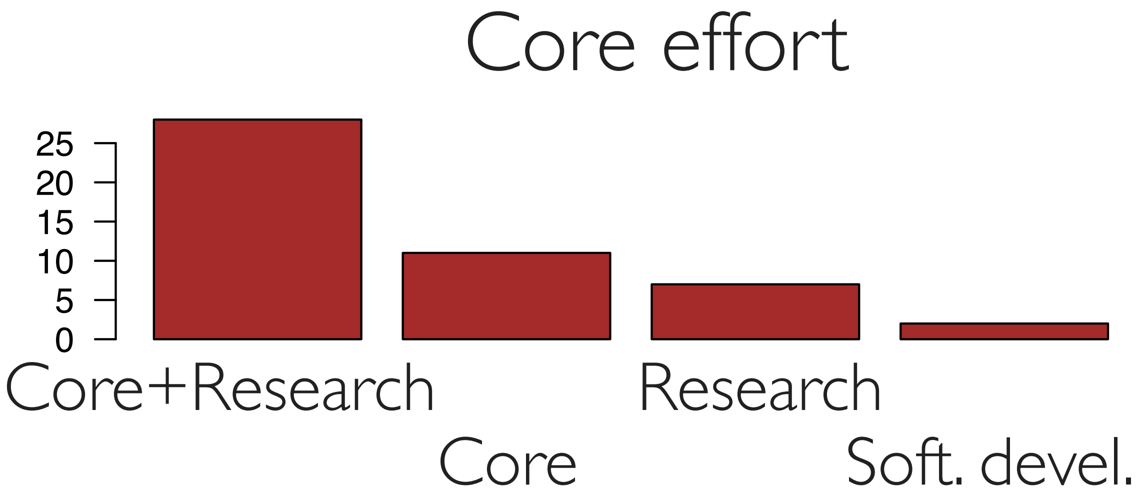

Continent

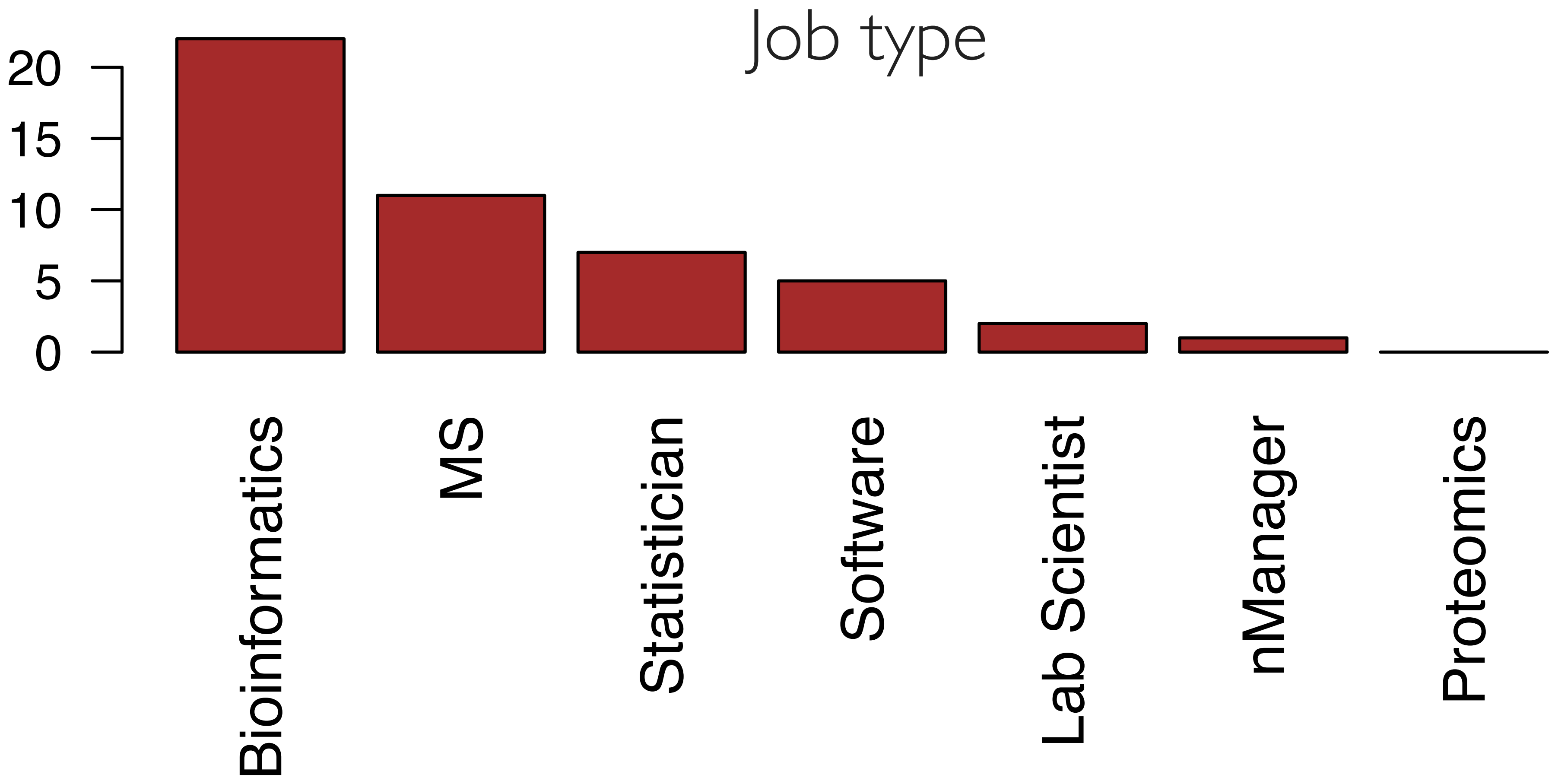

Country

51 SUBMISSIONS

49 readable submissions, 45 readable surveys



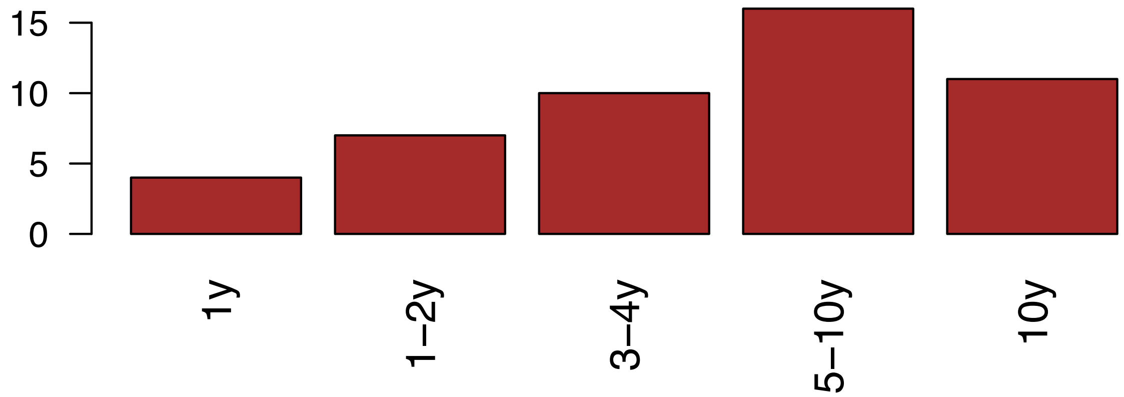




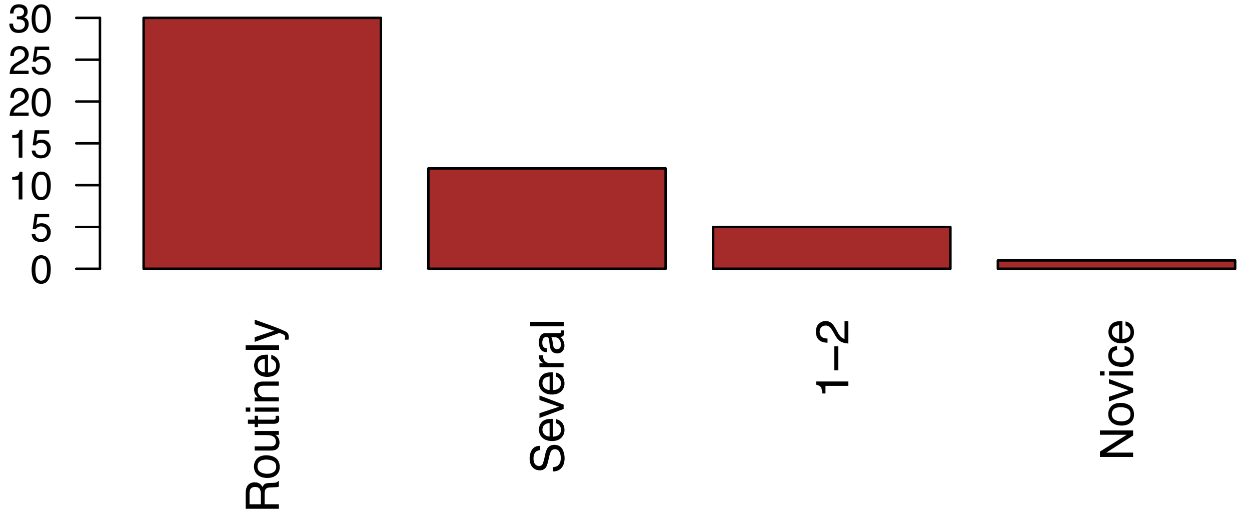
51 SUBMISSIONS

49 readable submissions, 45 readable surveys

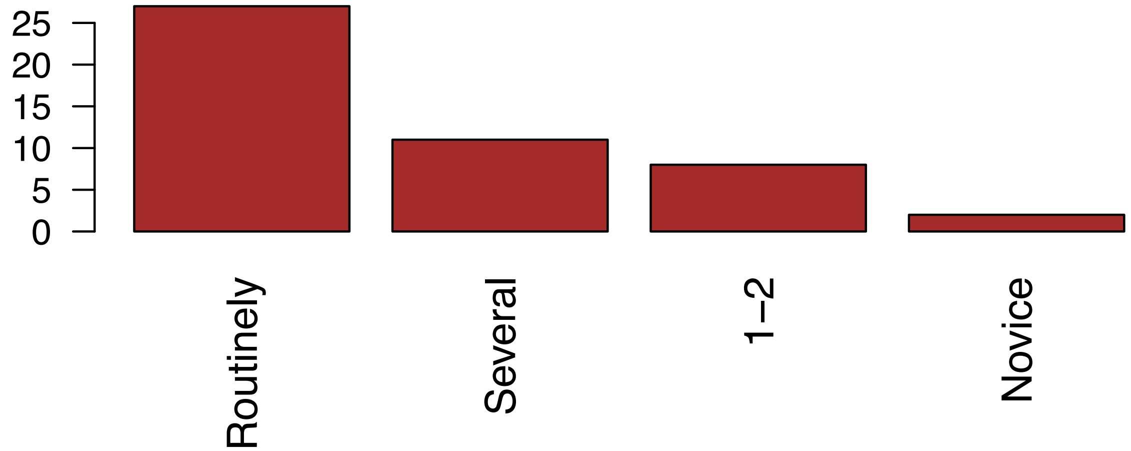
Years of experience in proteomics



Analysis of MS experiments

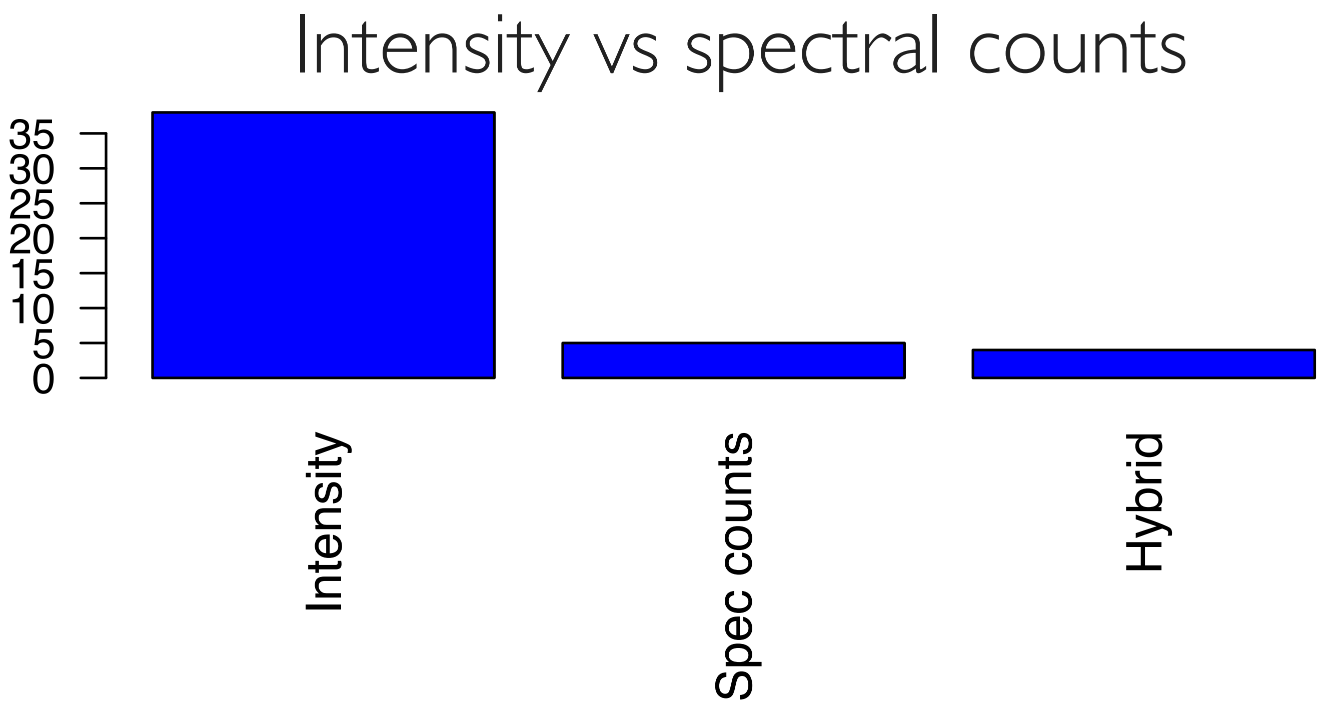


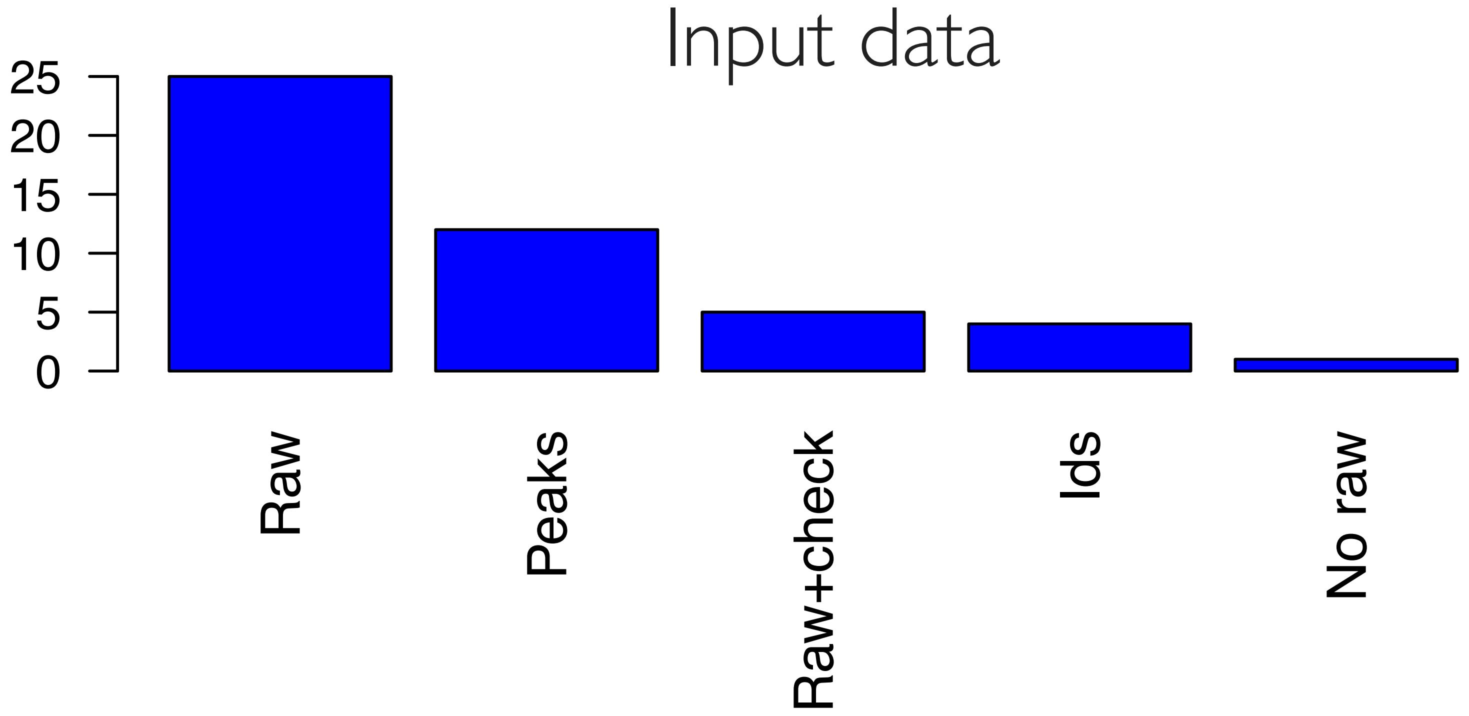
Analysis of quantitative label-free experiments



51 SUBMISSIONS

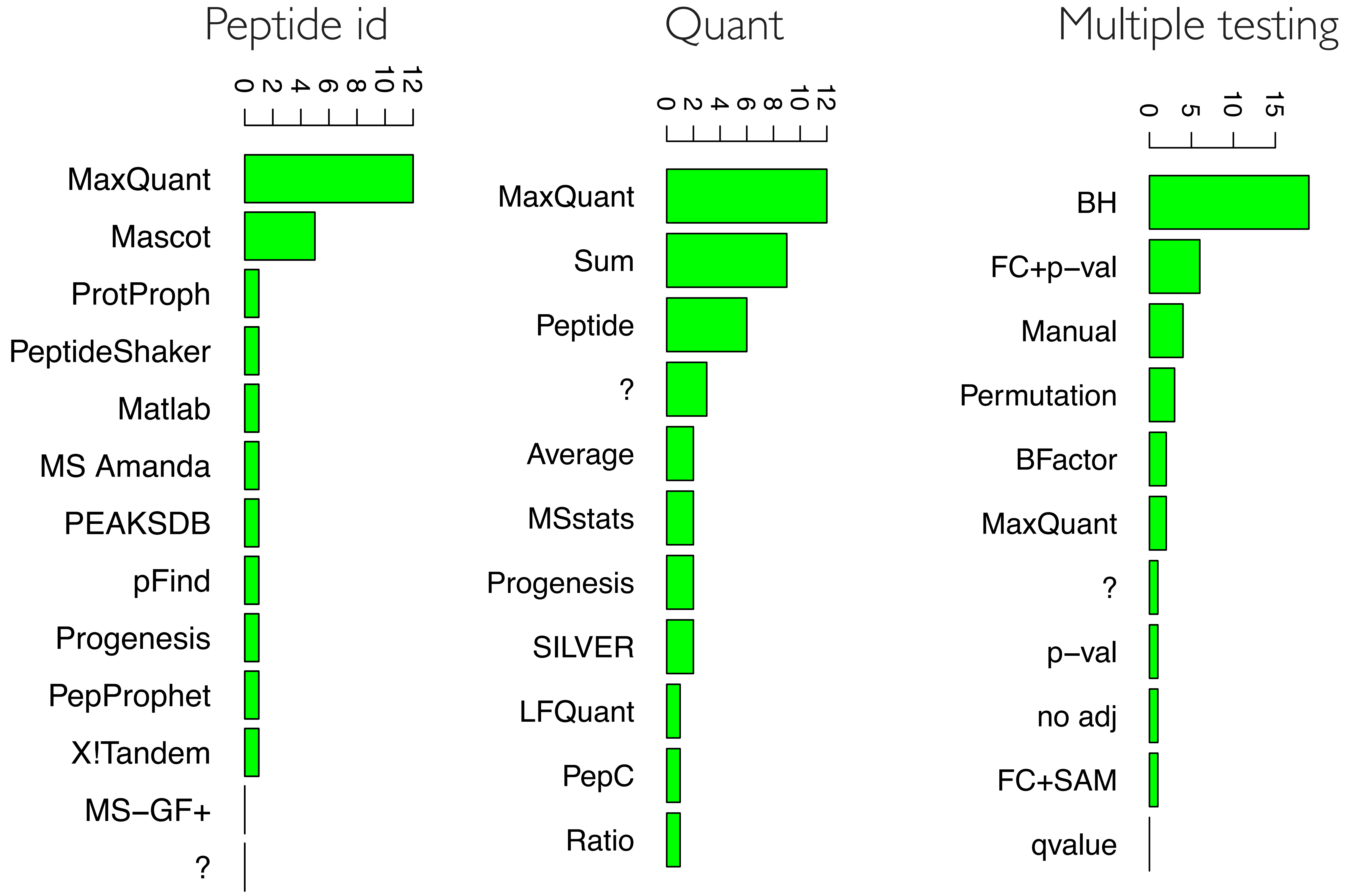
49 readable submissions, 45 readable surveys





51 SUBMISSIONS

49 readable submissions, 45 readable surveys

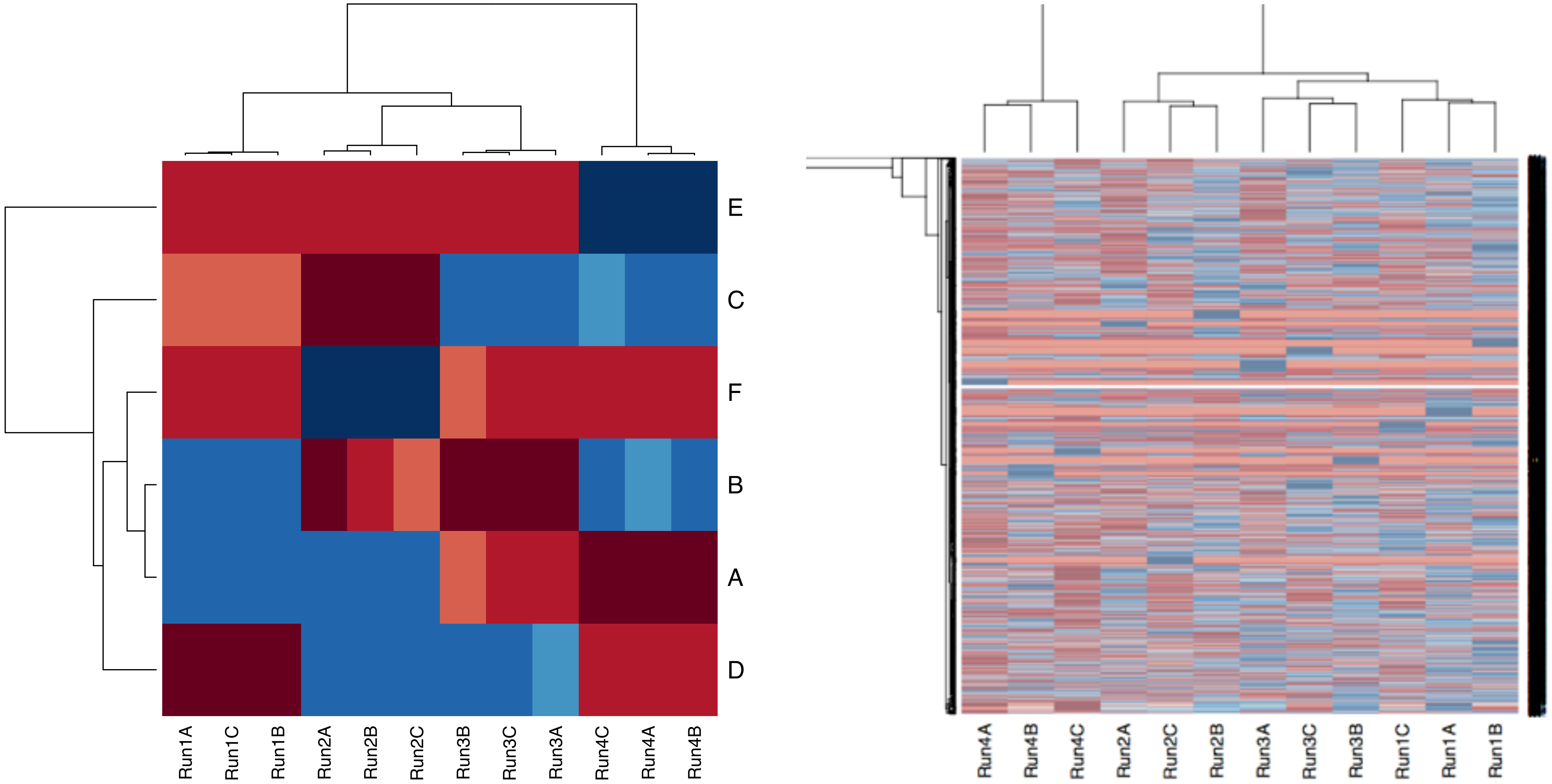


OUTLINE

- Study design
  - Experimental procedures
  - Data analysis
- Summary of the submissions
  - Participants
  - Methods
- Summary of the results
  - Comparative performance
  - Method characteristics

FIRST LOOK: IPRG ANALYSIS

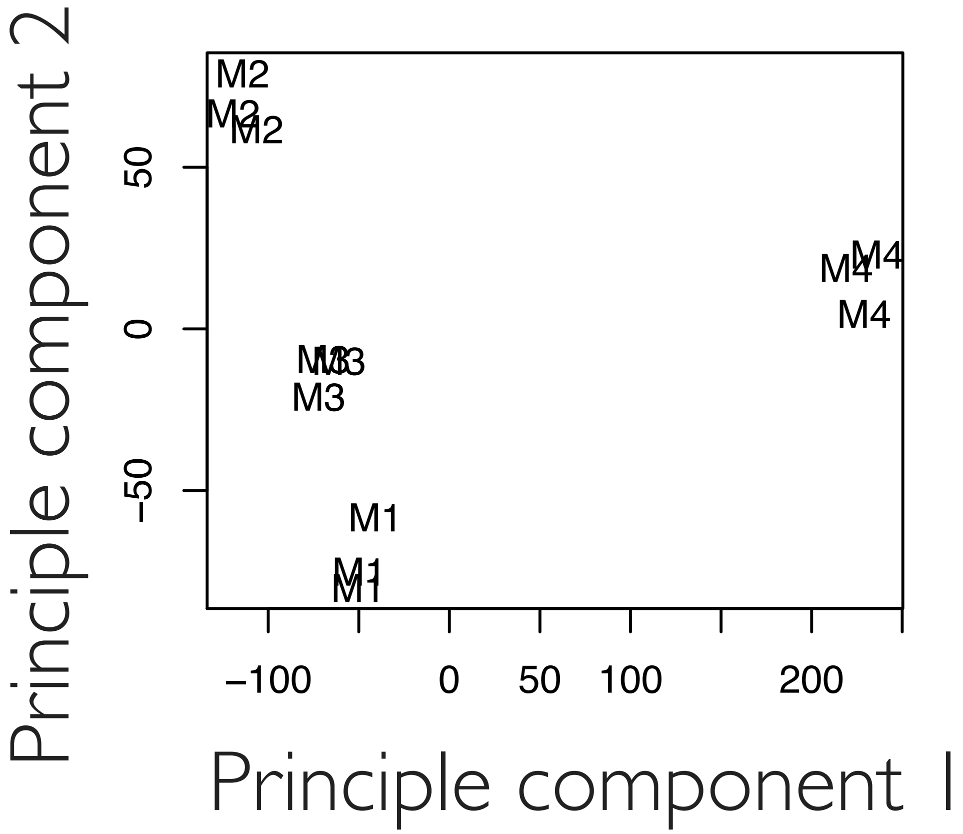
Spectral counts

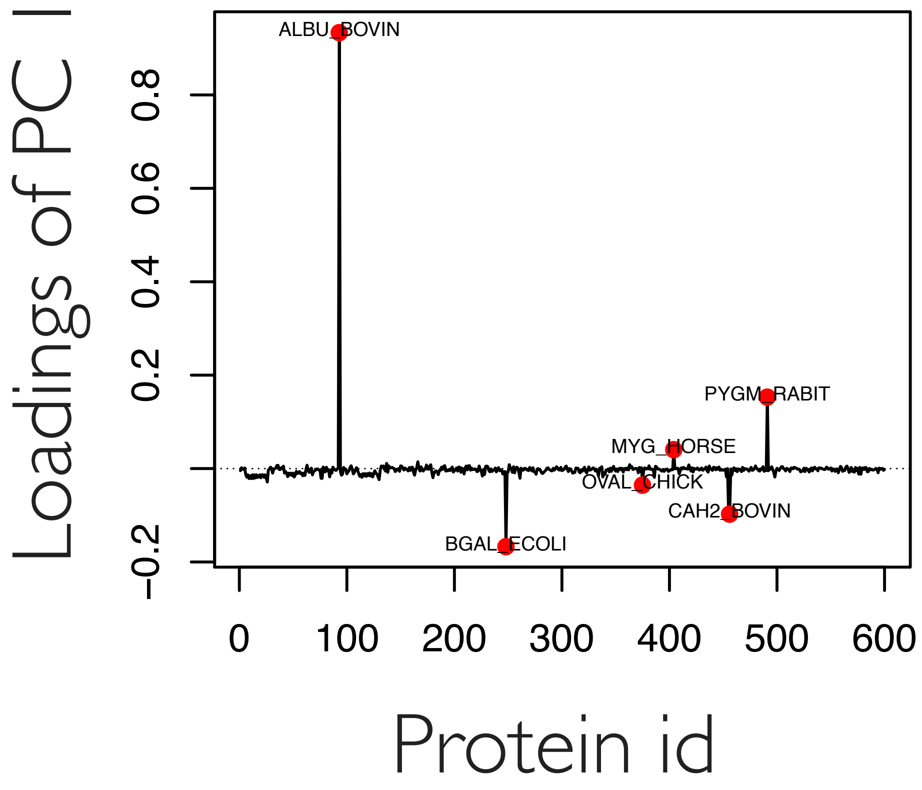


The patterns of changes are systematic and easy to find
(Latin squares are not really designed for blinded studies!)

FIRST LOOK: IPRG ANALYSIS

Spectral counts

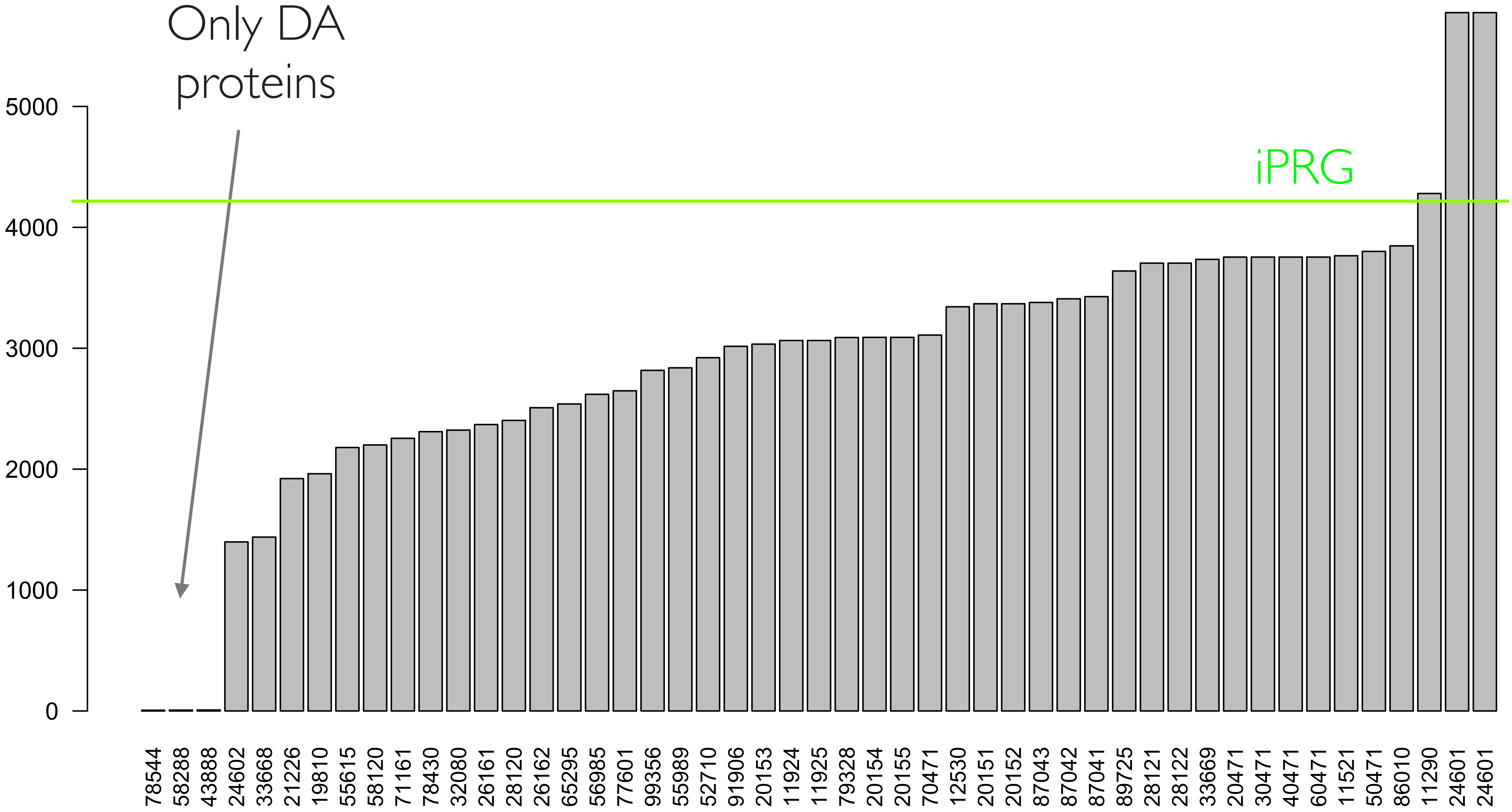




The patterns of changes are systematic and easy to find

Here we are interested in LR estimation
and testing and associated uncertainty

NUMBER OF REPORTED PROTEINS

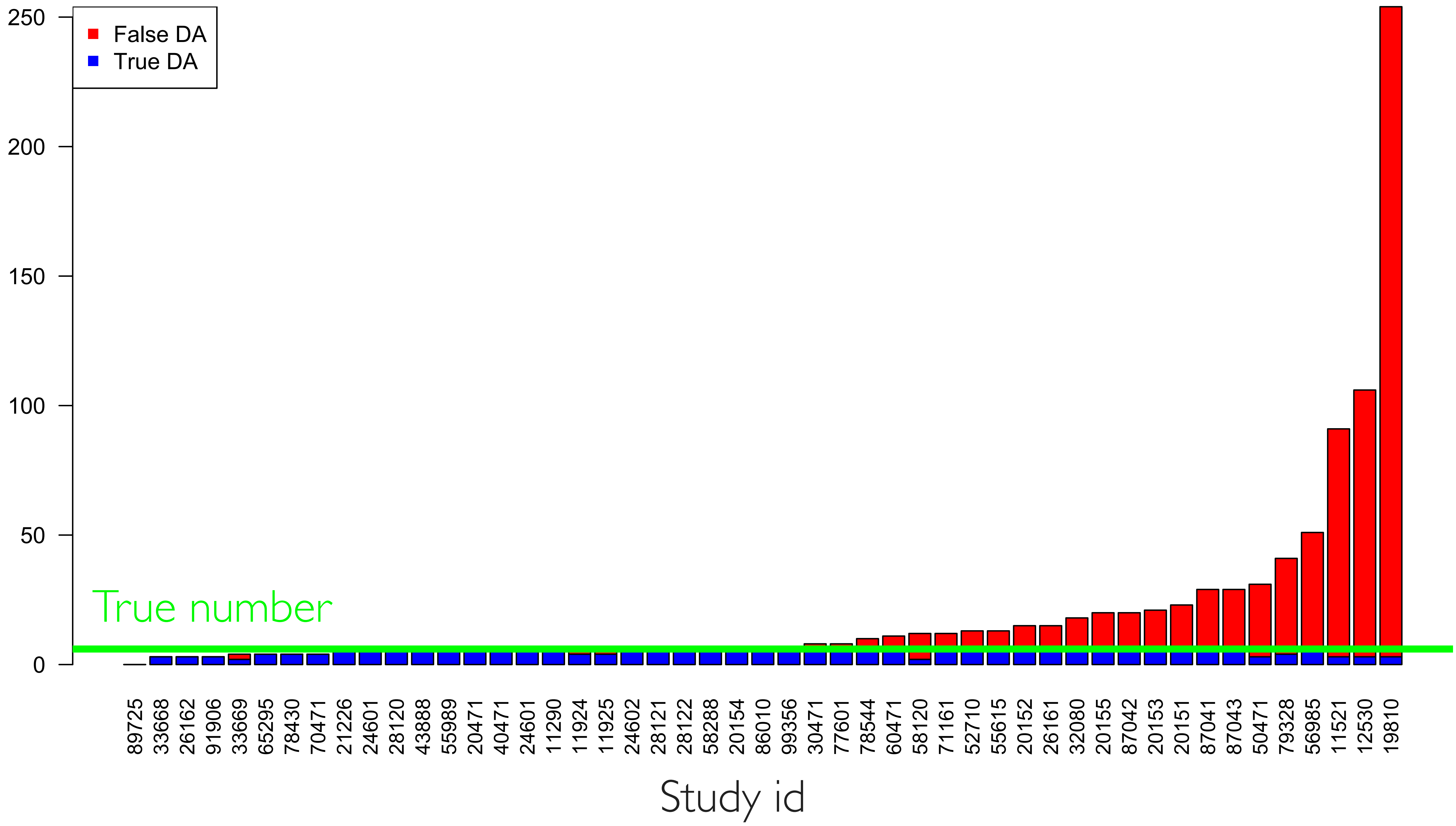


Only DA proteins

iPRG

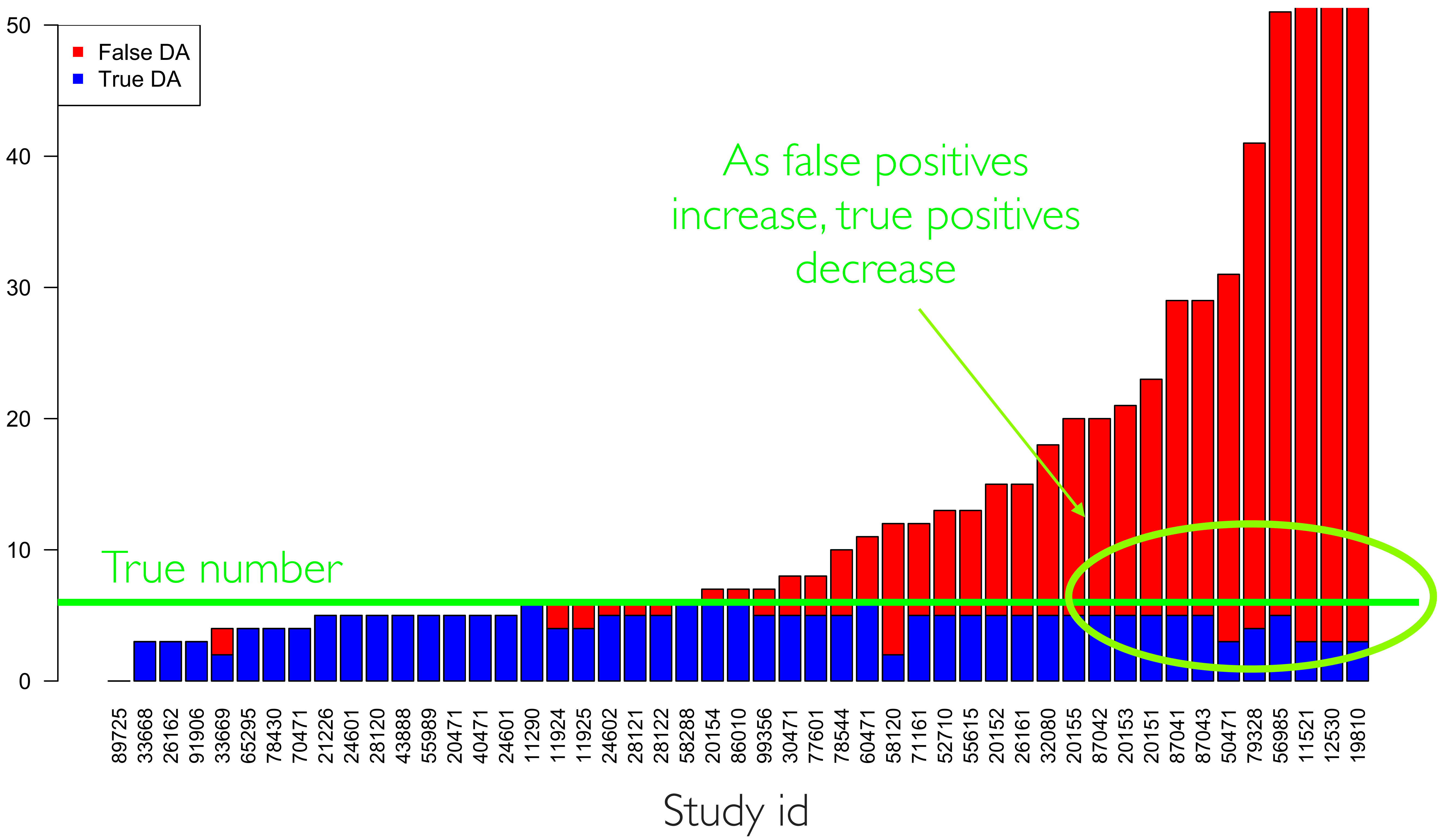
DIFFERENTIALLY ABUNDANT PROTEINS

Sample 1 vs sample 2



DIFFERENTIALLY ABUNDANT PROTEINS

Sample 1 vs sample 2

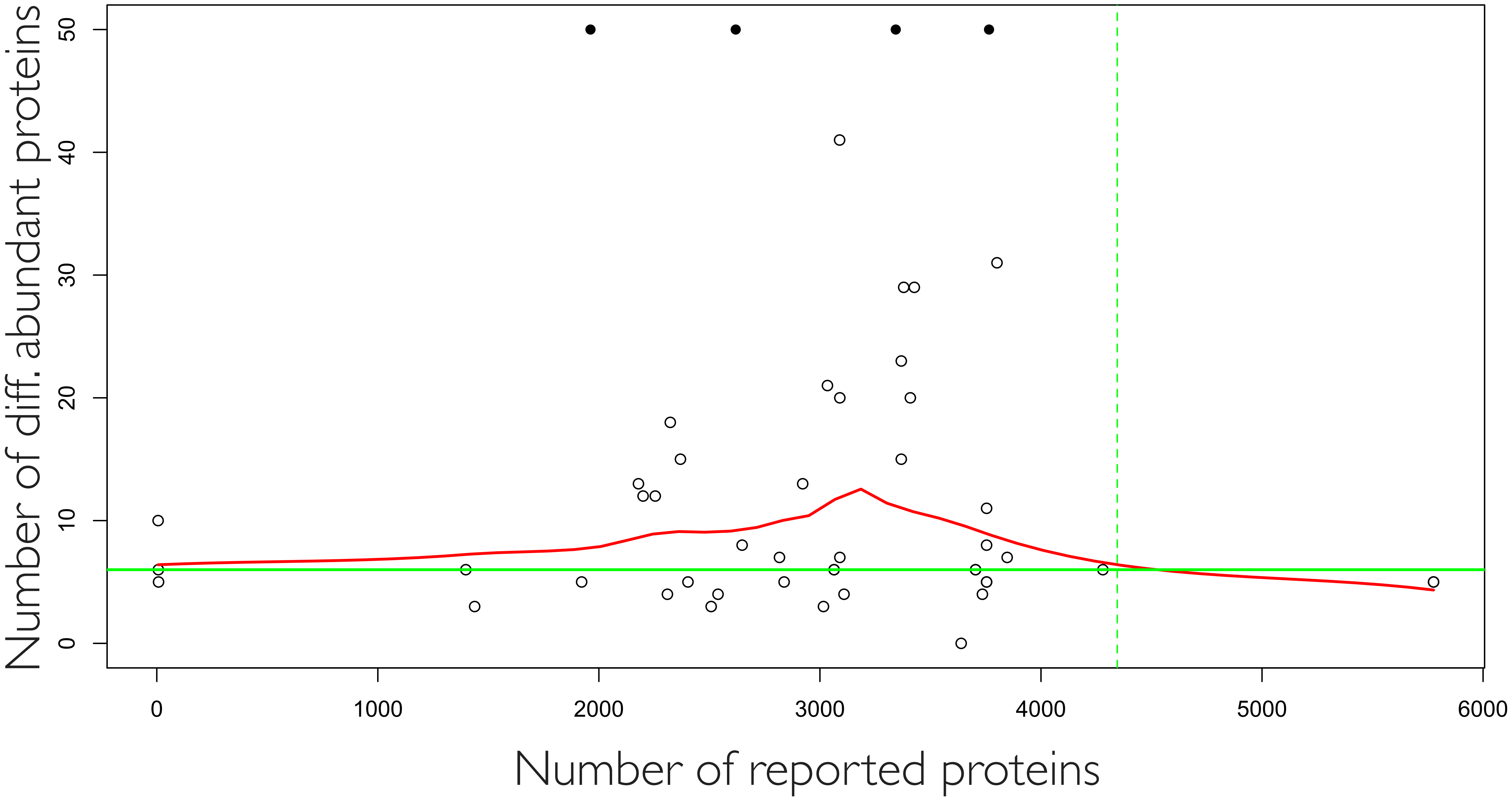


As false positives increase, true positives decrease

True number

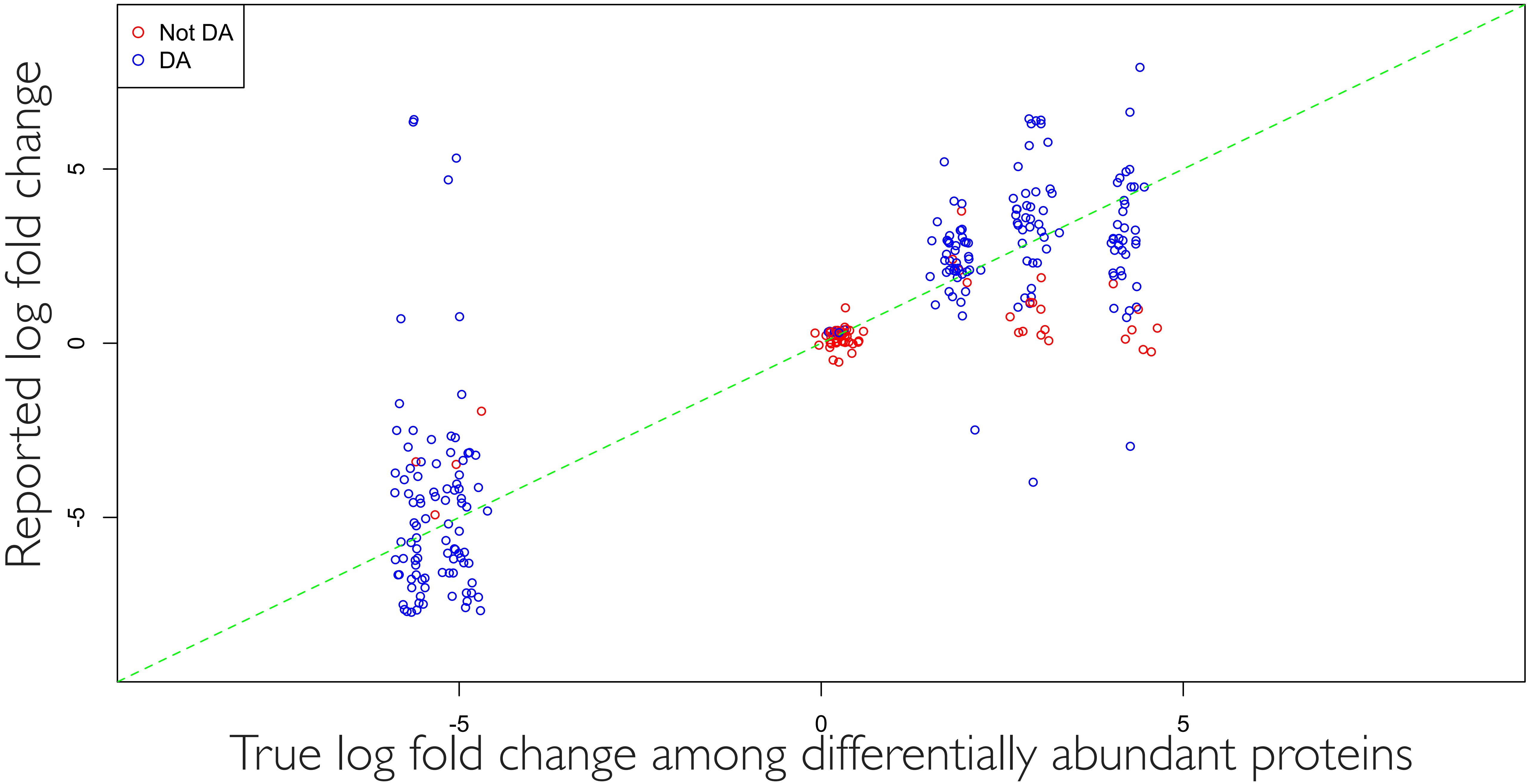
REPORTING MORE PROTEINS DID NOT HELP

Sample 1 vs sample 2



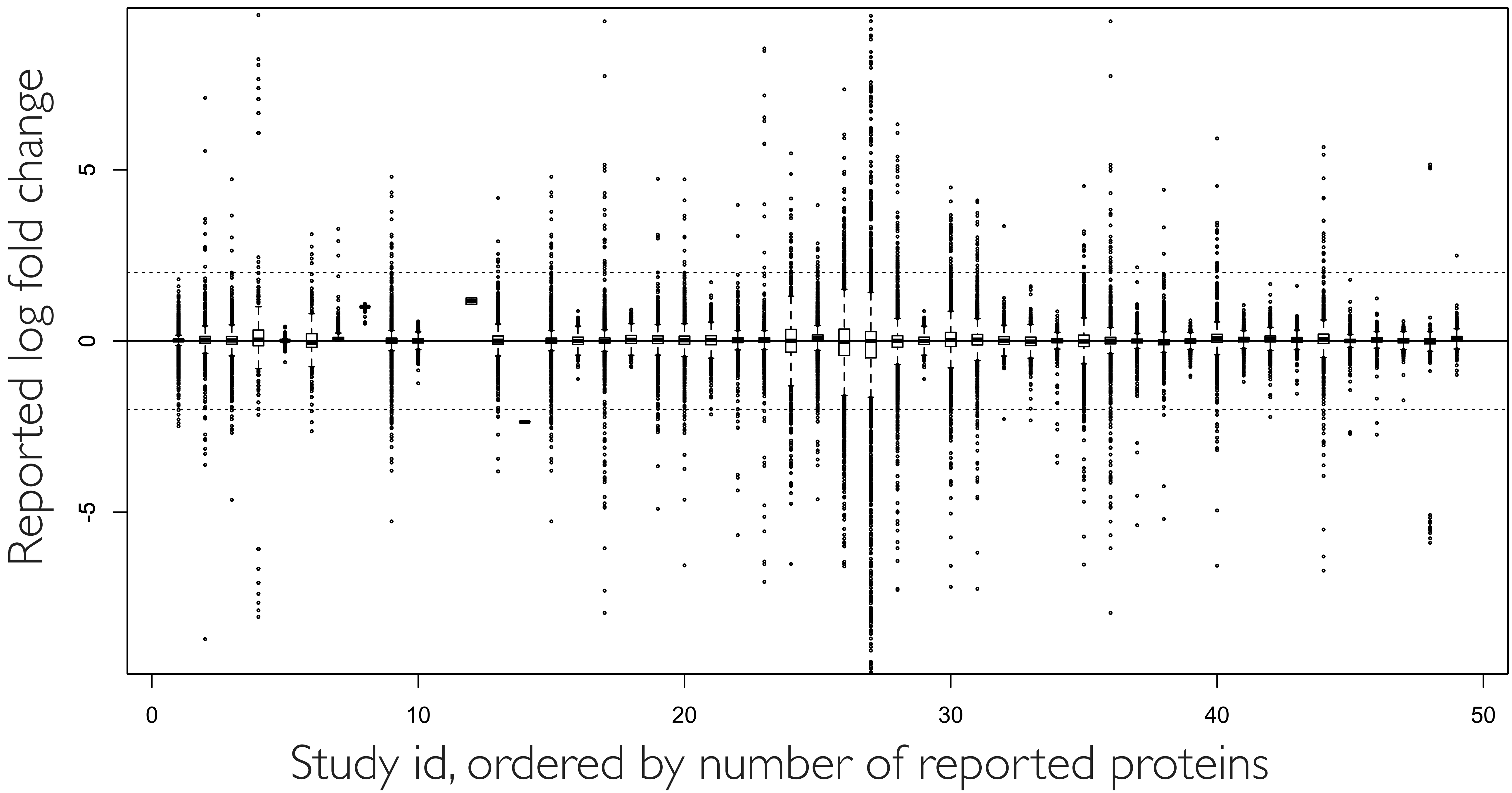
LOG-FC AMONG SPIKED PROTEINS

Sample 1 vs sample 2



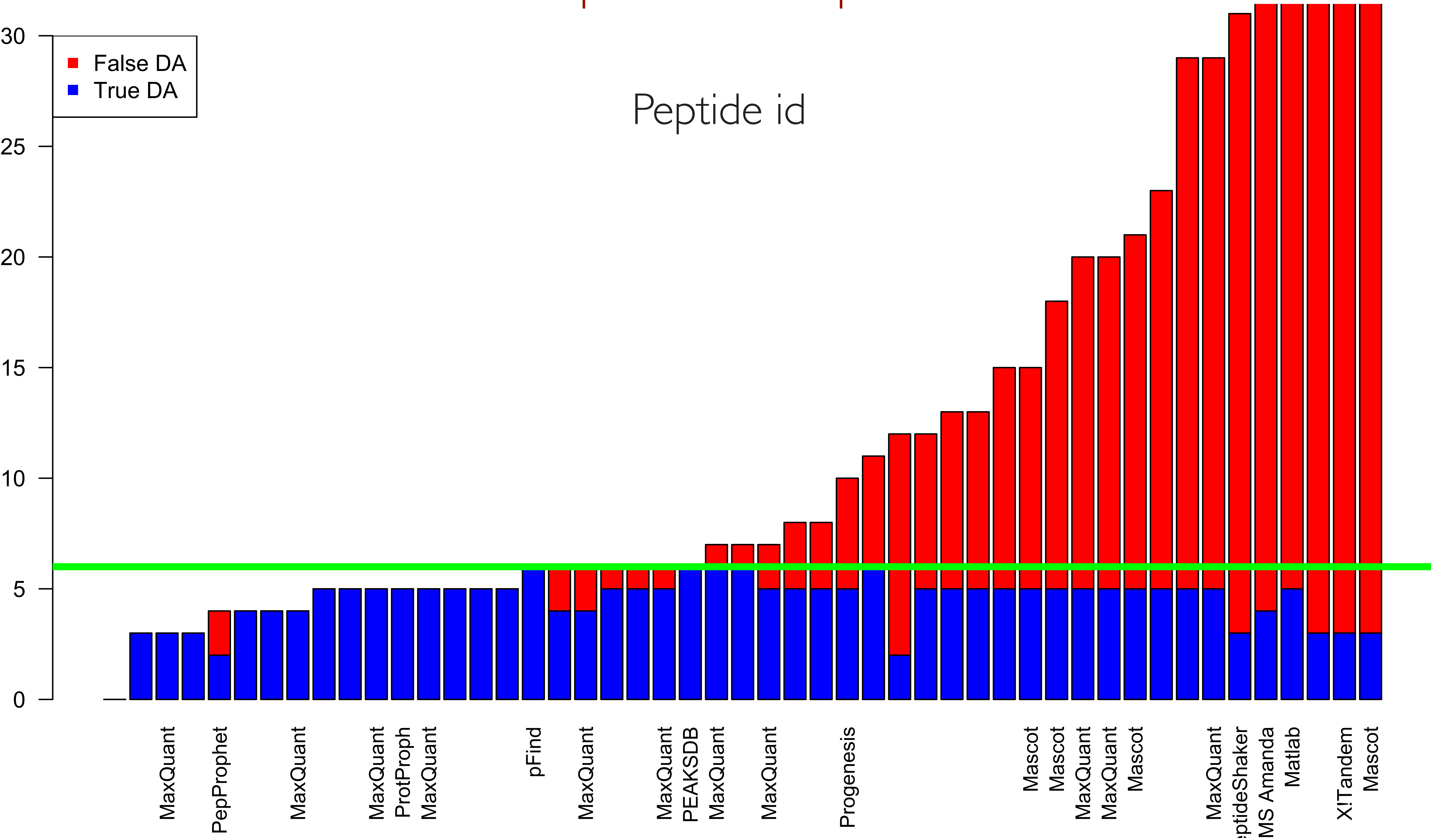
LOG-FC AMONG BACKGROUND PROTEINS

Sample 1 vs sample 2



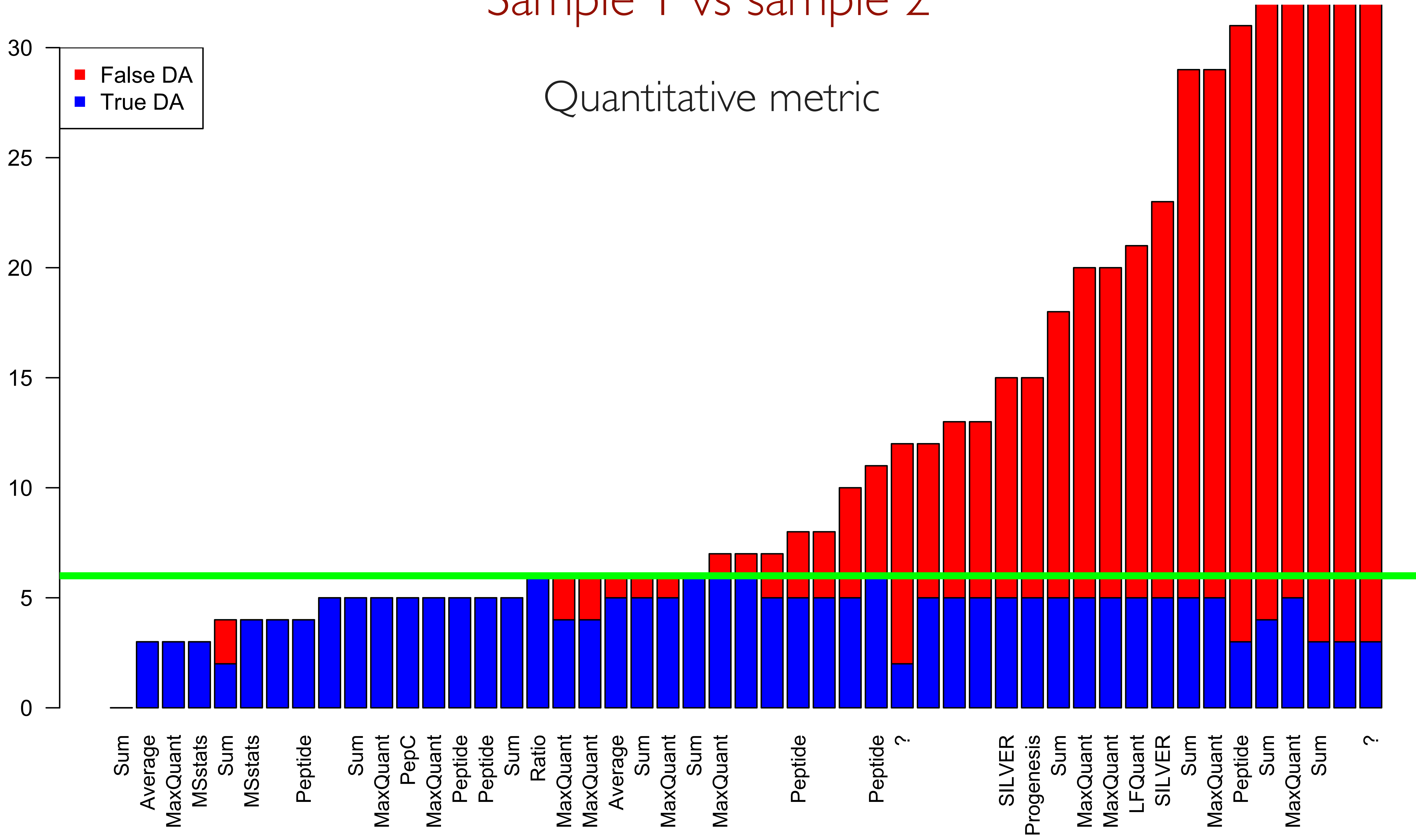
SAME TOOL CAN PRODUCE DIFFERENT RESULTS

Sample 1 vs sample 2



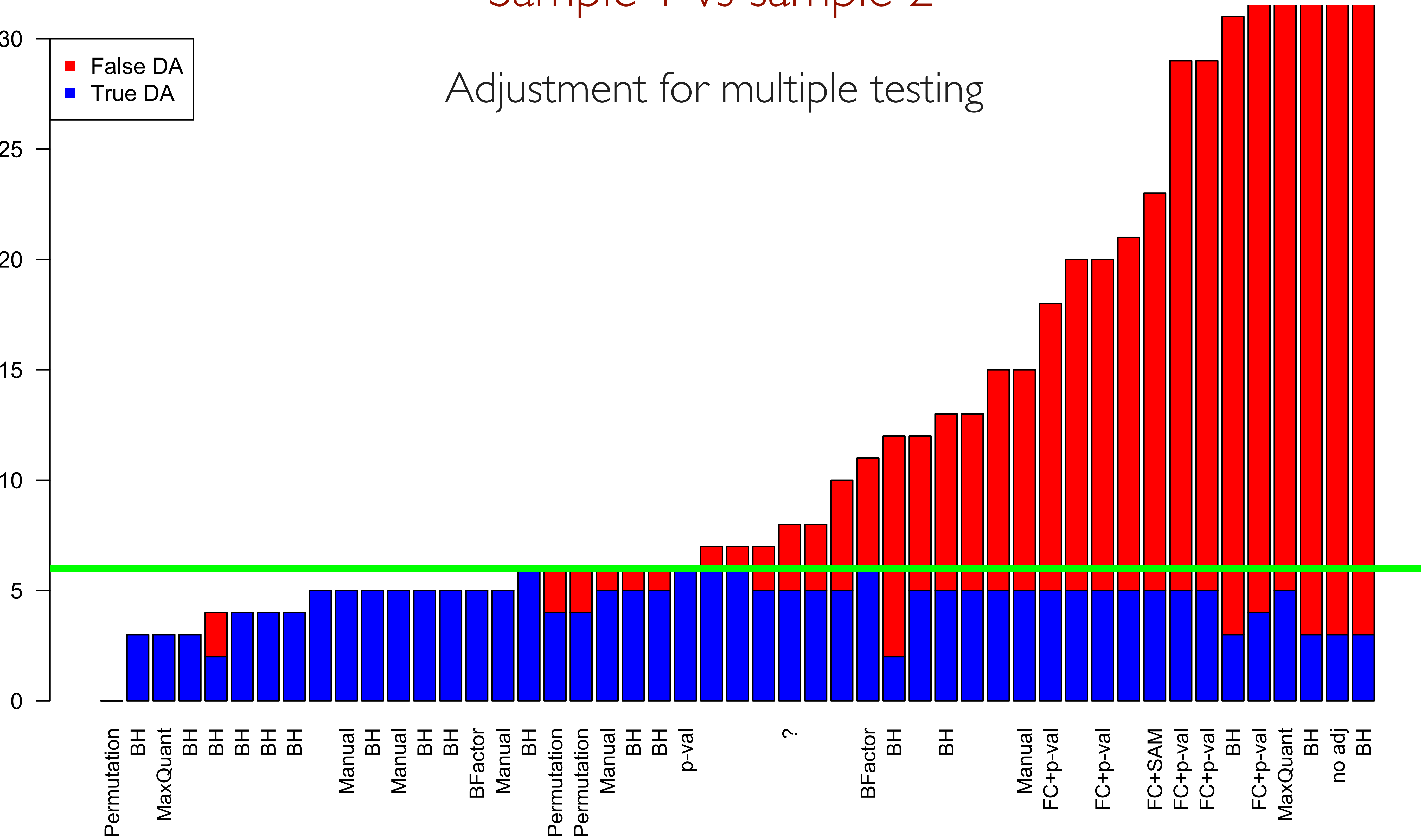
SAME TOOL CAN PRODUCE DIFFERENT RESULTS

Sample 1 vs sample 2



SAME TOOL CAN PRODUCE DIFFERENT RESULTS

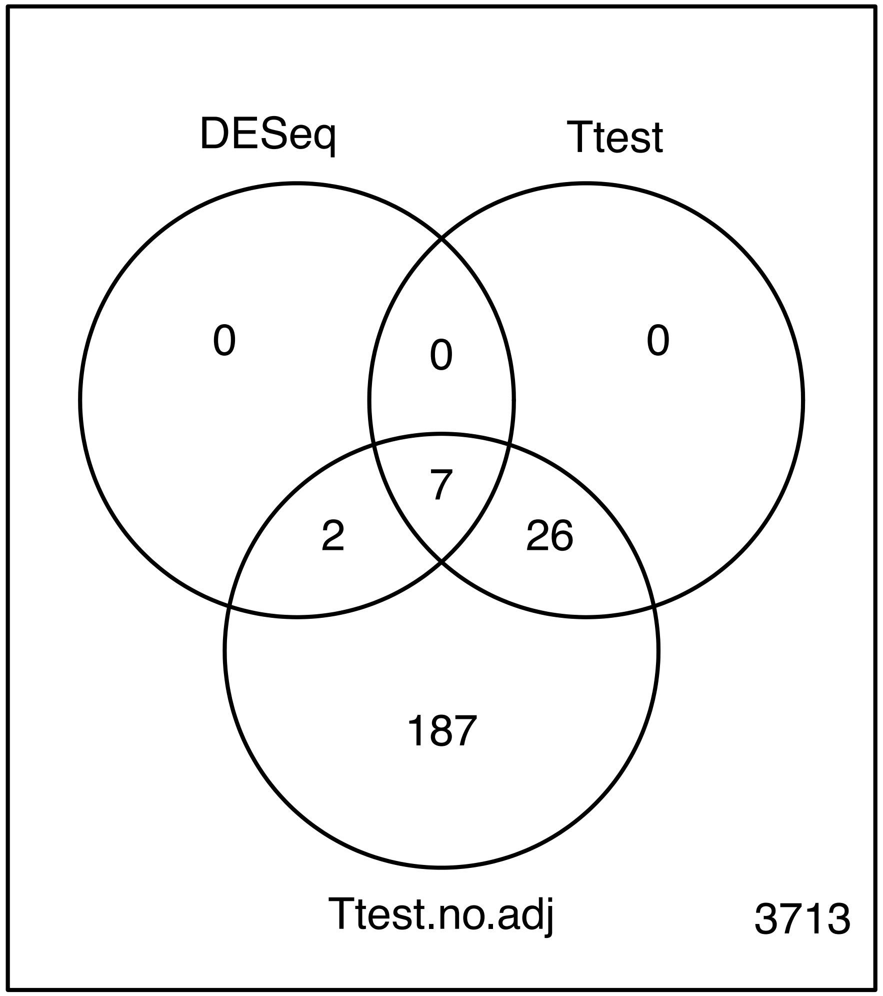
Sample 1 vs sample 2


Adjustment for multiple testing

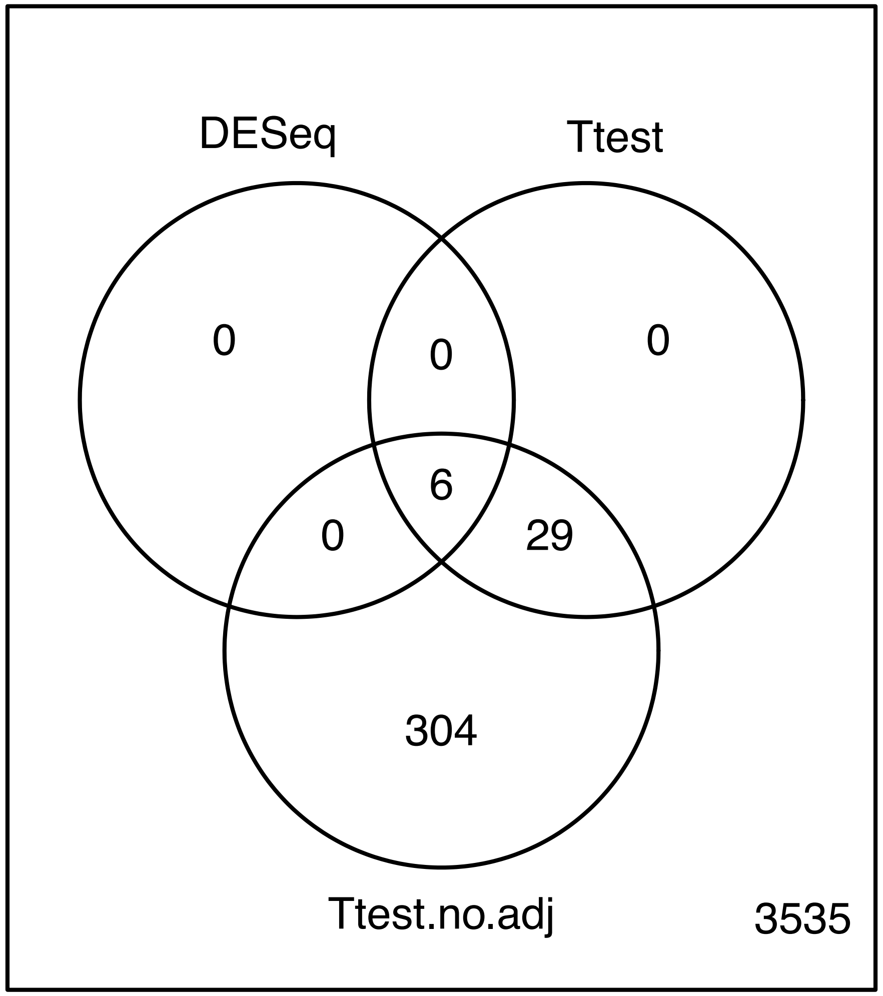
EXAMPLE: SPECTRAL COUNTS

iPRG analysis

Sample 1 vs sample 2



Sample 4 vs sample 1



T-test is not appropriate for low counts.

Lack of adjustment for multiple testing leads to a huge number of false positives

PRELIMINARY CONCLUSIONS

- A wide variety of approaches can be used
    - A same tool can produce good or bad results
- Some patterns emerge
    - Need more work to ensure specificity
    - Reporting more proteins is associated with more false positives, unless carefully done
    - Adjustment for multiple testing is importantSome patterns emerge
- More analyses of these submissions shortly