

# STATISTICS FOR (LABEL-FREE) PROTEOMICS

Olga Vitek

College of Science  
College of Computer and Information Science



Northeastern University

# WHY STATISTICS?

- Variation and uncertainty are unavoidable
  - *Technical variation*: sampling handling, storage, processing
  - *Instrumental variation*: elution time, ion suppression
  - *Signal processing*: peak boundaries, identity, intensity
  - *Biological variation*: variation in protein abundance
- Overall goal: effective, reproducible research
  - *Experimental design*: unbiased and efficient experiments
  - *Data analysis*: objective conclusions in presence of uncertainty
  - *Statistical tools*: re-analysis, peer review, reproducibility

*"Statistics: a body of methods for making wise decisions in the face of uncertainty." (W. A. Wallis)*

# WHY STATISTICS?



[nature.com](#) ▶ [journal home](#) ▶ [archive](#) ▶ [issue](#) ▶ [opinion and comment](#) ▶ [correspondence](#) ▶ [full text](#)

NATURE BIOTECHNOLOGY | OPINION AND COMMENT | CORRESPONDENCE



## Sequencing technology does not eliminate biological variability

Kasper D Hansen, Zhijin Wu, Rafael A Irizarry & Jeffrey T Leek

[Affiliations](#) | [Corresponding authors](#)

*Nature Biotechnology* 29, 572–573 (2011) | doi:10.1038/nbt.1910

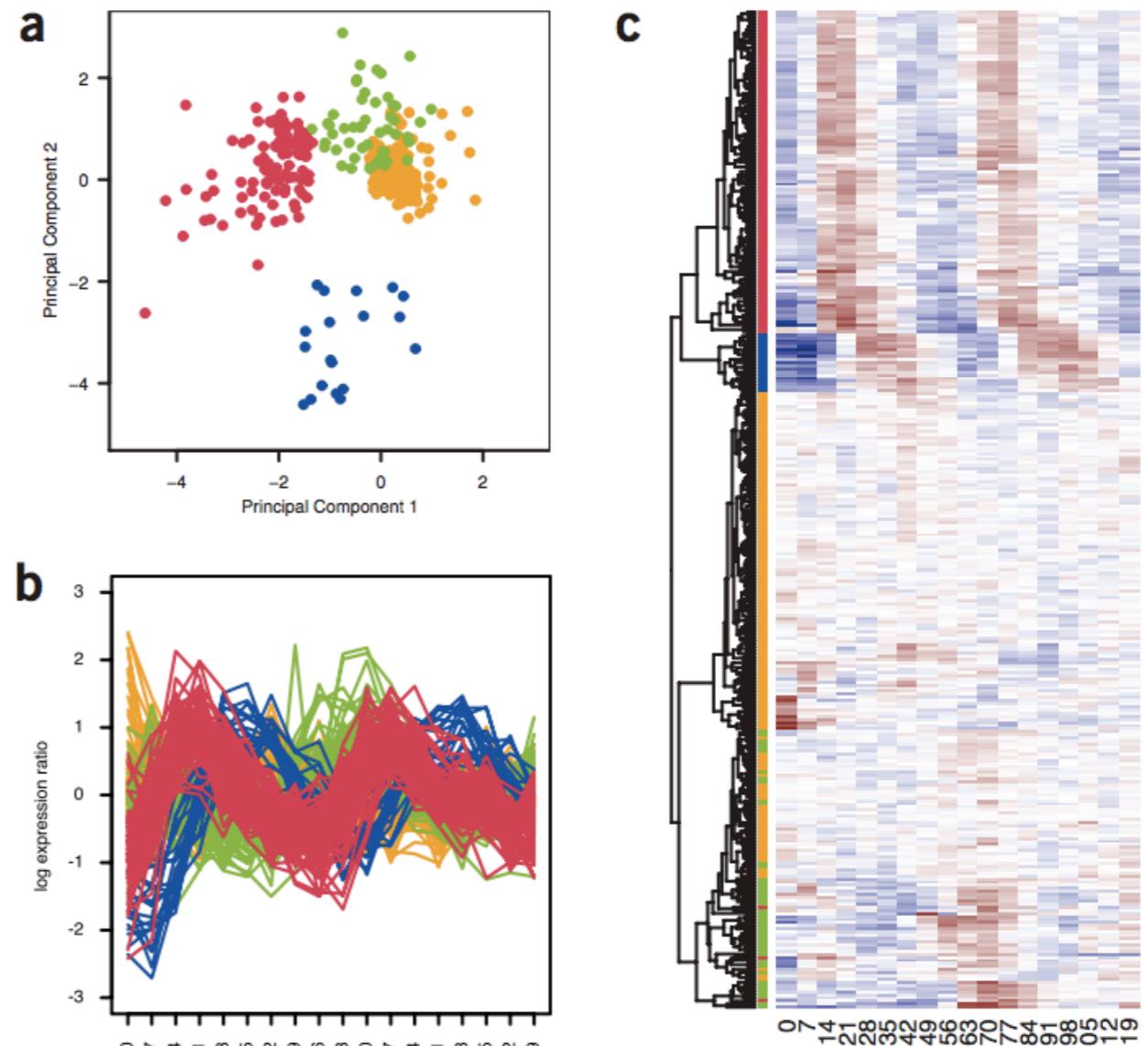
# OUTLINE

- Translate scientific question into statistics
  - Statistical terms for ‘biomarker’ (or ‘signature’)
- Experimental design
  - Replication, randomization, blocking
- Case study: iPRG 2015-2016
  - Study design and preliminary analysis

# STATISTICAL GOAL I: CLASS DISCOVERY

Discover proteins or subjects with similar patterns

- No known class labels
  - E.g., no ‘healthy’ or ‘disease’
  - All variation treated equally
  - No error rates
- Can’t find something meaningful if unsure what we look for
  - Best used for visualization

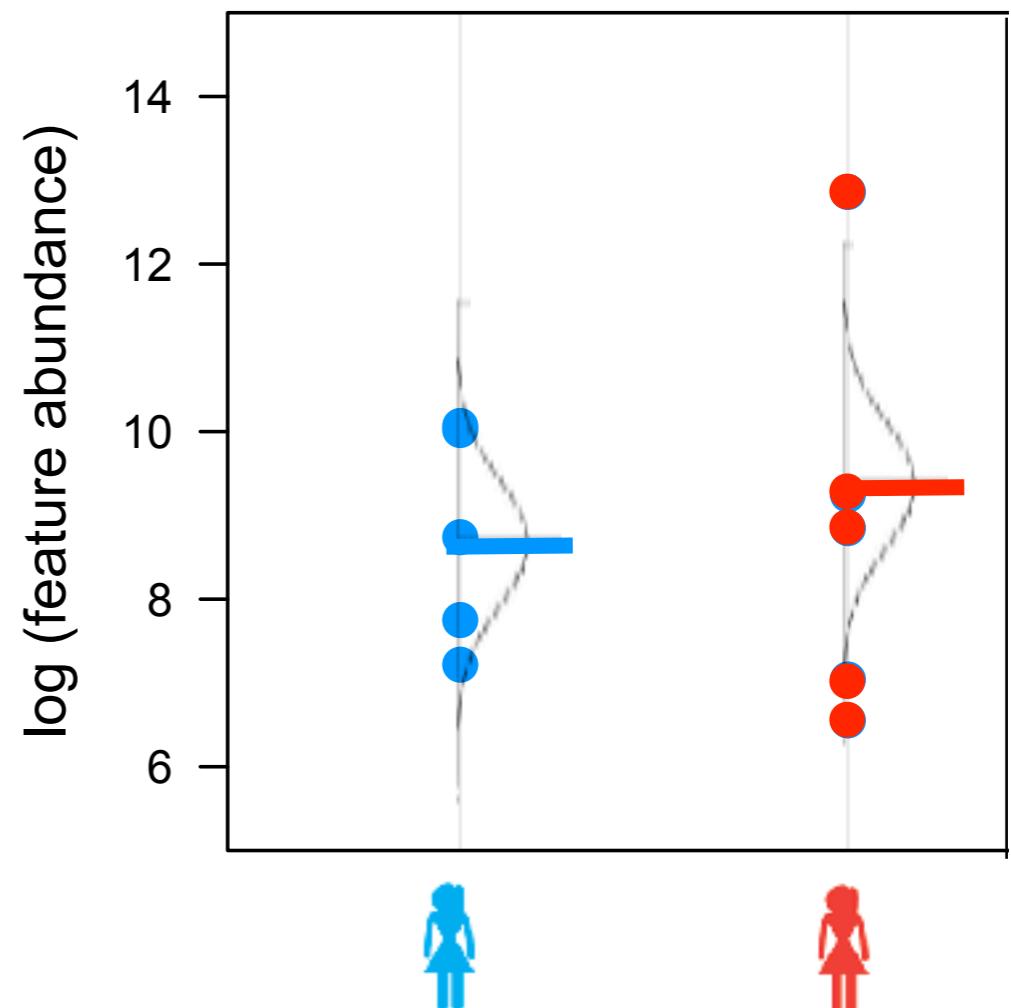


Gehlenborg *et al*, Nature Methods, 2010

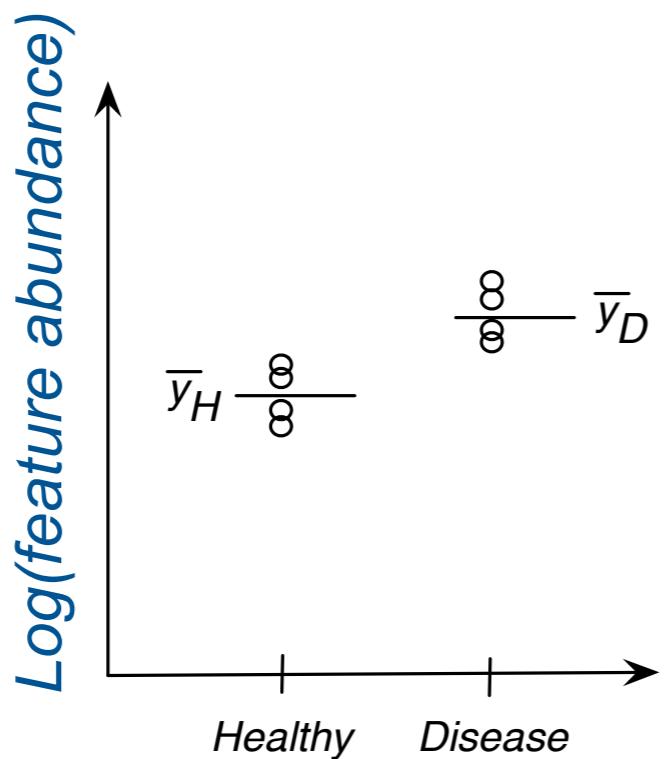
# STATISTICAL GOAL 2: CLASS COMPARISON

Compare mean abundances in subject groups

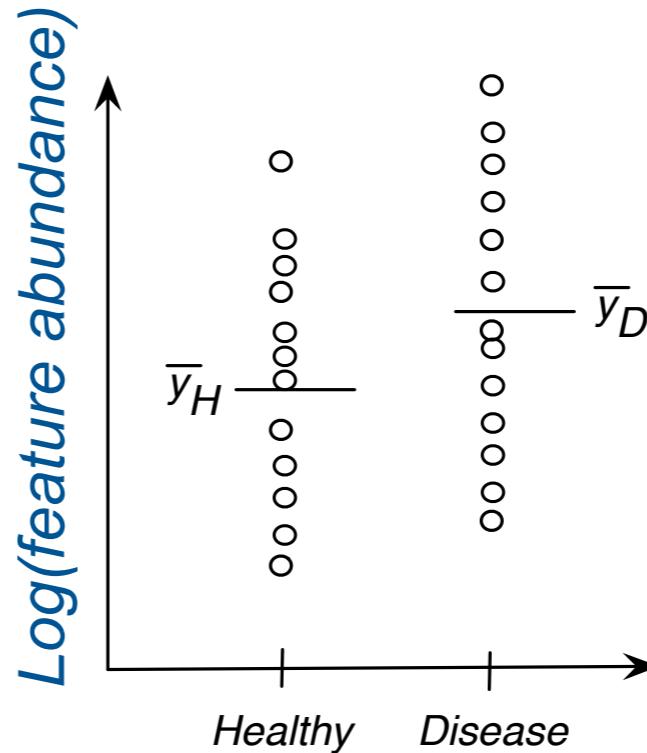
- Known class labels
  - Compare group averages
  - Report p-values, posterior probabilities etc
- Useful when compare groups of subjects
  - Best used for basic biology
  - Initial (Tier III) biomarker discovery screen



# DIFFERENTIALLY ABUNDANT PROTEINS ARE NOT ALWAYS BIOMARKERS



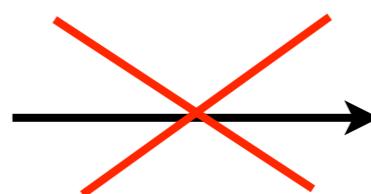
*Differentially abundant  
and predictive*



*Differentially abundant  
and not predictive*

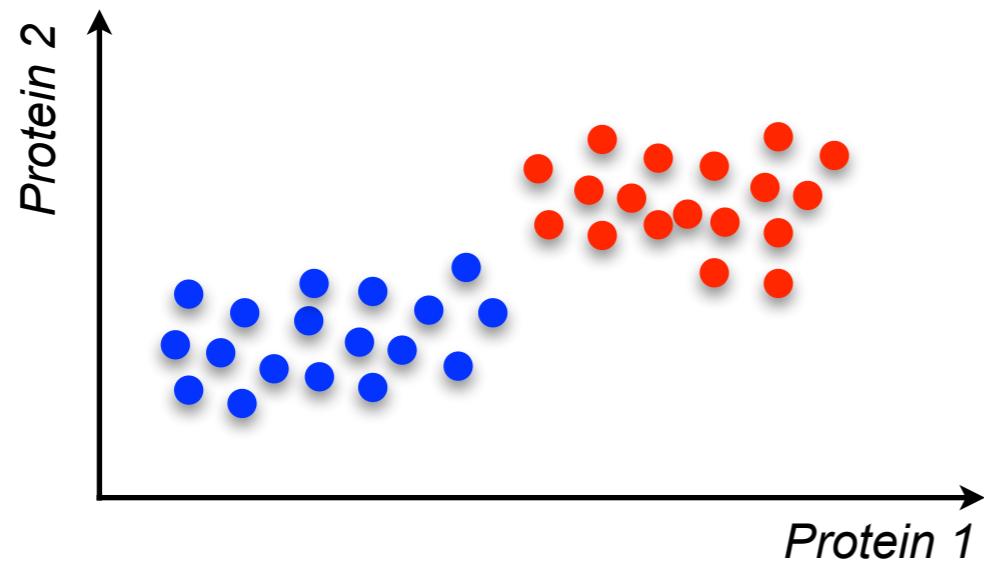
*Single protein:*

*Differentially  
abundant*

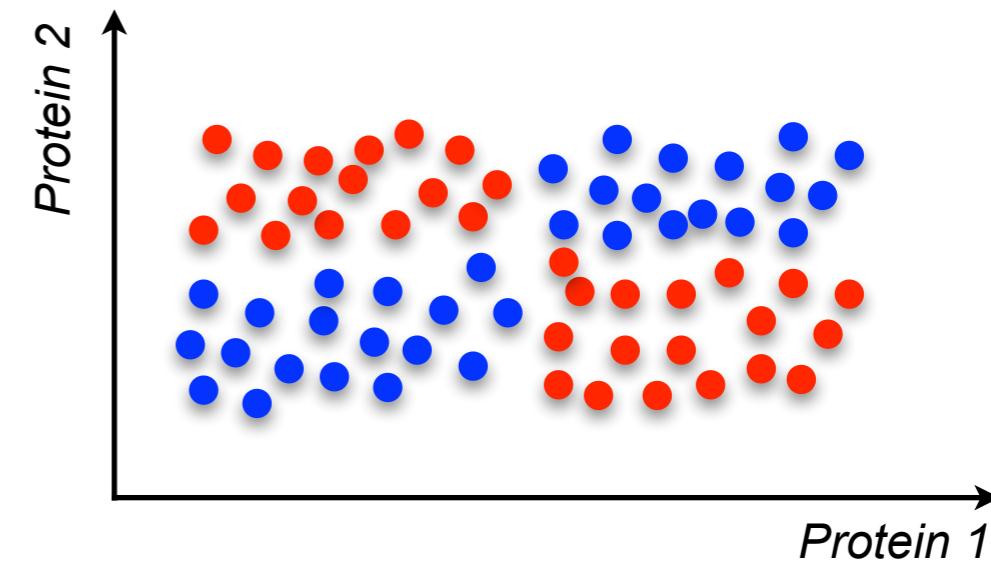


*Predictive*

# BIOMARKER PROTEINS ARE NOT ALWAYS DIFFERENTIALLY ABUNDANT



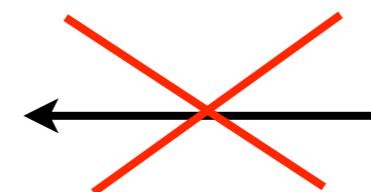
*Differentially  
abundant and  
predictive*



*Not differentially  
abundant but  
predictive*

*Single protein:*

*Differentially  
abundant*

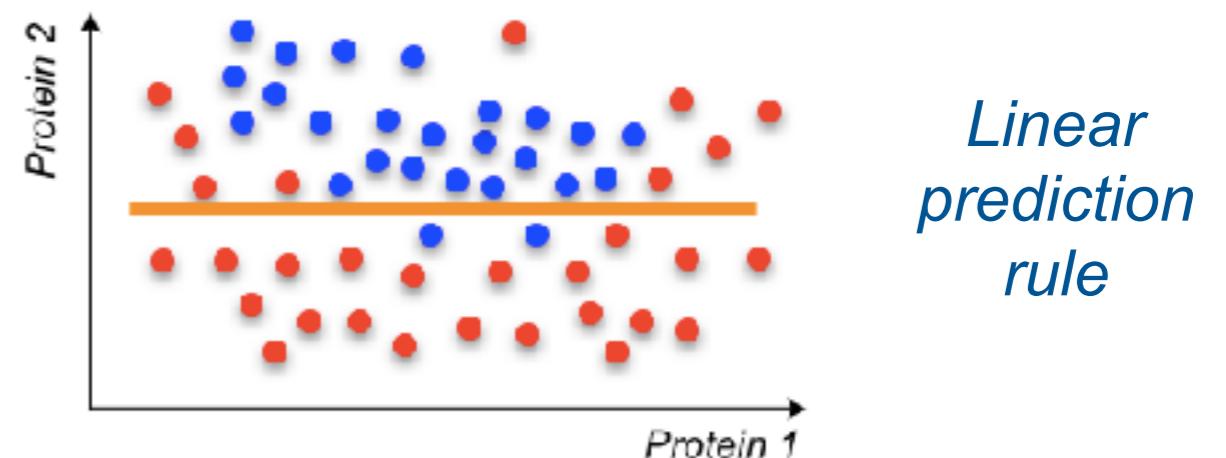


*Predictive*

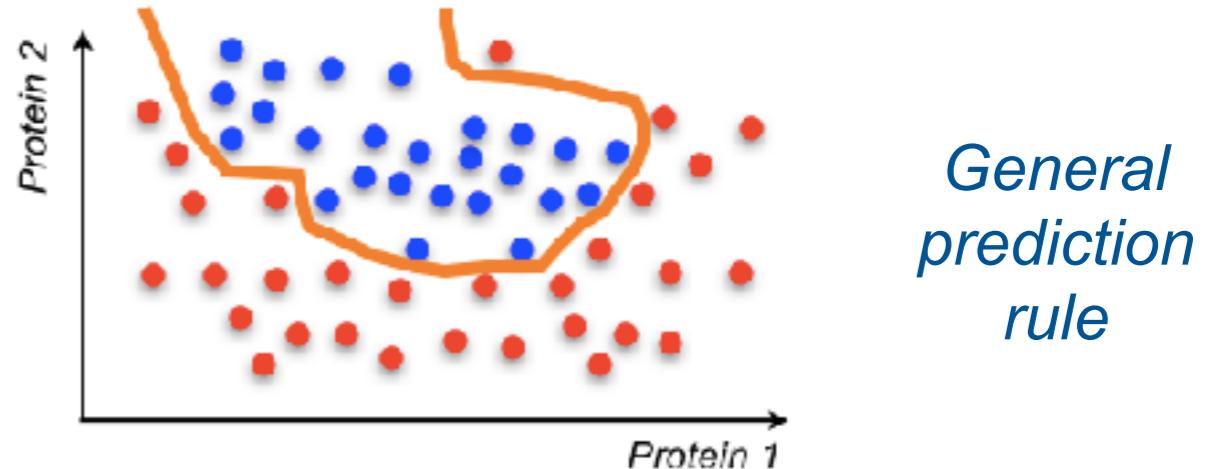
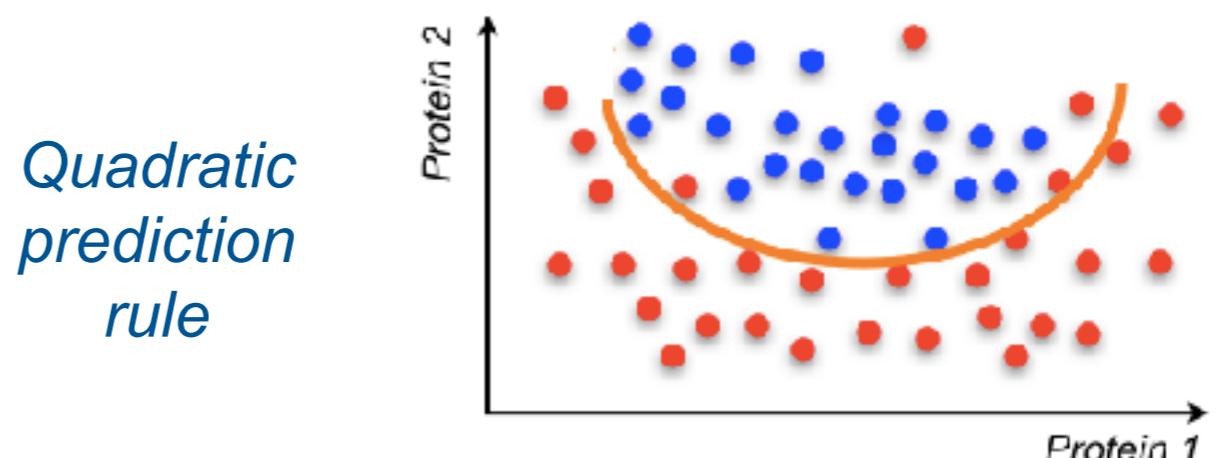
# STATISTICAL GOAL 3: CLASS PREDICTION

Classify each subject into a known group

- Known class labels
  - Predict individual subjects
  - Report misclassification error (sensitivity, specificity, predictive value etc)
- Useful when focus on an individual
  - Tier I or Tier II biomarker discovery studies



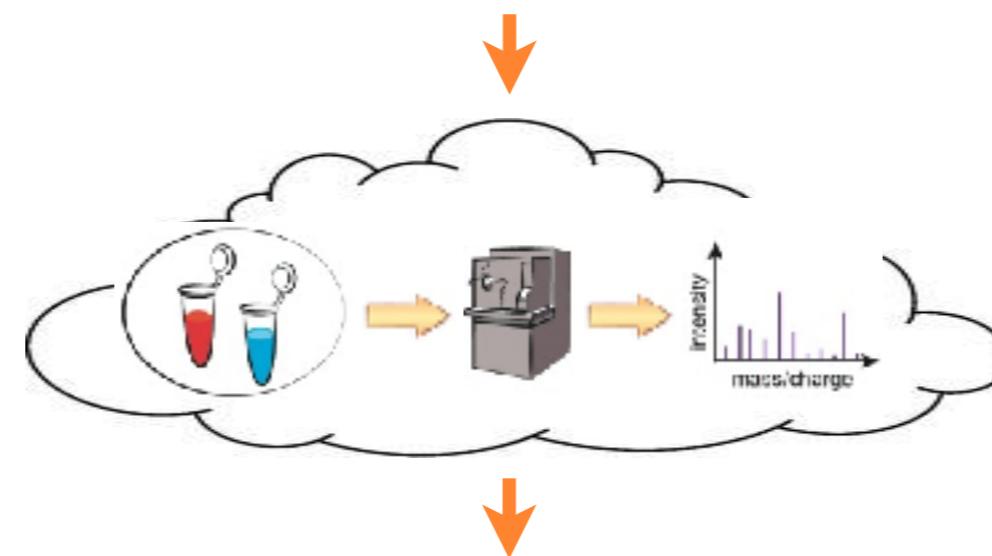
*Quadratic prediction rule*



*General prediction rule*

# ANALYSIS WORKFLOW

## Experimental design



## QC and normalization

## Statistical modeling

**GOAL 2: CLASS  
COMPARISON**

**GOAL 3: CLASS  
PREDICTION**

## Hypothesis testing

## Adjustment for multiple testing

## Experimental design

## Sample quantification

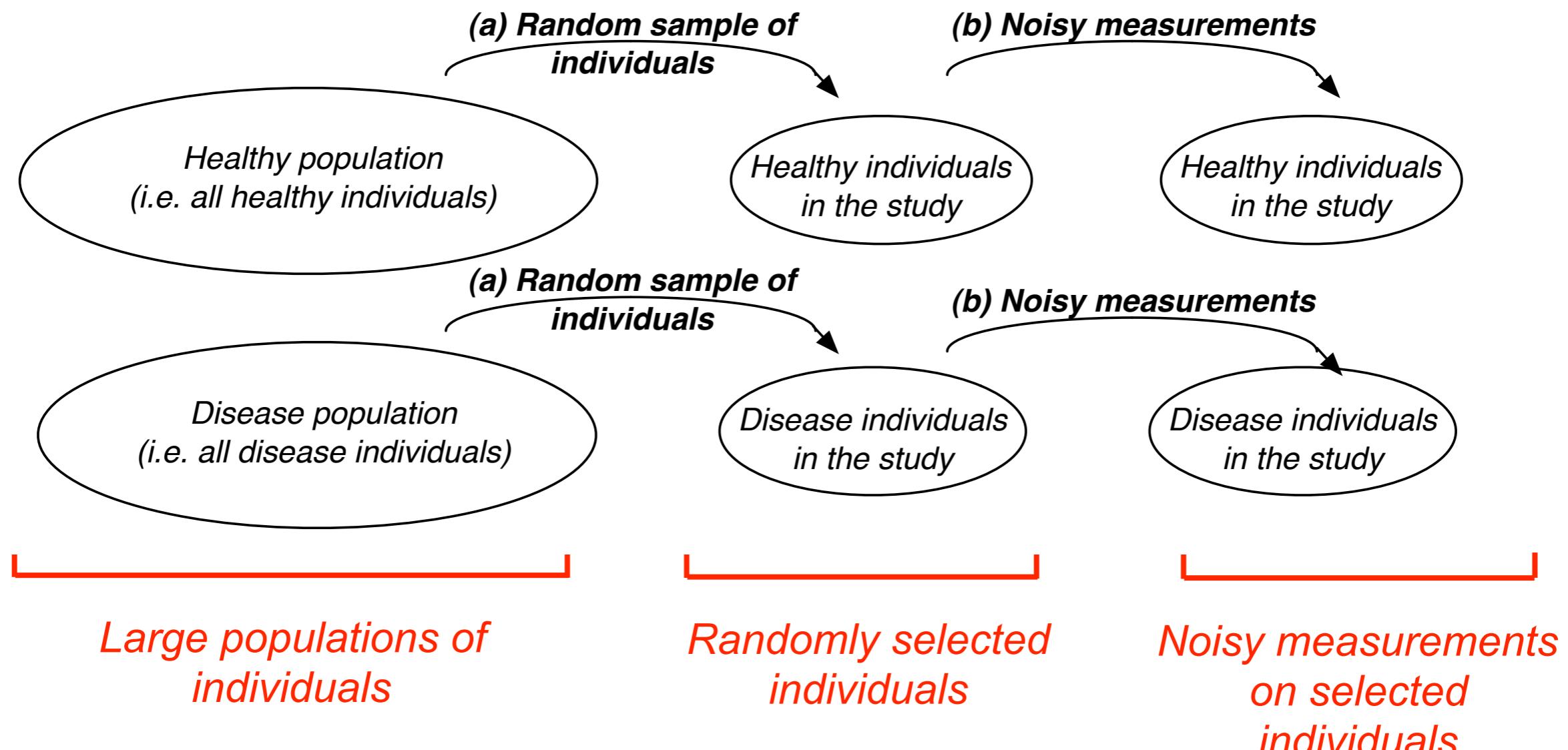
## Classifier development

## Performance evaluation

# OUTLINE

- Translate scientific question into statistics
  - Statistical terms for ‘biomarker’ (or ‘signature’)
- Experimental design
  - Replication, randomization, blocking
- Case study: iPRG 2015-2016
  - Study design and preliminary analysis

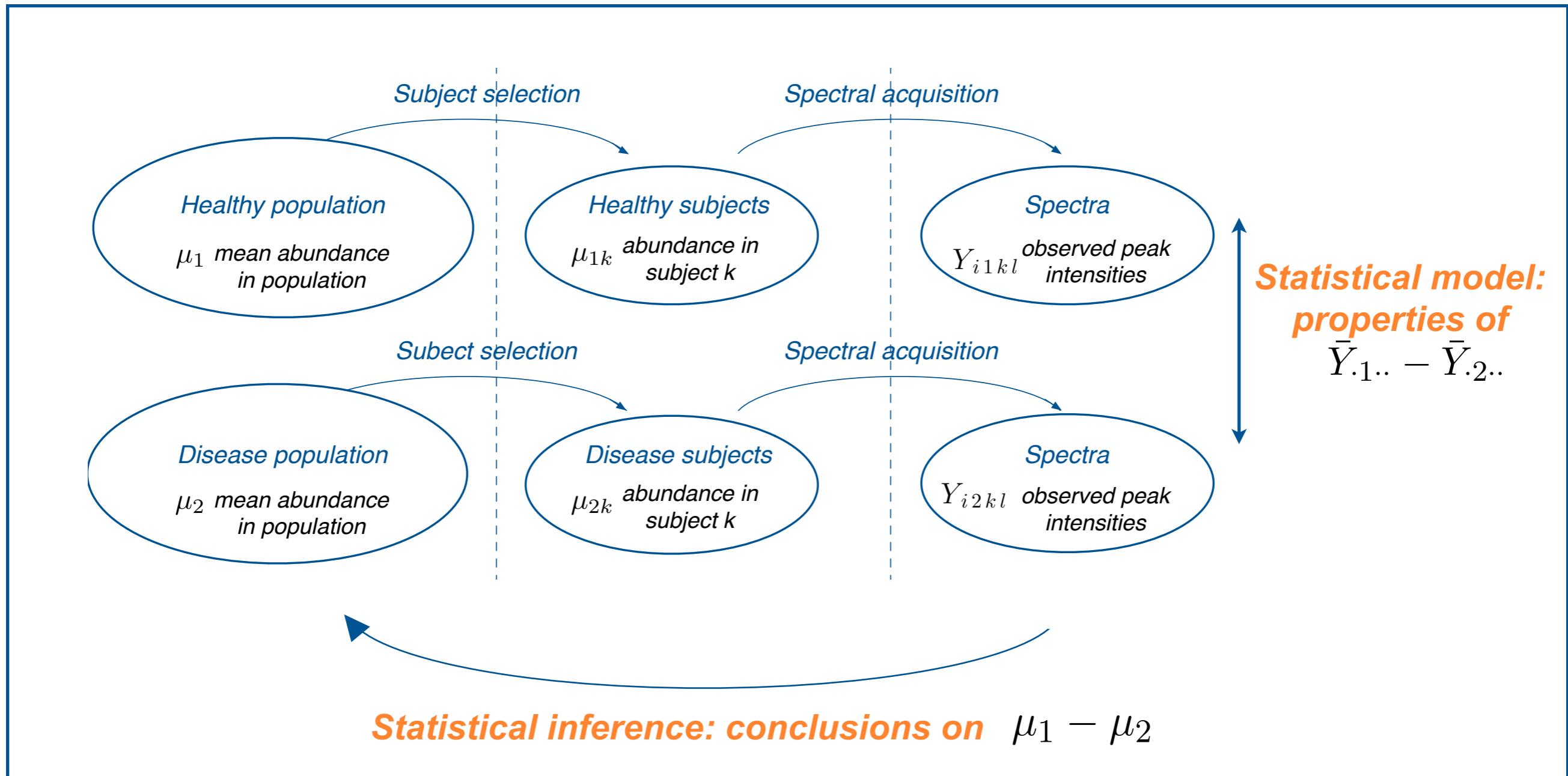
# A STATISTICIAN'S VIEW OF THE EXPERIMENT



## Dangers:

*Bias: conclusions systematically differ from truth*  
*Inefficiency: unnecessary variation in the data*

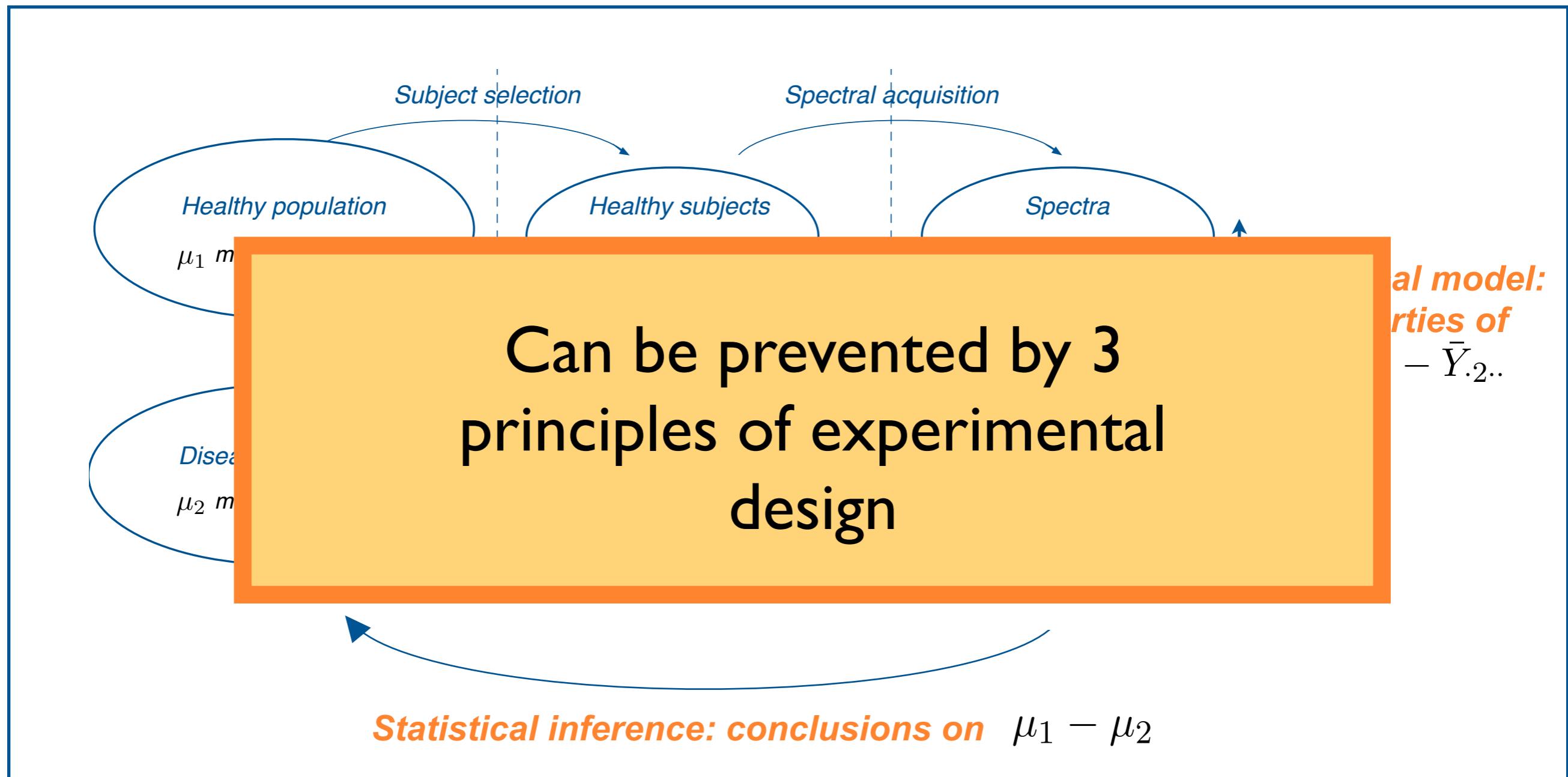
# DEFINITION OF BIAS AND INEFFICIENCY



**Bias:**  $\bar{Y}_{1..} - \bar{Y}_{2..}$  systematically different from  $\mu_{1k} - \mu_{2k}$

**Inefficiency:** Large  $Var(\bar{Y}_{1..} - \bar{Y}_{2..})$

# DEFINITION OF BIAS AND INEFFICIENCY

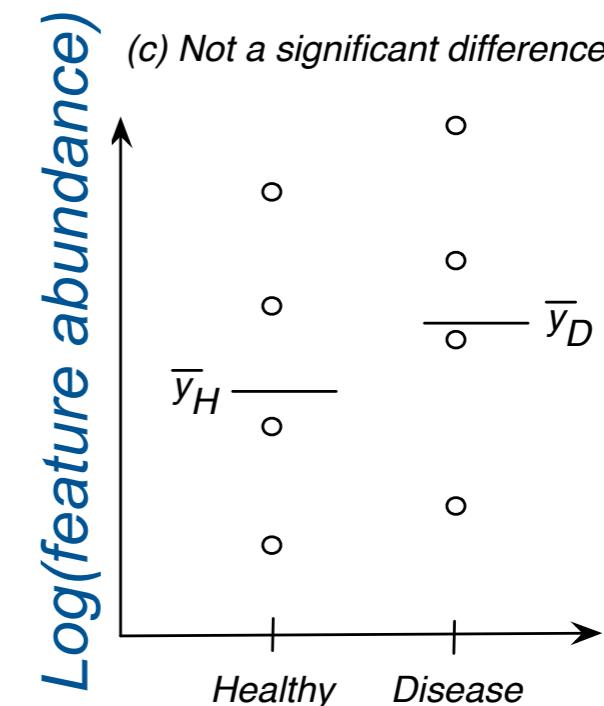
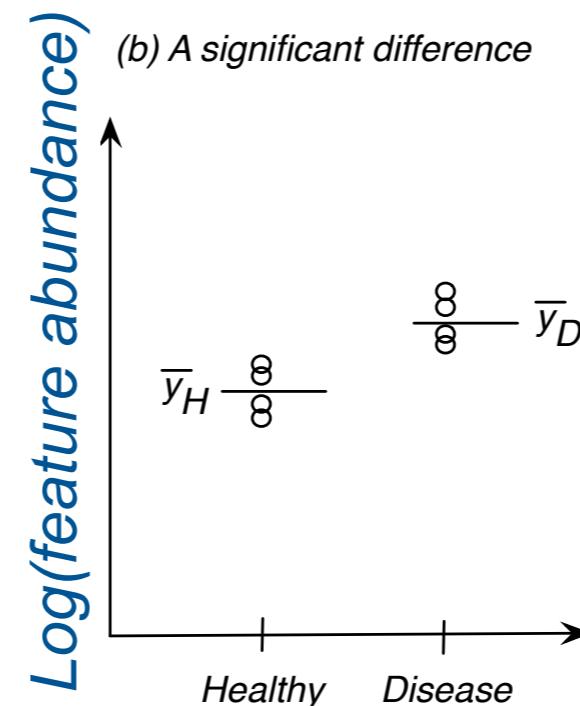
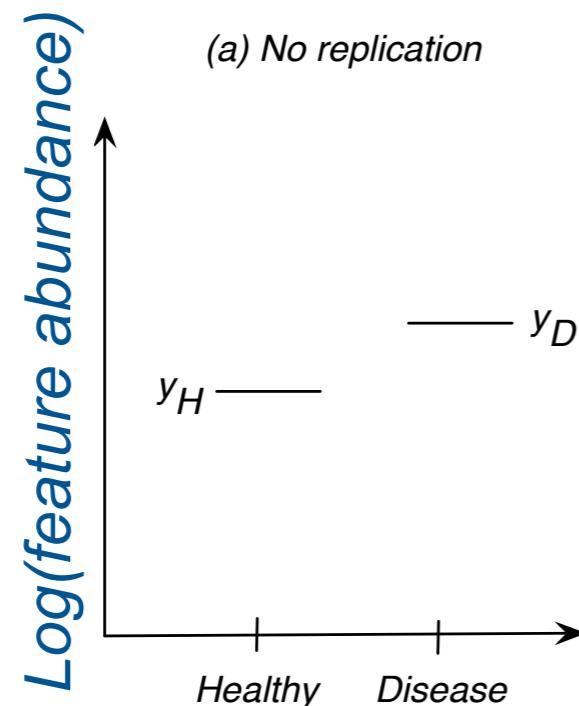
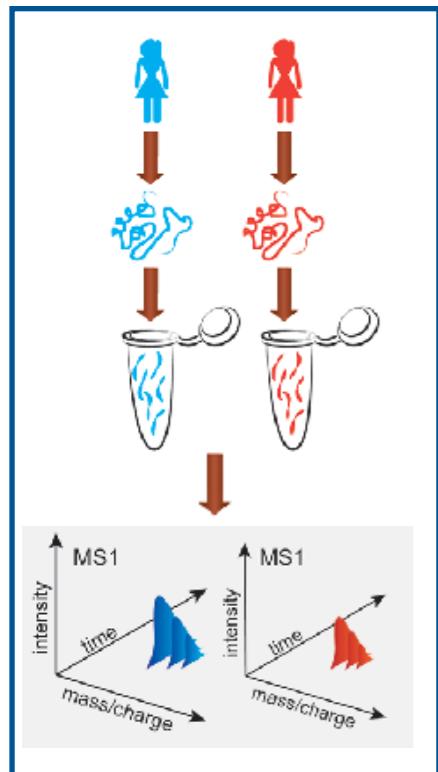


**Bias:**  $\bar{Y}_{.1..} - \bar{Y}_{.2..}$  systematically different from  $\mu_{1k} - \mu_{2k}$

**Inefficiency:** Large  $Var(\bar{Y}_{.1..} - \bar{Y}_{.2..})$

# PRINCIPLE I: REPLICATION

(1) carries out the inference and (2) minimizes inefficiencies

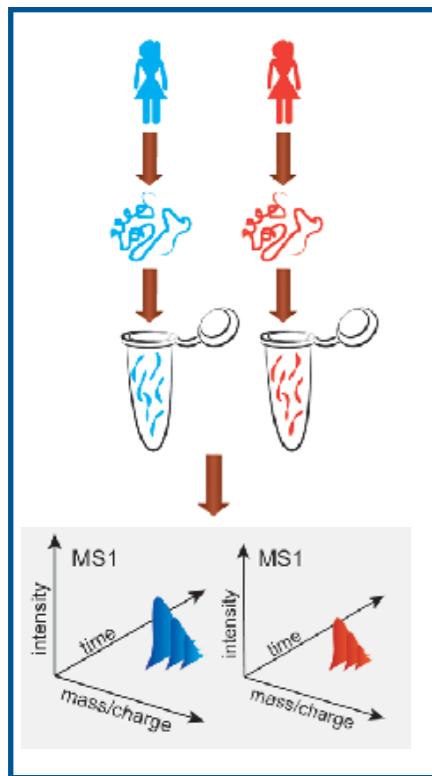


Two levels of randomness imply two types of replication:

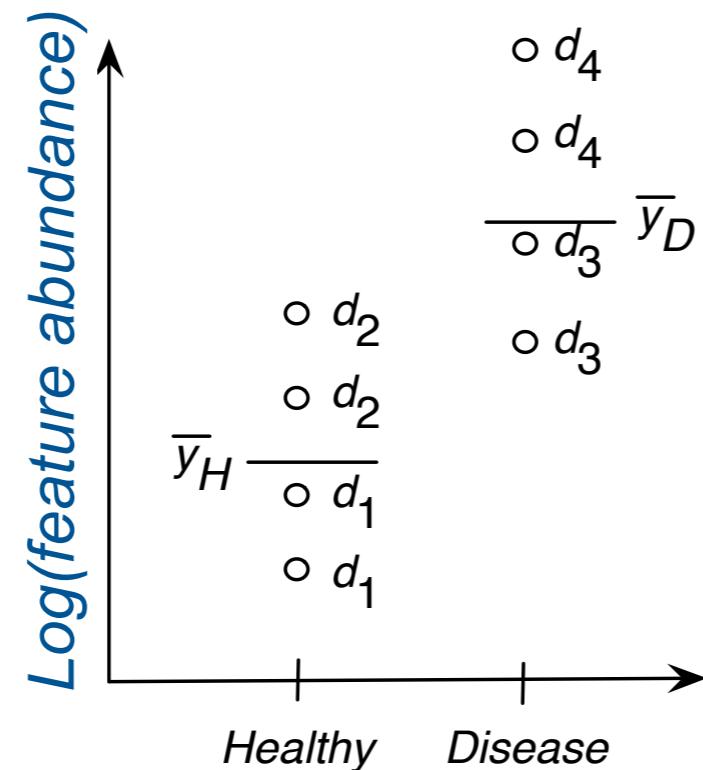
- ◆ *Biological replicates*: selecting multiple subjects from the population
- ◆ *Technical replicates*: multiple runs per subject

# PRINCIPLE 2: RANDOMIZATION

Prevents bias

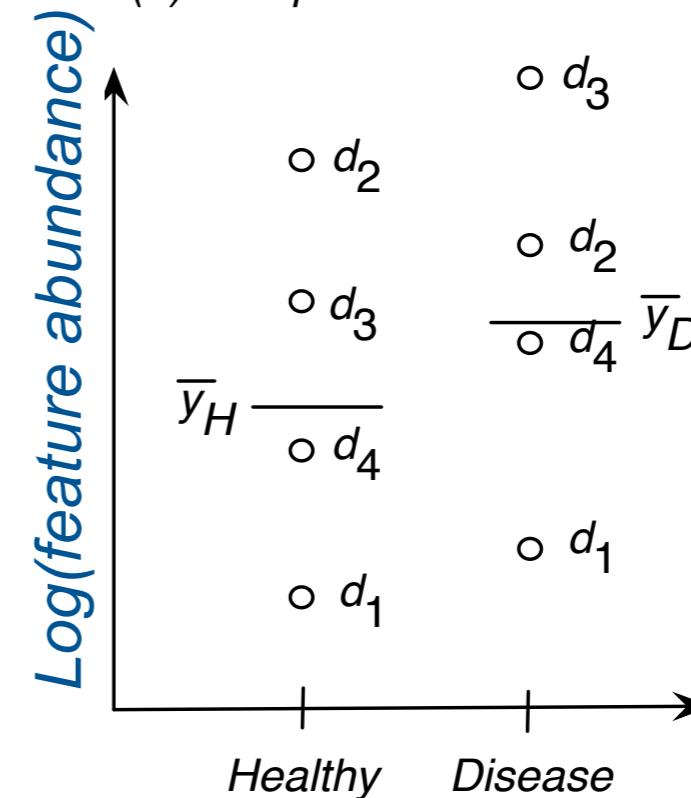


(a) Sequential acquisition



No randomization  
= confounding  
= bias

(b) Complete randomization



Complete randomization  
= no bias

Two levels of randomness imply two types of randomization:

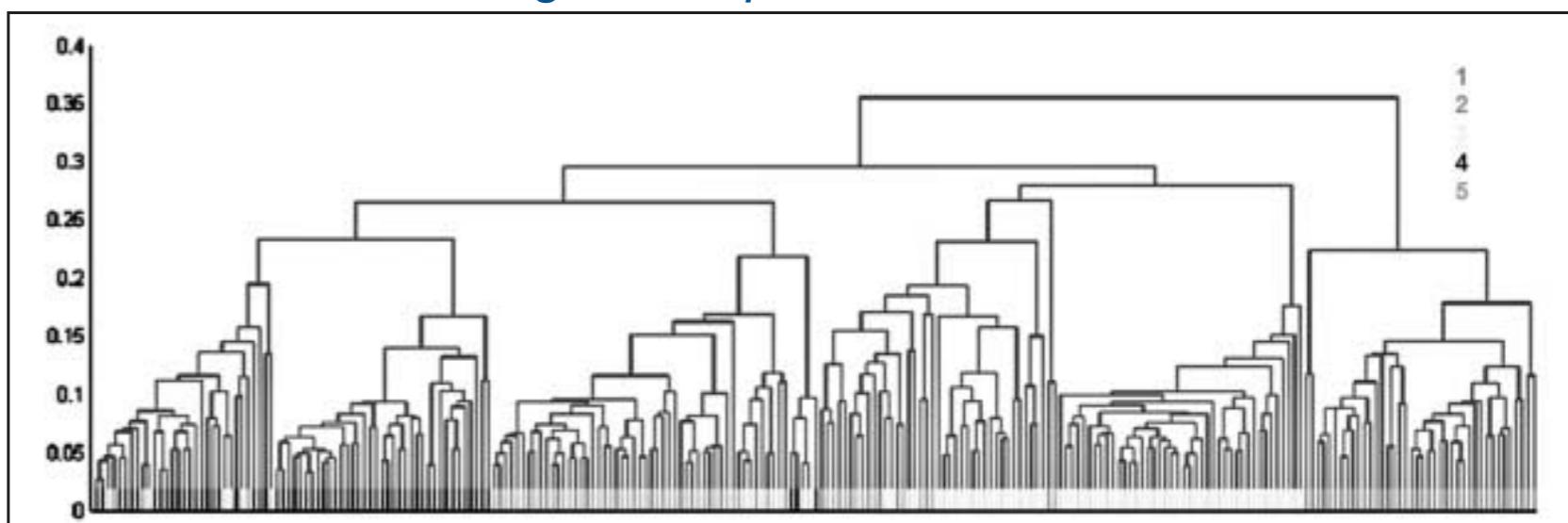
- ◆ *Biological replicates*: random selection of subjects from the population
- ◆ *Technical replicates*: random allocation of samples to all processing steps

# EXAMPLE: LACK OF RANDOMIZATION

Hu, Coombes, Morris, Baggerly, *Briefings in Functional Genomics*, 2005

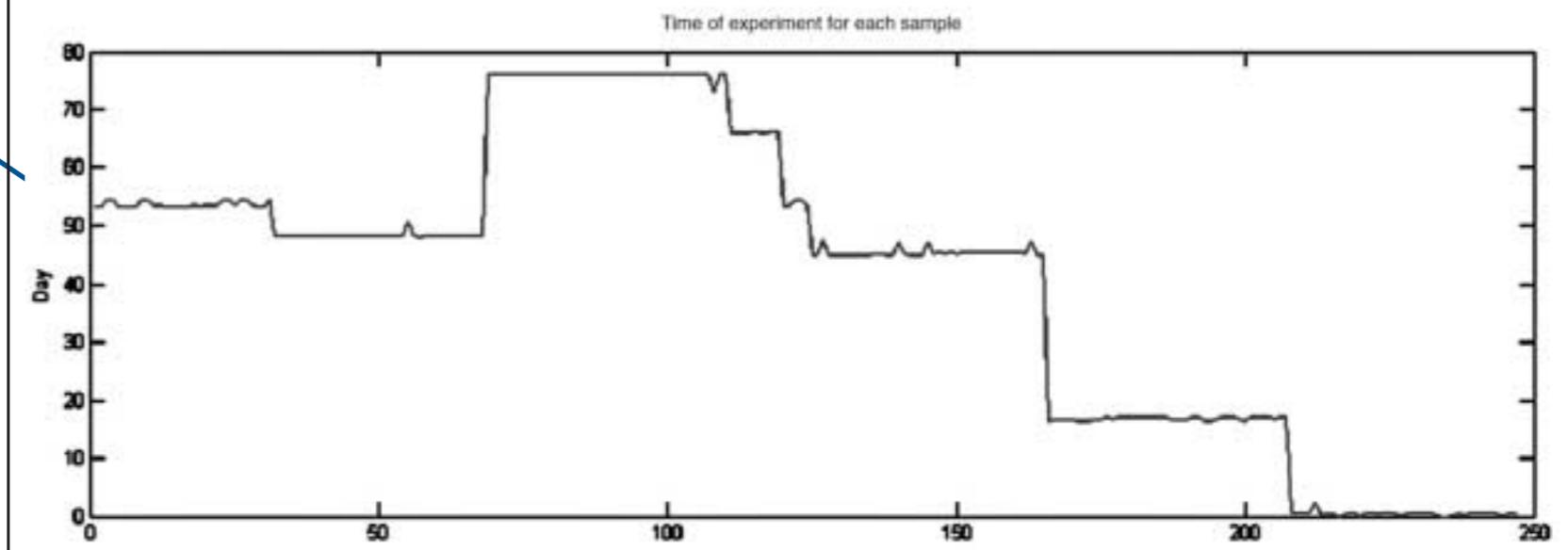
- Serum samples with five types of cancer
- SELDI-TOF MS
  - ◆ normalized, peak picked

*Hierarchical clustering of samples*



*Cancer subtype  
confounded with  
time*

*Same time-  
based clustering  
on the QC  
samples!*



# BEWARE OF BIG EFFECTS THEY ARE LIKELY TO REFLECT FLAWS OF THE DESIGN

- Study of gene expression between Asians and Europeans
- Found that 78% of genes are differentially
  - Asians were profiles in one year, and Europeans in another
  - The difference therefore likely reflects a batch effect



Journal home > Archive > Letter > Full Text

| Journal content                              |
|--|
| <a href="#">+ Journal home</a>               |
| <a href="#">+ Advance online publication</a> |
| <a href="#">+ Current issue</a>              |
| <a href="#">+ Archive</a>                    |
| <a href="#">+ Focuses and Supplements</a>    |

## Letter

*Nature Genetics* 39, 226 - 231 (2007)  
Published online: 7 January 2007 | doi:10.1038/ng1955

### Common genetic variants account for differences in gene expression among ethnic groups

Richard S. Spielman<sup>1</sup>, Laurel A. Bastone<sup>2</sup>, Joshua T. Burdick<sup>3</sup>, Michael Morley<sup>3</sup>, Warren J. Ewens<sup>4</sup> & Vivian G. Cheung<sup>1,3,5</sup>



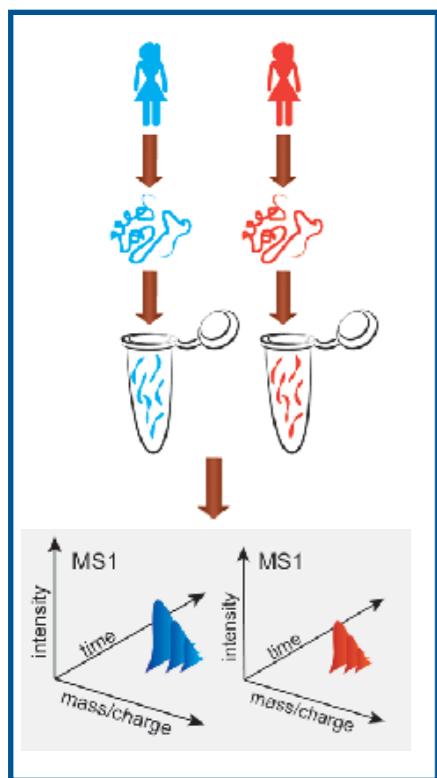
| This issue                        |
|-----------------------------------|
| <a href="#">Table of contents</a> |
| <a href="#">Previous article</a>  |

**Source: a blog by Jeff Leek, Biostatistics, John Hopkins University**

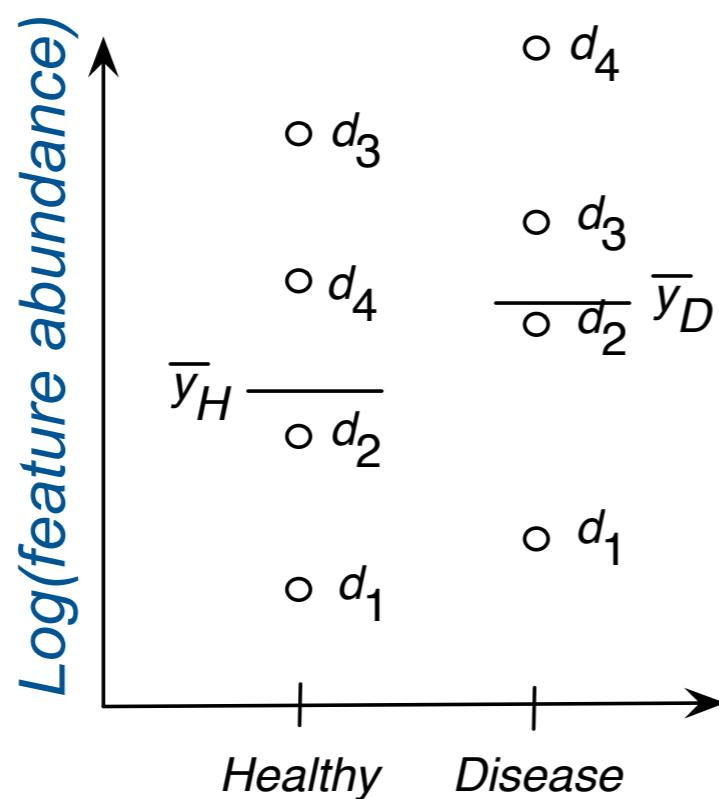
<http://simplystatistics.org/2016/02/01/a-menagerie-of-messed-up-data-analyses-and-how-to-avoid-them/>

# PRINCIPLE 3: BLOCKING

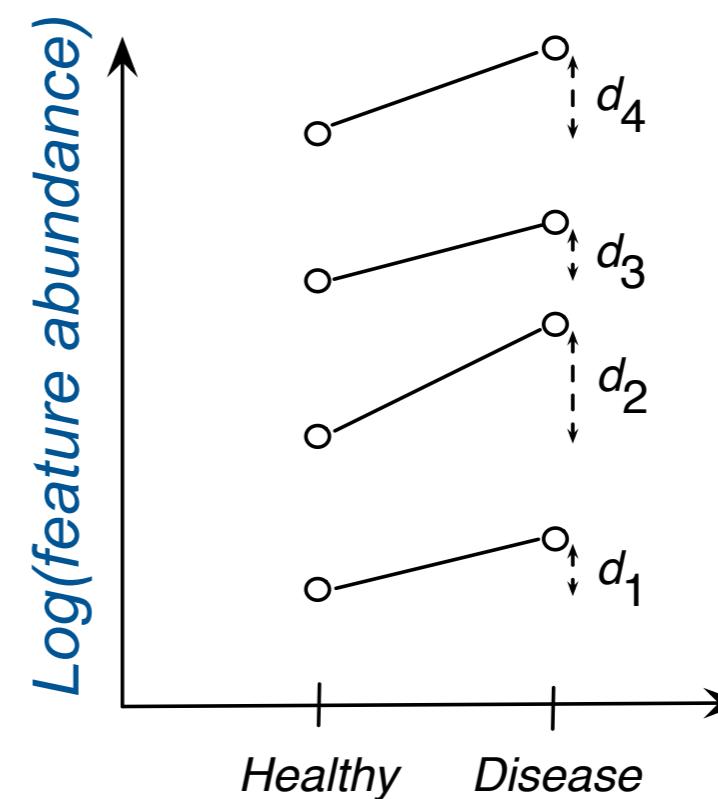
## Helps reduce both bias and inefficiency



(b) Complete randomization



(c) Day = block



Complete randomization  
= inflated variance

Block-randomization  
= restriction on randomization  
= systematic allocation

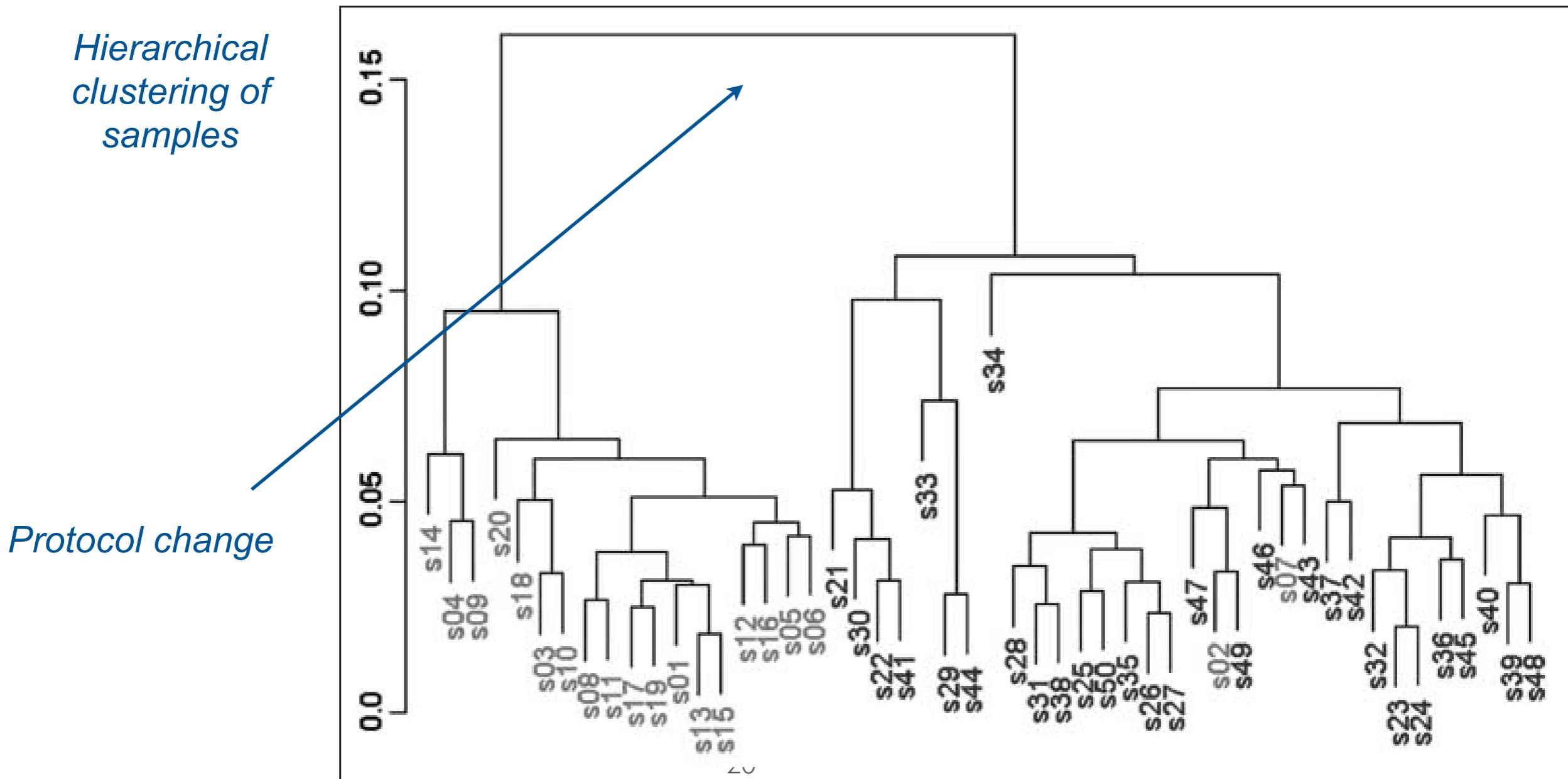
Two levels of randomness imply two types of blocks:

- ◆ *Biological replicates*: subjects having similar characteristics (e.g. age)
- ◆ *Technical replicates*: samples processed together (e.g. in a same day)

# EXAMPLE: LACK OF BLOCKING

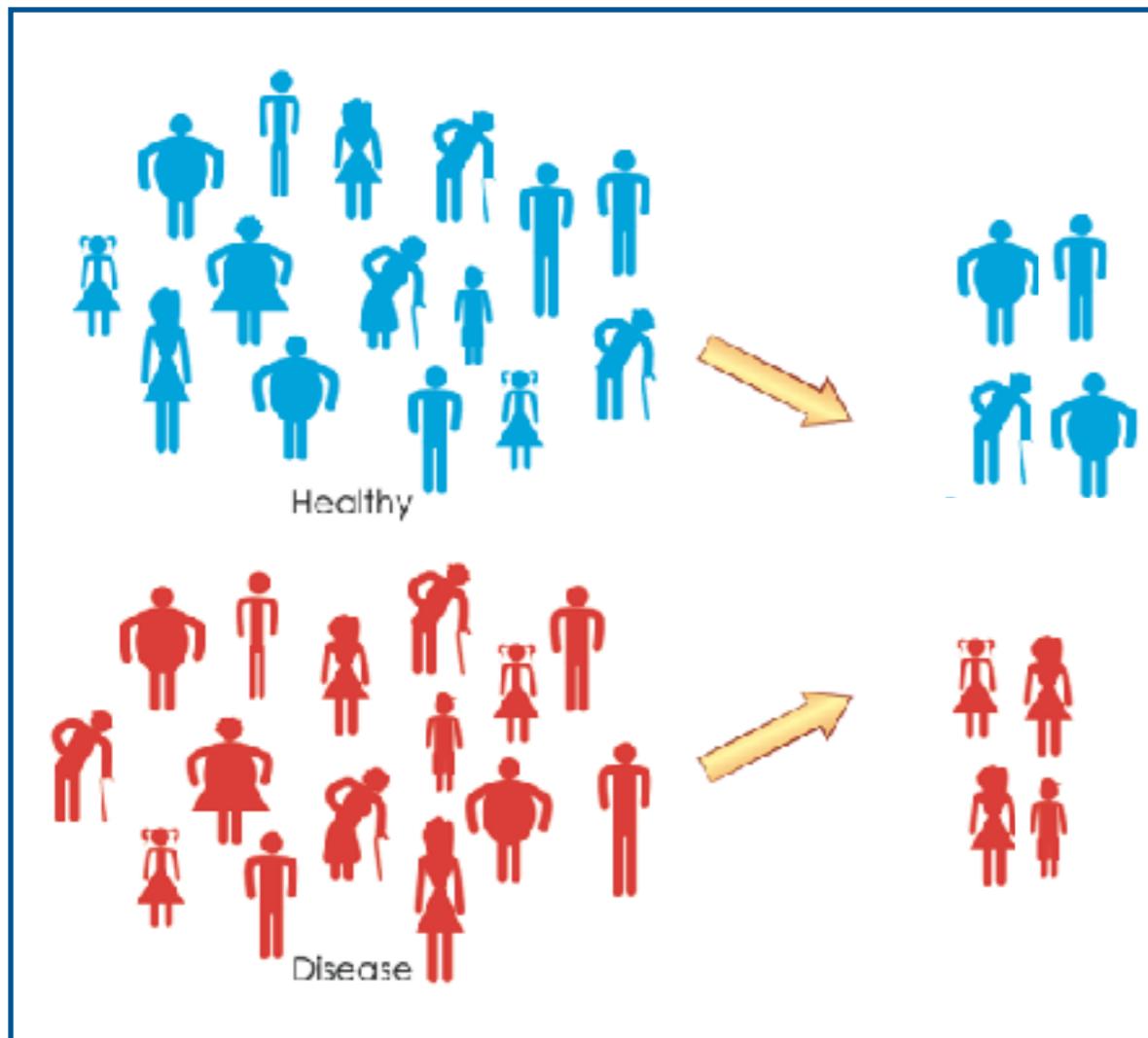
Hu, Coombes, Morris, Baggerly, *Briefings in Functional Genomics*, 2005

- Serum samples with two types of cancer
- SELDI-TOF MS, 3 fractions
  - ◆ normalized, peak picked

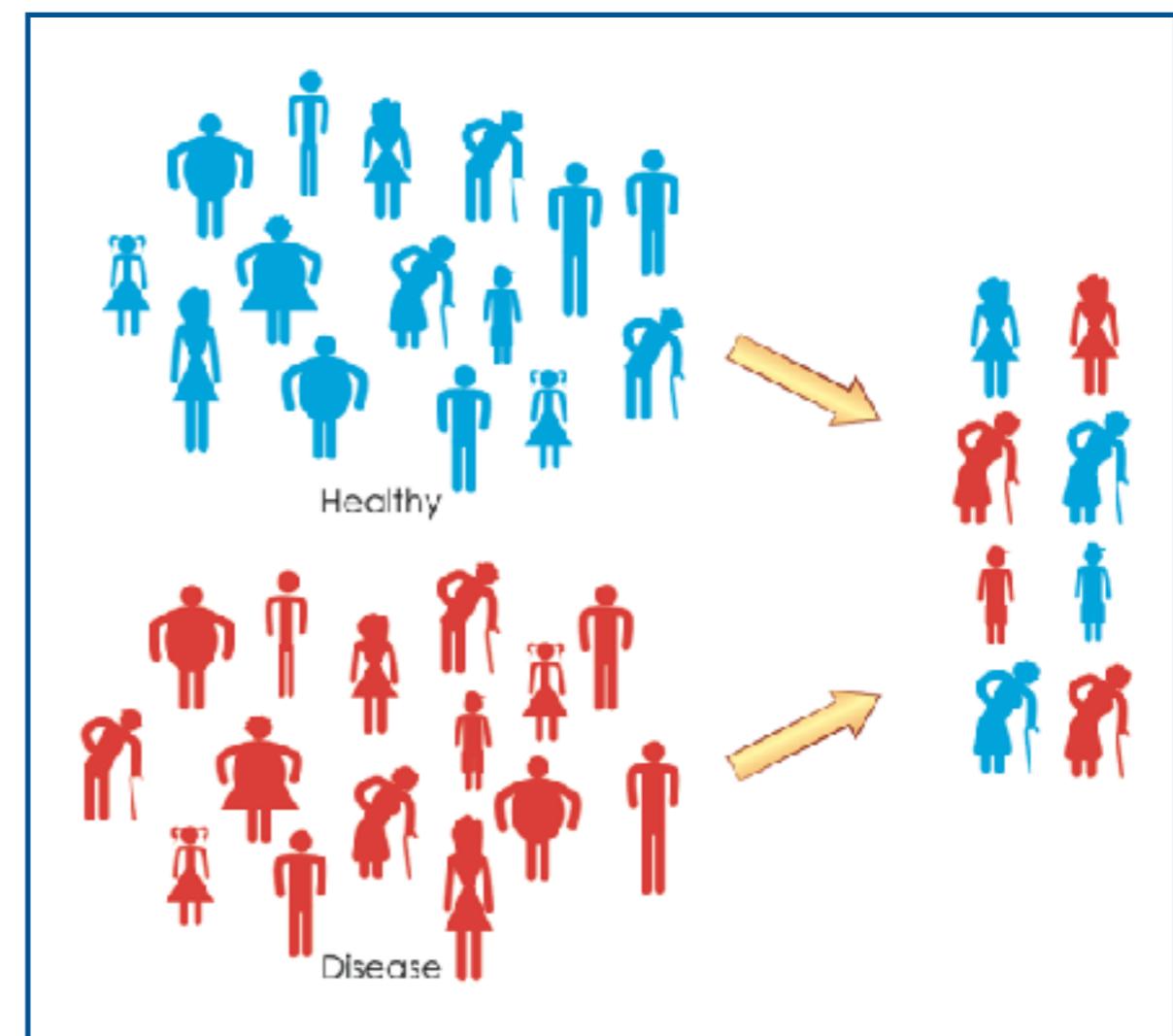


# MATCHING

Blocking with respect to biological risk factors



Complete randomization  
= inflated variance



Block-randomization  
= restriction on randomization  
= systematic allocation

# EXAMPLE

## Block-randomized selection of subjects from repository

|                |                       | Disease group |               |                 |        |       |
|----------------|-----------------------|---------------|---------------|-----------------|--------|-------|
|                |                       | Control       | Stable angina | Unstable angina | NSTEMI | STEMI |
| Stratification | $\geq 58$ y.o; Female | 354           | 300           | 49              | 39     | 29    |
|                | $\geq 58$ y.o; Male   | 701           | 843           | 143             | 86     | 54    |
|                | < 58 y.o; Female      | 80            | 56            | 5               | 5      | 8     |
|                | < 58 y.o; Male        | 264           | 190           | 34              | 23     | 27    |

*Counts in the initial repository of samples*

|                |                       | Disease group |               |                 |        |       |
|----------------|-----------------------|---------------|---------------|-----------------|--------|-------|
|                |                       | Control       | Stable angina | Unstable angina | NSTEMI | STEMI |
| Stratification | $\geq 58$ y.o; Female | 3             | 3             | 3               | 3      | 3     |
|                | $\geq 58$ y.o; Male   | 3             | 3             | 3               | 3      | 3     |
|                | < 58 y.o; Female      | 2             | 2             | 2               | 2      | 2     |
|                | < 58 y.o; Male        | 2             | 2             | 2               | 2      | 2     |

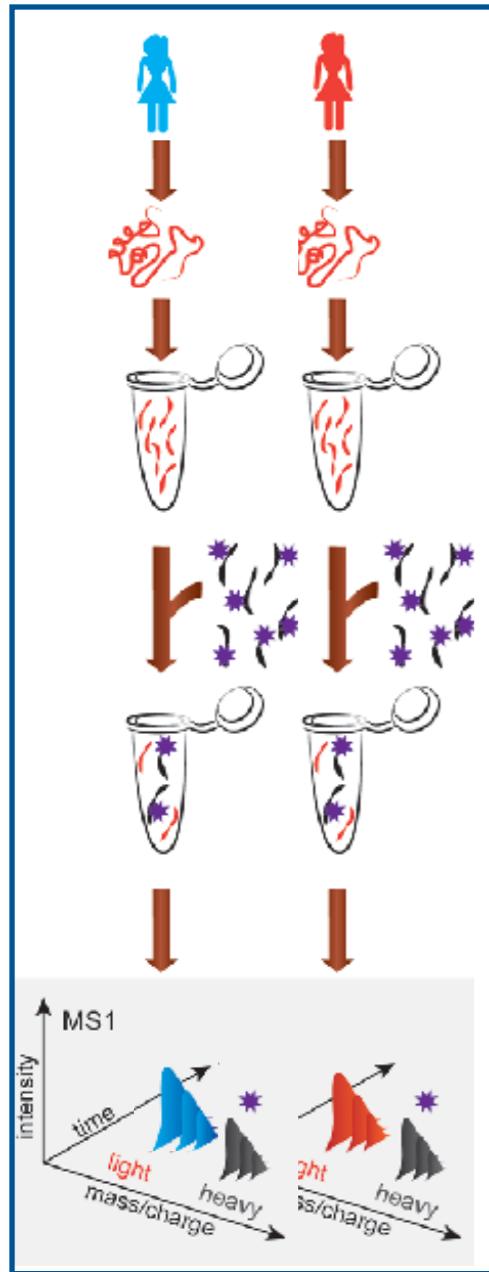
*Counts of subjects included in the study*

*Mass spectra acquired without technical replication*

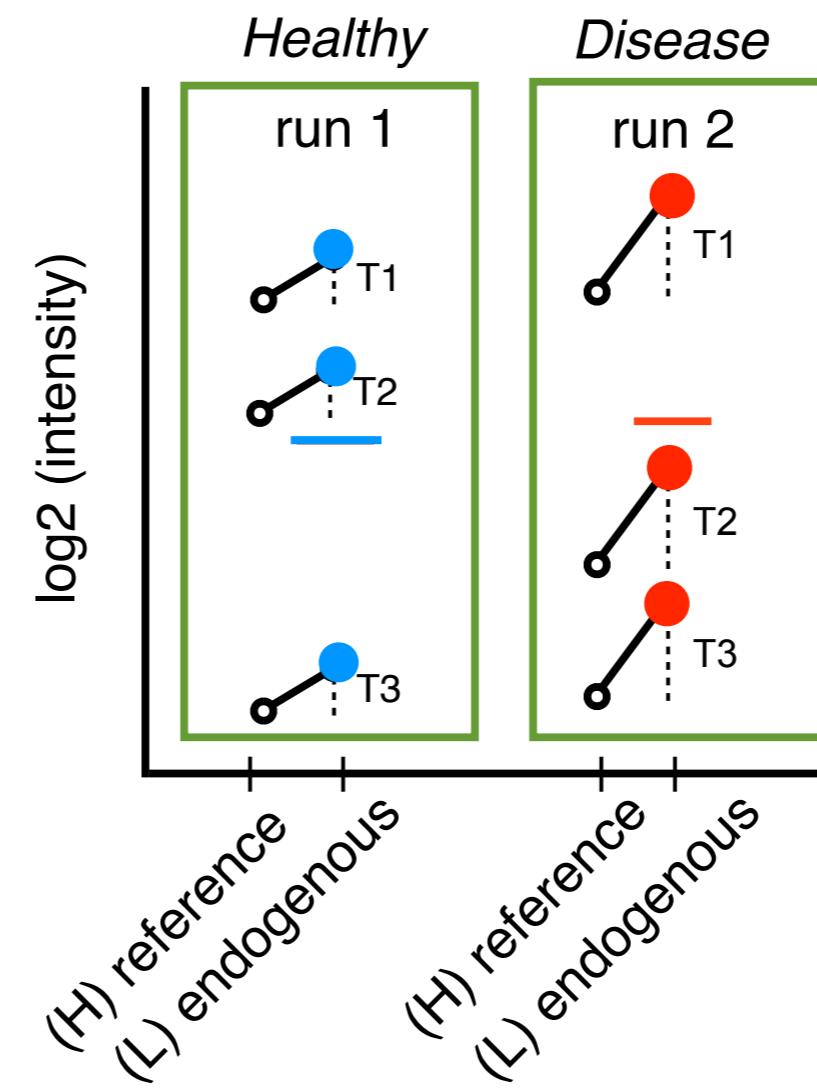
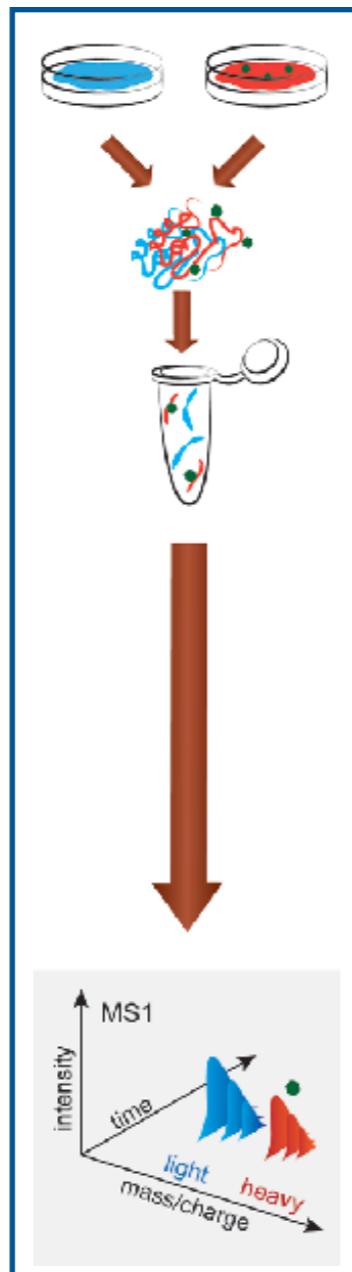
# LABELING (=MULTIPLEXING)

Blocking with respect to mass spectrometry run

## Synthetic standards



## SILAC



- T transition
- average of log-intensities
- paired log-intensities of L and H
- ⋮ difference of log-intensities of L and H  
= log-ratios of L over H
- run

Multiplexing reduces both bias and variance  
(assuming that extra sample handling does not introduce extra variation)

# Martin Krzywinski & Naomi Altman

**nature methods**

Techniques for life scientists and chemists

[nature.com](#) ▶ [journal home](#) ▶ [archive](#) ▶ [issue](#) ▶ [this month](#) ▶ [abstract](#)

*NATURE METHODS* | THIS MONTH

## Points of significance: Importance of being uncertain

[Martin Krzywinski & Naomi Altman](#)

[Affiliations](#)

*Nature Methods* 10, 809–810 (2013) | doi:10.1038/nmeth.2613

## Points of significance: Comparing samples—part I

[Martin Krzywinski & Naomi Altman](#)

## Points of Significance: Error bars

[Martin Krzywinski & Naomi Altman](#)

[Affiliations](#)

*Nature Methods* 10, 921–922 (2013) | doi:10.1038/nmeth.2659

## Points of significance: Power and sample size

[Martin Krzywinski & Naomi Altman](#)

[Affiliations](#)

*Nature Methods* 10, 1139–1140 (2013) | doi:10.

## Points of significance: Significance, *P* values and *t*-tests

[Martin Krzywinski & Naomi Altman](#)

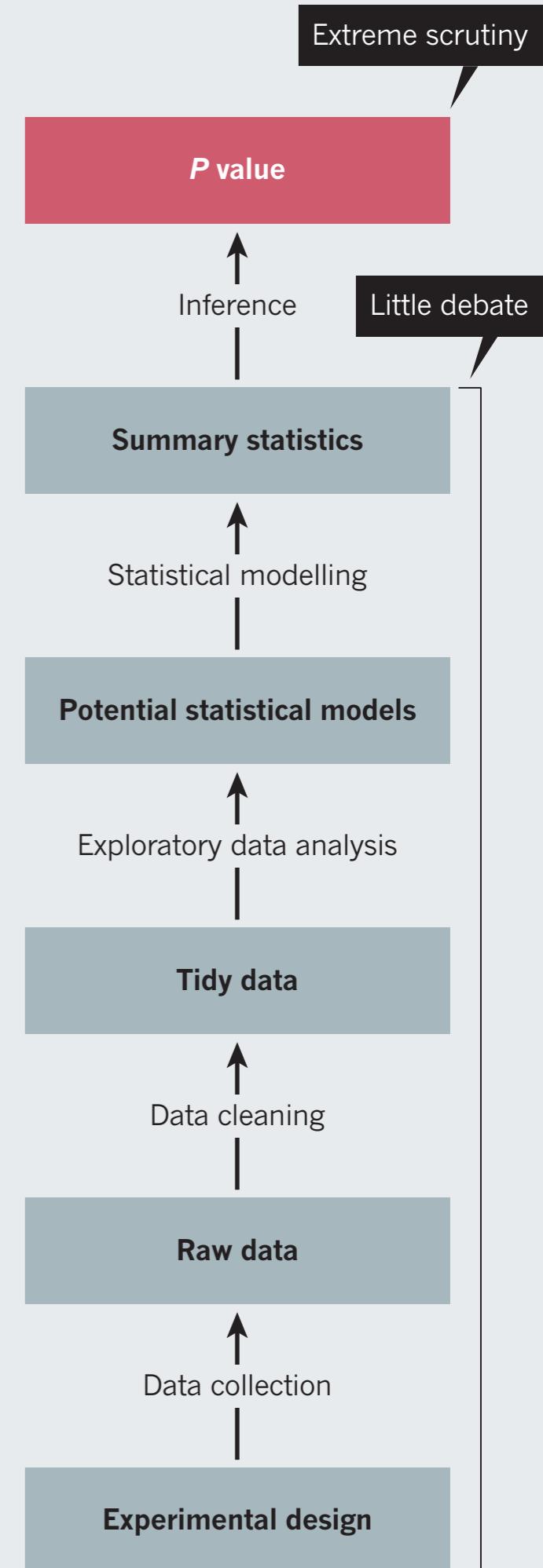
**NATURE | COMMENT**

## Statistics: *P* values are just the tip of the iceberg

**Jeffrey T. Leek & Roger D. Peng**

28 April 2015

Statistical  
considerations are key  
at every step



# OUTLINE

- Translate scientific question into statistics
  - Statistical terms for ‘biomarker’ (or ‘signature’)
- Experimental design
  - Replication, randomization, blocking
- Case study: iPRG 2015-2016
  - Study design and preliminary analysis

# ABRF IPRG STUDY 2015

## Detection of differentially abundant proteins in controlled mixture

| Name                   | Origin              | Molecular Weight | Samples |     |    |     |
|------------------------|---------------------|------------------|---------|-----|----|-----|
|                        |                     |                  | 1       | 2   | 3  | 4   |
| A Ovalbumin            | Chicken Egg White   | 45KD             | 65      | 55  | 15 | 2   |
| B Myoglobin            | Equine Heart        | 17KD             | 55      | 15  | 2  | 65  |
| C Phosphorylase b      | Rabbit Muscle       | 97KD             | 15      | 2   | 65 | 55  |
| D Beta-Galactosidase   | Escherichia Coli    | 116KD            | 2       | 65  | 55 | 15  |
| E Bovine Serum Albumin | Bovine Serum        | 66KD             | 11      | 0.6 | 10 | 500 |
| F Carbonic Anhydrase   | Bovine Erythrocytes | 29KD             | 10      | 500 | 11 | 0.6 |

*Spiked into a constant background: tryptic digests of S. cerevisiae*

Choi et al., “ABRF Proteome Informatics Research Group (iPRG) 2015 Study: Detection of differentially abundant proteins in label-free quantitative LC-MS/MS experiments”, *Journal of Proteome Research*, 2017.

# EXPERIMENTAL PROCEDURES

- **Background**

- ◆ 200ng of tryptic digests of *S. cerevisiae*

- **Spectral acquisition**

- ◆ *Three technical replicates per sample*
- ◆ *Randomized order*
- ◆ Separation:
  - ◆ Thermo nLC 1000 system
  - ◆ 110-min linear gradient
- ◆ Spectral acquisition:
  - ◆ *DDA profile mode in Orbitrap*
  - ◆ Resolution 70,000 for MS and 17,500 for MS/MS
  - ◆ MSI scan range 300-1650 m/z
  - ◆ Normalized collision energy 27%
  - ◆ Singly charged ions excluded

# SPECTRAL PROCESSING

## ● MS/MS identification

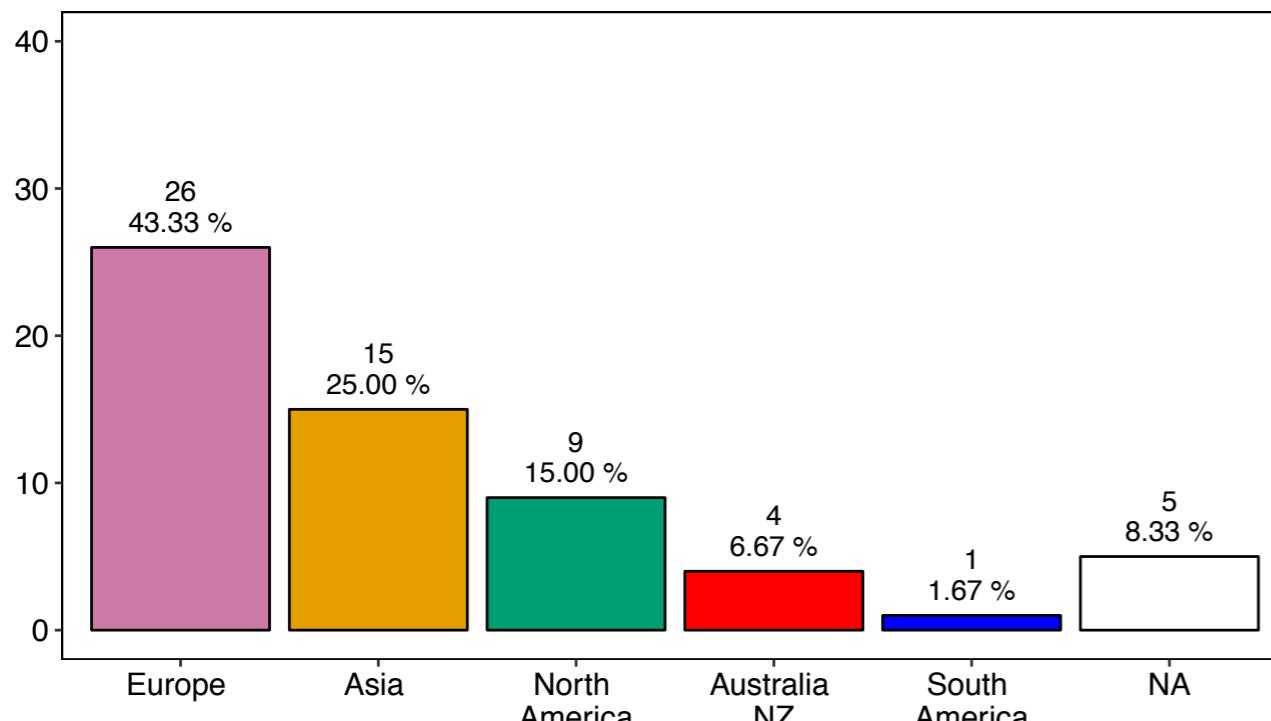
- ◆ Database search: OMSSA7, MS-GF+, Comet
- ◆ Q-value: target-decoy.
  - ◆ No filtering!
- ◆ 5,766 proteins, ~26,242 PSMs/run
- ◆ 48% of proteins had 1 or 2 PSMs
- ◆ 29% of proteins had 15 or more PSMs

## ● MSI quantification with Skyline

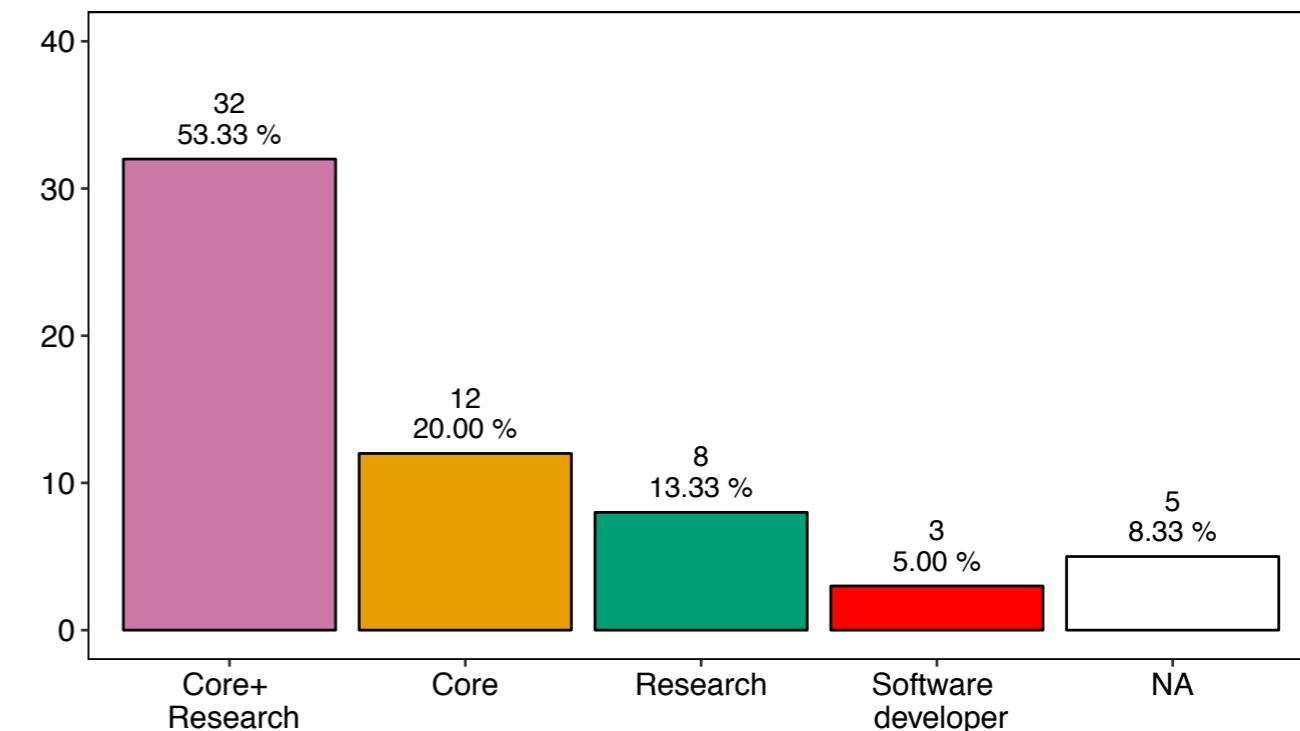
- ◆ Original processing
  - ◆ 3,766 proteins, 29,575 features
    - ◆ median 5 features/protein
- ◆ Post-study processing
  - ◆ 3,027 proteins, 34,783 features
    - ◆ median 7 features/protein

# STUDY PARTICIPANTS

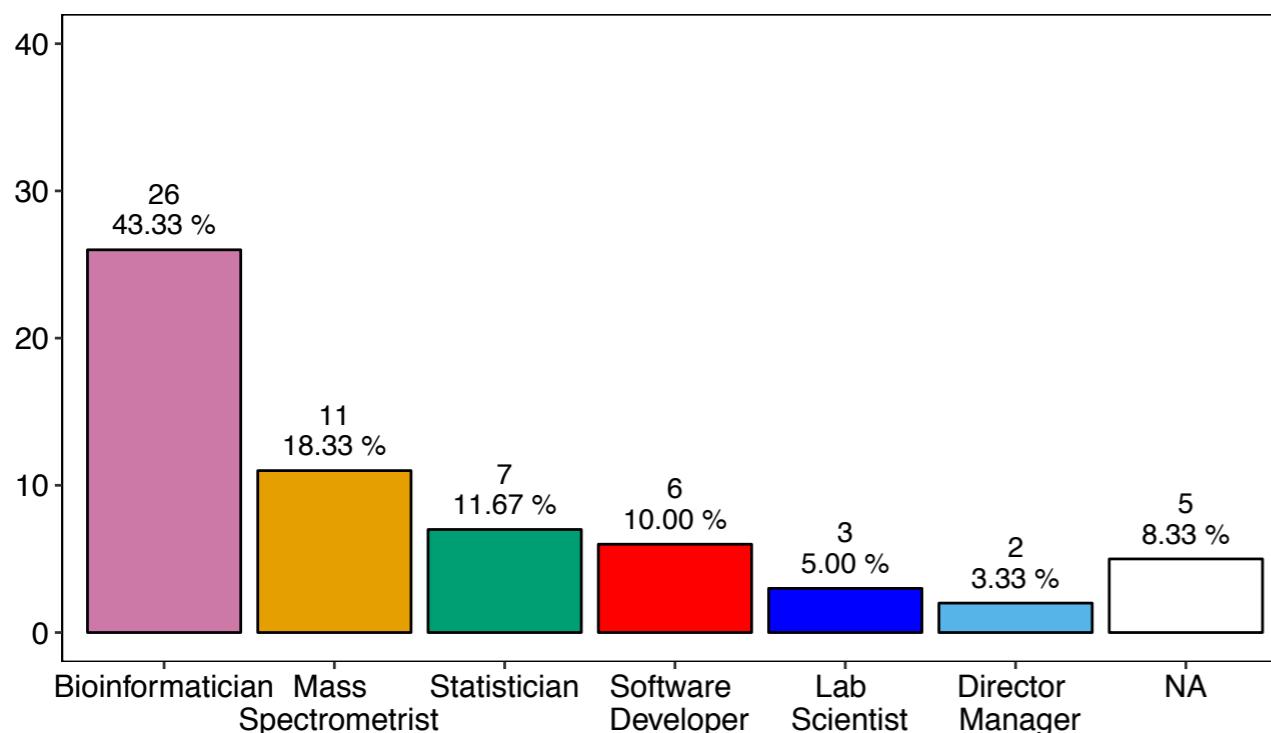
(A) Continent



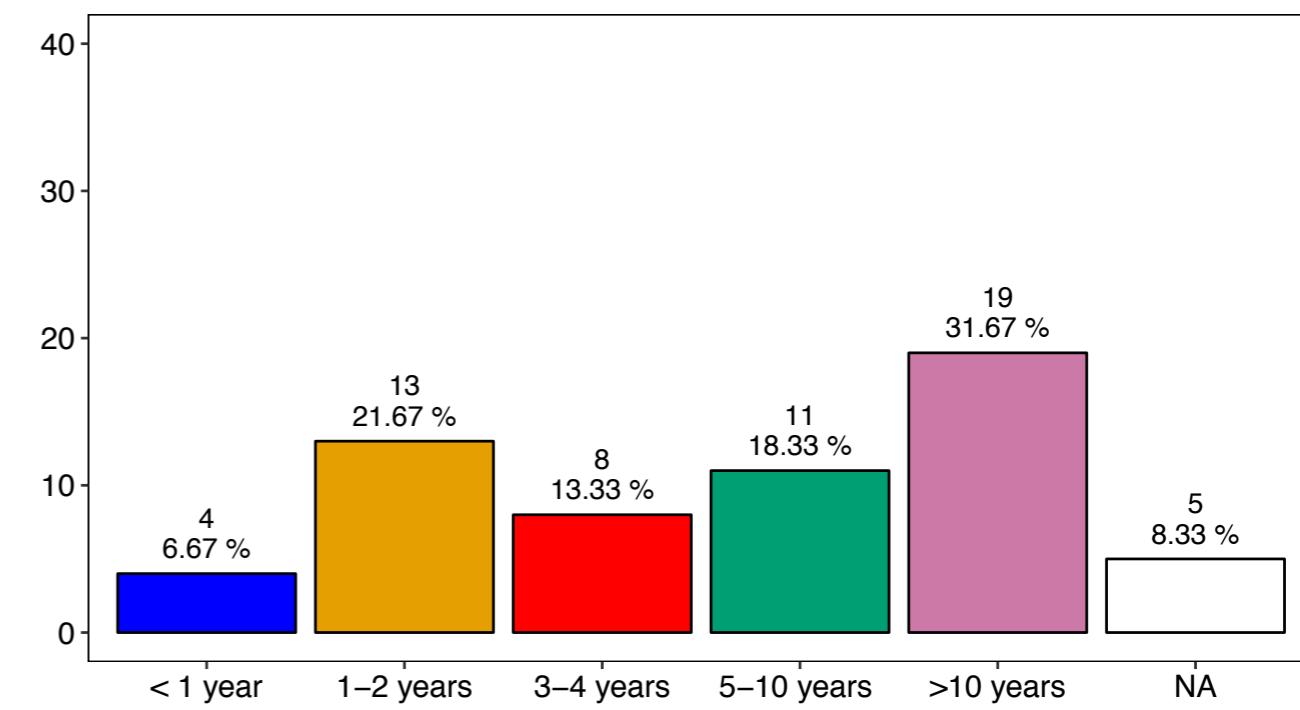
(B) Facility type



(C) Job description

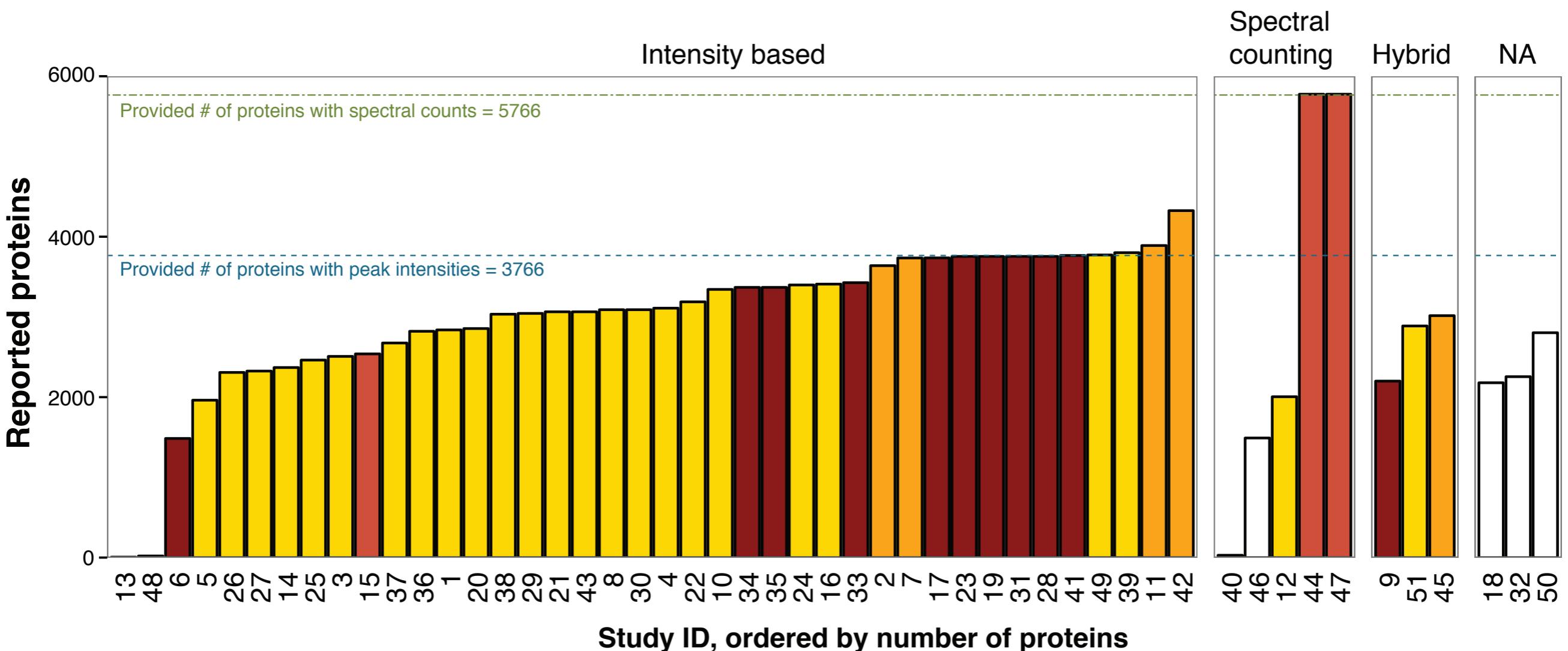
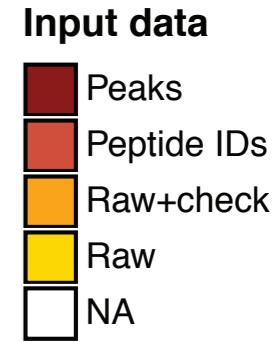


(D) Experience in label-free proteomics



# DIVERSE SUBMISSIONS

*INPUT, PROTEIN NUMBER,  
AND CHOICE OF QUANTIFICATION*

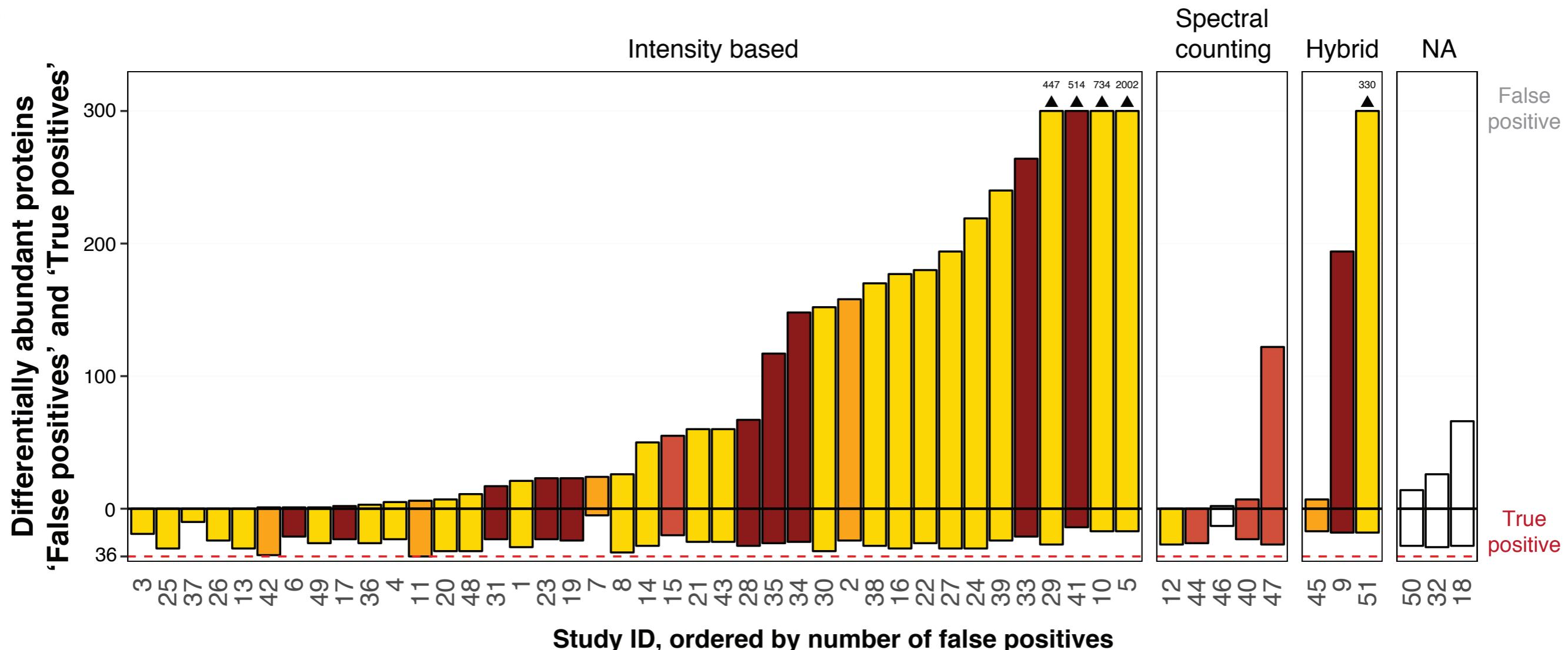


# DIVERSE SUBMISSIONS

## ACCURACY OF DETECTING DIFFERENTIAL ABUNDANCE

**Input data**

- Peaks
- Peptide IDs
- Raw+check
- Raw
- NA

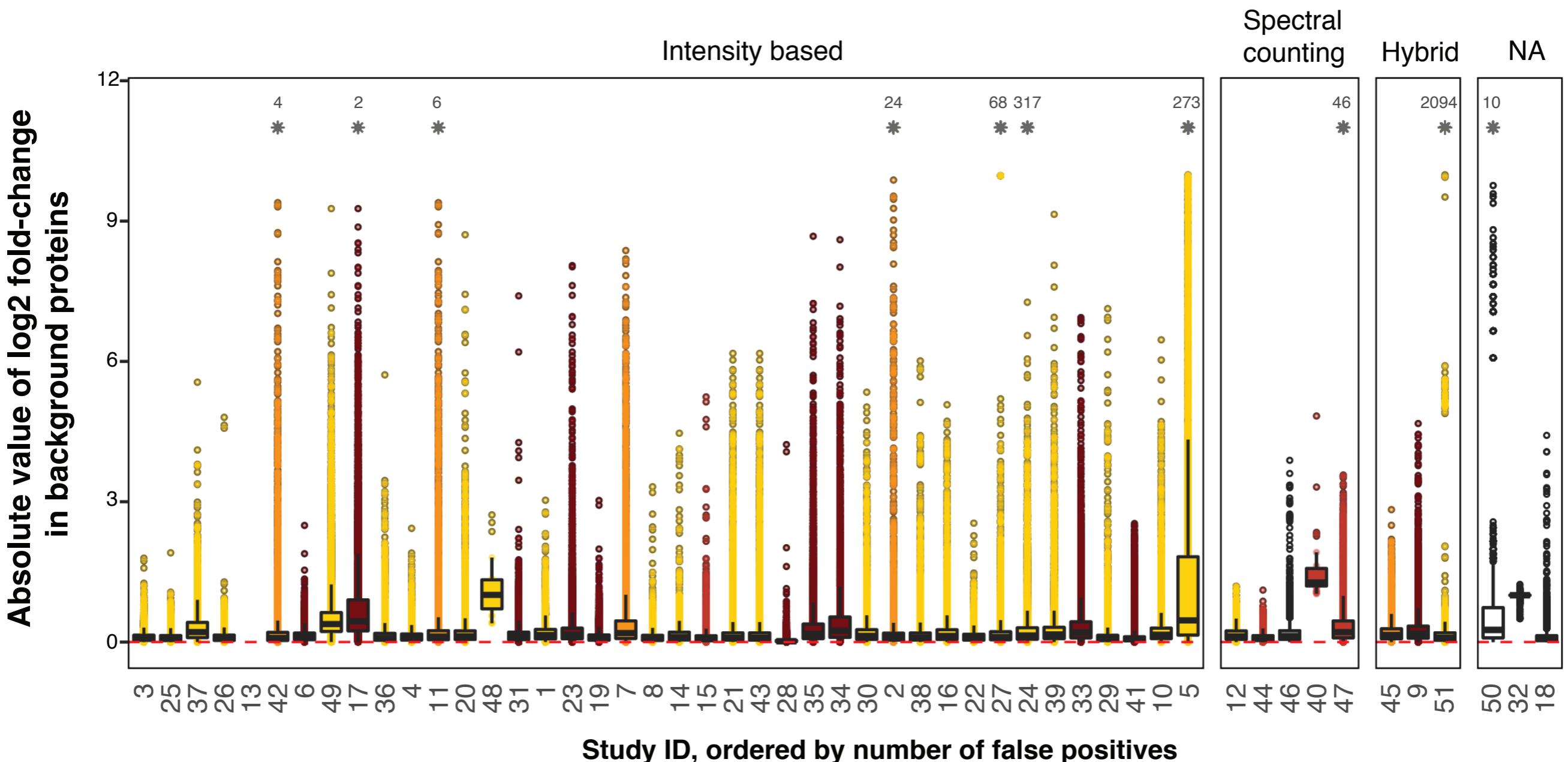


# DIVERSE SUBMISSIONS

## ACCURACY OF ESTIMATING FOLD CHANGE

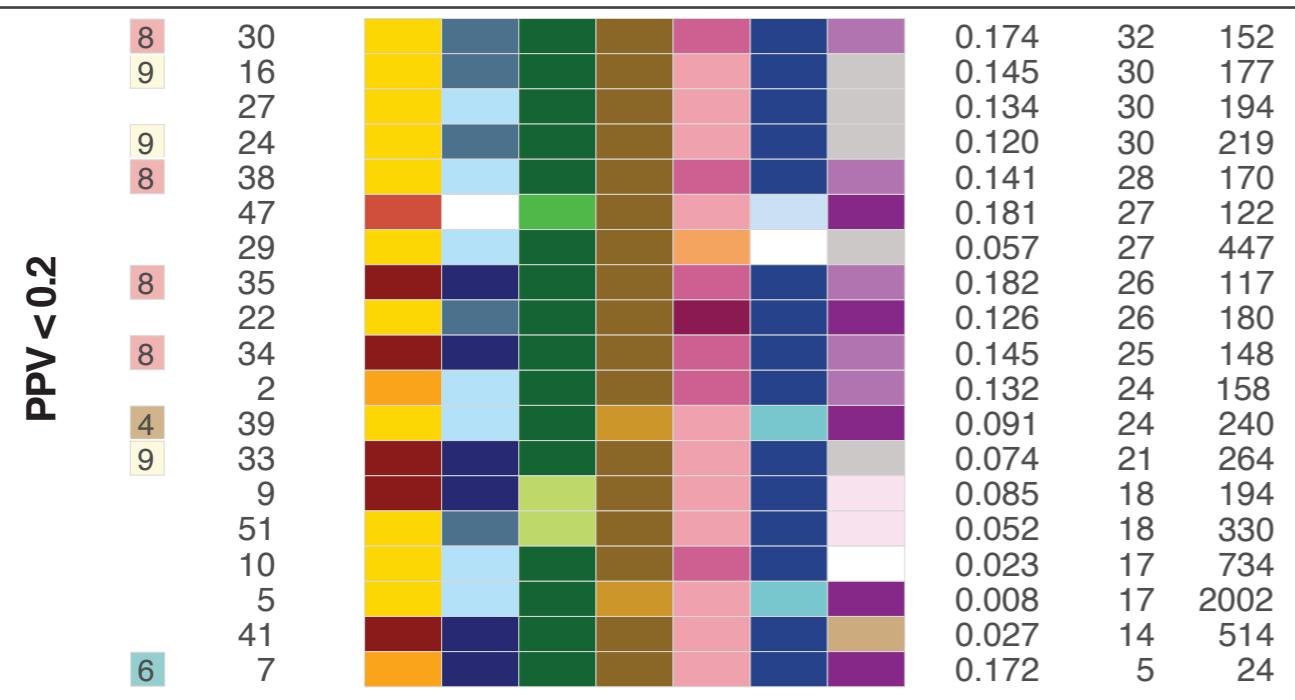
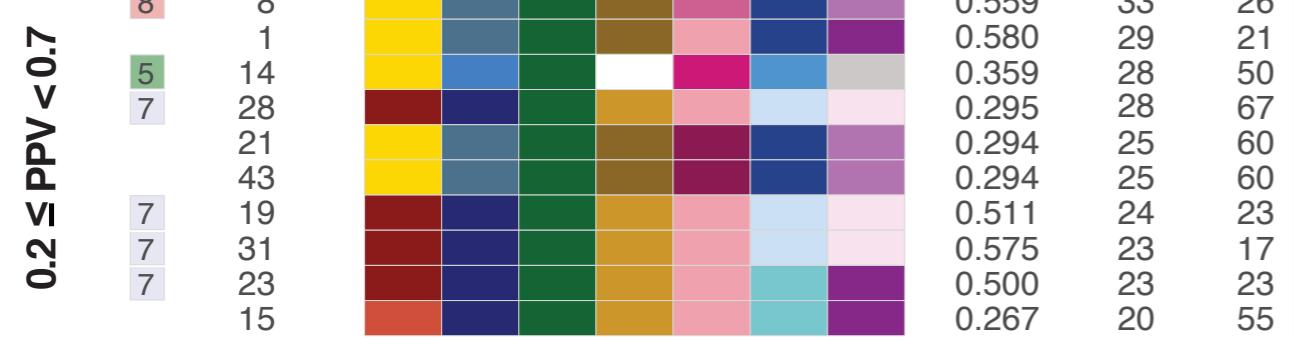
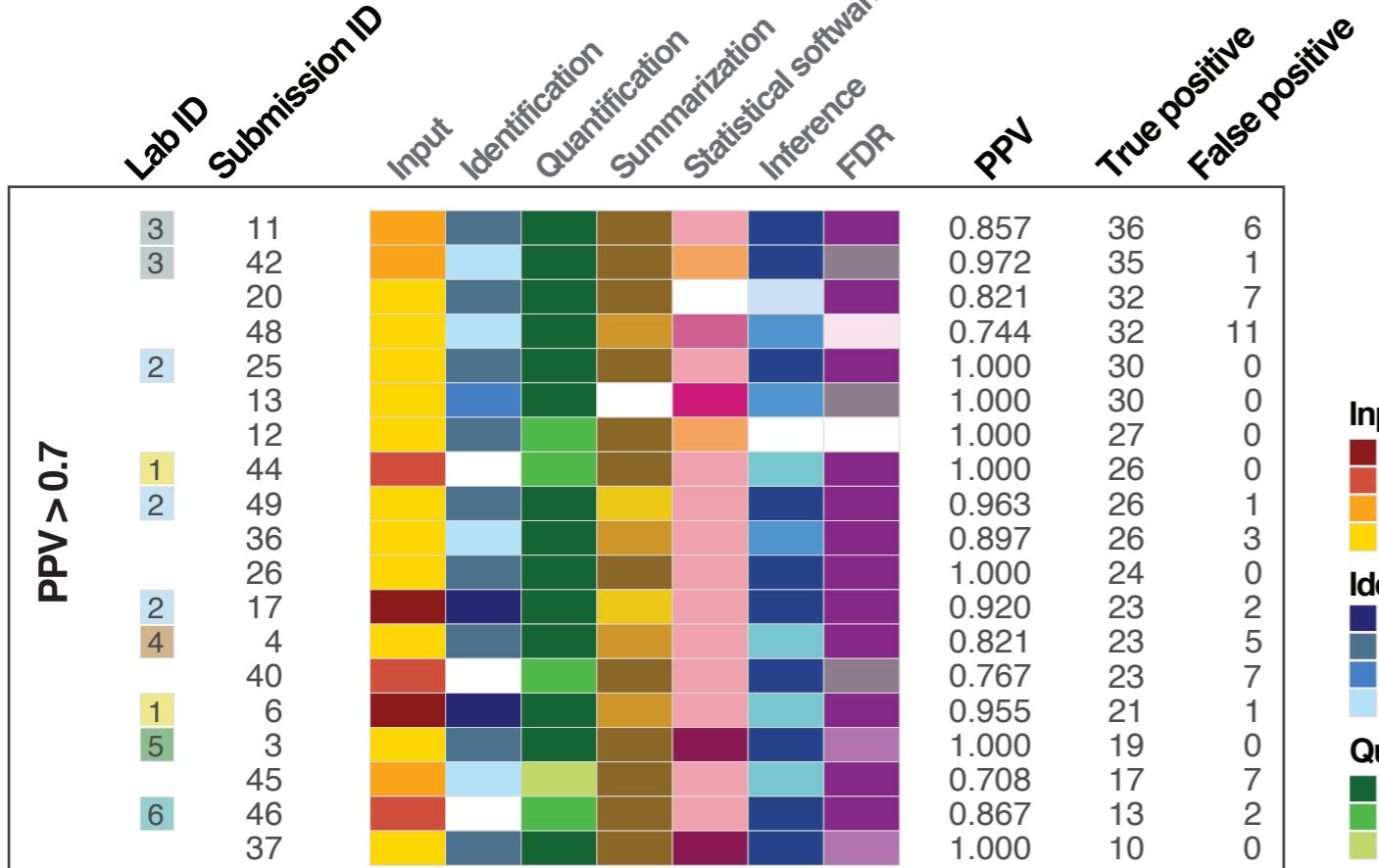
**Input data**

- Peaks
- Peptide IDs
- Raw+check
- Raw
- NA



# SUMMARY OF SUBMISSIONS

USER EXPERTISE IS KEY



## Input

- Peaks
- Peptide ids
- Raw+check
- Raw

## Identification

- Skyline
- MaxQuant
- Progenesis
- Others

## Quantification

- Feature intensity
- Spectral counting
- Hybrid

## Summarization

- Protein summarization / Protein-level inference
- Peptide summarization / Protein-level inference
- Peptide summarization / Peptide-level inference

## Statistical software

- Persus
- Progenesis QI
- Others
- R, Excel, MatLab, Python
- In-house scripts

## Inference

- t-test / SAM's t test
- ANOVA
- Linear (mixed-effects) model
- Others

## FDR

- Benjamini Hochberg
- Permutation FDR
- Others
- Manual validation
- FC cutoff
- No adjustment

No information

# USER EXPERTISE IS KEY

| PPV > 0.7 | Lab ID | Submission ID | Process Flow |                |                |               |                      |           | PPV   | True positive | False positive |
|-----------|--------|---------------|--------------|----------------|----------------|---------------|----------------------|-----------|-------|---------------|----------------|
|           |        |               | Input        | Identification | Quantification | Summarization | Statistical software | Inference |       |               |                |
| 3         | 11     |               |              |                |                |               |                      |           | 0.857 | 36            | 6              |
| 3         | 42     |               |              |                |                |               |                      |           | 0.972 | 35            | 1              |
| 20        | 20     |               |              |                |                |               |                      |           | 0.821 | 32            | 7              |
| 48        | 48     |               |              |                |                |               |                      |           | 0.744 | 32            | 11             |
| 25        | 25     |               |              |                |                |               |                      |           | 1.000 | 30            | 0              |
| 13        | 13     |               |              |                |                |               |                      |           | 1.000 | 30            | 0              |
| 12        | 12     |               |              |                |                |               |                      |           | 1.000 | 27            | 0              |
| 44        | 44     |               |              |                |                |               |                      |           | 1.000 | 26            | 0              |
| 2         | 49     |               |              |                |                |               |                      |           | 0.963 | 26            | 1              |
| 36        | 36     |               |              |                |                |               |                      |           | 0.897 | 26            | 3              |
| 26        | 26     |               |              |                |                |               |                      |           | 1.000 | 24            | 0              |
| 17        | 17     |               |              |                |                |               |                      |           | 0.920 | 23            | 2              |
| 4         | 4      |               |              |                |                |               |                      |           | 0.821 | 23            | 5              |
| 40        | 40     |               |              |                |                |               |                      |           | 0.767 | 23            | 7              |
| 6         | 6      |               |              |                |                |               |                      |           | 0.955 | 21            | 1              |
| 5         | 3      |               |              |                |                |               |                      |           | 1.000 | 19            | 0              |
| 45        | 45     |               |              |                |                |               |                      |           | 0.708 | 17            | 7              |
| 6         | 46     |               |              |                |                |               |                      |           | 0.867 | 13            | 2              |
| 37        | 37     |               |              |                |                |               |                      |           | 1.000 | 10            | 0              |

Positive predictive value =

# true differentially abundant proteins

# claimed differentially abundant proteins

Good

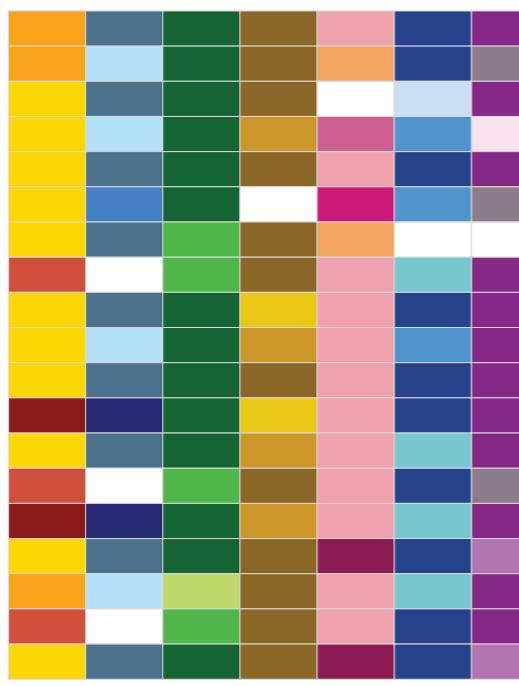
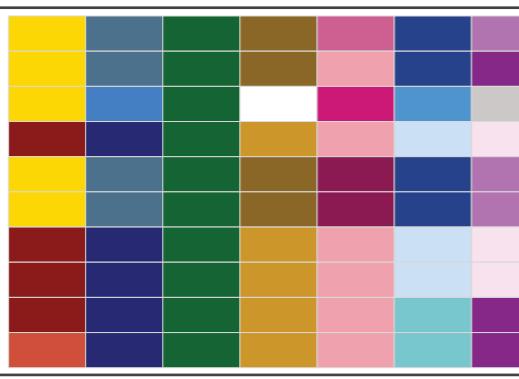
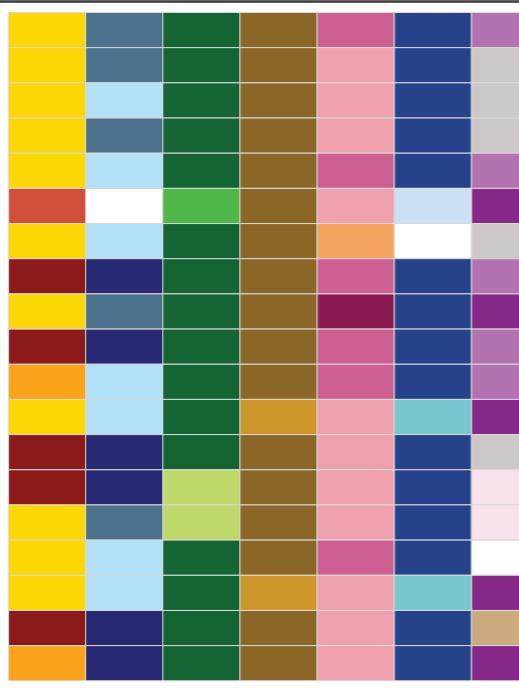
| 0.2 ≤ PPV < 0.7 | Lab ID | Submission ID | Process Flow |                |                |               |                      |           | PPV   | True positive | False positive |
|-----------------|--------|---------------|--------------|----------------|----------------|---------------|----------------------|-----------|-------|---------------|----------------|
|                 |        |               | Input        | Identification | Quantification | Summarization | Statistical software | Inference |       |               |                |
| 8               | 8      |               |              |                |                |               |                      |           | 0.559 | 33            | 26             |
| 1               | 1      |               |              |                |                |               |                      |           | 0.580 | 29            | 21             |
| 5               | 14     |               |              |                |                |               |                      |           | 0.359 | 28            | 50             |
| 7               | 28     |               |              |                |                |               |                      |           | 0.295 | 28            | 67             |
|                 | 21     |               |              |                |                |               |                      |           | 0.294 | 25            | 60             |
|                 | 43     |               |              |                |                |               |                      |           | 0.294 | 25            | 60             |
| 7               | 19     |               |              |                |                |               |                      |           | 0.511 | 24            | 23             |
| 7               | 31     |               |              |                |                |               |                      |           | 0.575 | 23            | 17             |
| 7               | 23     |               |              |                |                |               |                      |           | 0.500 | 23            | 23             |
|                 | 15     |               |              |                |                |               |                      |           | 0.267 | 20            | 55             |

Bad

| PPV < 0.2 | Lab ID | Submission ID | Process Flow |                |                |               |                      |           | PPV   | True positive | False positive |
|-----------|--------|---------------|--------------|----------------|----------------|---------------|----------------------|-----------|-------|---------------|----------------|
|           |        |               | Input        | Identification | Quantification | Summarization | Statistical software | Inference |       |               |                |
| 8         | 30     |               |              |                |                |               |                      |           | 0.174 | 32            | 152            |
| 9         | 16     |               |              |                |                |               |                      |           | 0.145 | 30            | 177            |
|           | 27     |               |              |                |                |               |                      |           | 0.134 | 30            | 194            |
| 9         | 24     |               |              |                |                |               |                      |           | 0.120 | 30            | 219            |
| 8         | 38     |               |              |                |                |               |                      |           | 0.141 | 28            | 170            |
|           | 47     |               |              |                |                |               |                      |           | 0.181 | 27            | 122            |
|           | 29     |               |              |                |                |               |                      |           | 0.057 | 27            | 447            |
| 8         | 35     |               |              |                |                |               |                      |           | 0.182 | 26            | 117            |
|           | 35     |               |              |                |                |               |                      |           | 0.126 | 26            | 180            |
| 8         | 22     |               |              |                |                |               |                      |           | 0.145 | 25            | 148            |
|           | 34     |               |              |                |                |               |                      |           | 0.132 | 24            | 158            |
| 2         | 2      |               |              |                |                |               |                      |           | 0.091 | 24            | 240            |
| 4         | 39     |               |              |                |                |               |                      |           | 0.074 | 21            | 264            |
| 9         | 33     |               |              |                |                |               |                      |           | 0.085 | 18            | 194            |
|           | 51     |               |              |                |                |               |                      |           | 0.052 | 18            | 330            |
| 10        | 10     |               |              |                |                |               |                      |           | 0.023 | 17            | 734            |
| 5         | 5      |               |              |                |                |               |                      |           | 0.008 | 17            | 2002           |
| 6         | 41     |               |              |                |                |               |                      |           | 0.027 | 14            | 514            |
|           | 7      |               |              |                |                |               |                      |           | 0.172 | 5             | 24             |

Very bad

# USER EXPERTISE IS KEY

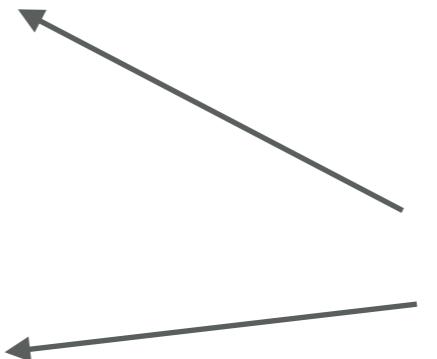
|   | Lab ID | Submission ID | Input | Identification | Quantification | Summarization | Statistical software | Inference | FDR | PPV | True positive | False positive |
|---|--------|---------------|-------|----------------|----------------|---------------|----------------------|-----------|-----|-----|---------------|----------------|
| <b>PPV &gt; 0.7</b><br>        | 3      | 11            | 0.857 | 36             | 6              |               |                      |           |     |     |               |                |
|   | 3      | 42            | 0.972 | 35             | 1              |               |                      |           |     |     |               |                |
|   | 20     | 0.821         | 32    | 7              |                |               |                      |           |     |     |               |                |
|   | 48     | 0.744         | 32    | 11             |                |               |                      |           |     |     |               |                |
| 2   | 25     | 1.000         | 30    | 0              |                |               |                      |           |     |     |               |                |
|   | 13     | 1.000         | 30    | 0              |                |               |                      |           |     |     |               |                |
|   | 12     | 1.000         | 27    | 0              |                |               |                      |           |     |     |               |                |
| 1   | 44     | 1.000         | 26    | 0              |                |               |                      |           |     |     |               |                |
| 2   | 49     | 0.963         | 26    | 1              |                |               |                      |           |     |     |               |                |
|   | 36     | 0.897         | 26    | 3              |                |               |                      |           |     |     |               |                |
|   | 26     | 1.000         | 24    | 0              |                |               |                      |           |     |     |               |                |
| 2   | 17     | 0.920         | 23    | 2              |                |               |                      |           |     |     |               |                |
| 4   | 4      | 0.821         | 23    | 5              |                |               |                      |           |     |     |               |                |
|   | 40     | 0.767         | 23    | 7              |                |               |                      |           |     |     |               |                |
| 1   | 6      | 0.955         | 21    | 1              |                |               |                      |           |     |     |               |                |
| 5   | 3      | 1.000         | 19    | 0              |                |               |                      |           |     |     |               |                |
|   | 45     | 0.708         | 17    | 7              |                |               |                      |           |     |     |               |                |
| 6   | 46     | 0.867         | 13    | 2              |                |               |                      |           |     |     |               |                |
|   | 37     | 1.000         | 10    | 0              |                |               |                      |           |     |     |               |                |
| <b>0.2 ≤ PPV &lt; 0.7</b><br> | 8      | 8             | 0.559 | 33             | 26             |               |                      |           |     |     |               |                |
|   | 1      | 0.580         | 29    | 21             |                |               |                      |           |     |     |               |                |
| 5   | 14     | 0.359         | 28    | 50             |                |               |                      |           |     |     |               |                |
| 7   | 28     | 0.295         | 28    | 67             |                |               |                      |           |     |     |               |                |
|   | 21     | 0.294         | 25    | 60             |                |               |                      |           |     |     |               |                |
|   | 43     | 0.294         | 25    | 60             |                |               |                      |           |     |     |               |                |
| 7   | 19     | 0.511         | 24    | 23             |                |               |                      |           |     |     |               |                |
| 7   | 31     | 0.575         | 23    | 17             |                |               |                      |           |     |     |               |                |
| 7   | 23     | 0.500         | 23    | 23             |                |               |                      |           |     |     |               |                |
|   | 15     | 0.267         | 20    | 55             |                |               |                      |           |     |     |               |                |
| <b>PPV &lt; 0.2</b><br>      | 8      | 30            | 0.174 | 32             | 152            |               |                      |           |     |     |               |                |
| 9   | 16     | 0.145         | 30    | 177            |                |               |                      |           |     |     |               |                |
|   | 27     | 0.134         | 30    | 194            |                |               |                      |           |     |     |               |                |
| 9   | 24     | 0.120         | 30    | 219            |                |               |                      |           |     |     |               |                |
| 8   | 38     | 0.141         | 28    | 170            |                |               |                      |           |     |     |               |                |
|   | 47     | 0.181         | 27    | 122            |                |               |                      |           |     |     |               |                |
|   | 29     | 0.057         | 27    | 447            |                |               |                      |           |     |     |               |                |
| 8   | 35     | 0.182         | 26    | 117            |                |               |                      |           |     |     |               |                |
|   | 22     | 0.126         | 26    | 180            |                |               |                      |           |     |     |               |                |
| 8   | 34     | 0.145         | 25    | 148            |                |               |                      |           |     |     |               |                |
|   | 2      | 0.132         | 24    | 158            |                |               |                      |           |     |     |               |                |
| 4   | 39     | 0.091         | 24    | 240            |                |               |                      |           |     |     |               |                |
| 9   | 33     | 0.074         | 21    | 264            |                |               |                      |           |     |     |               |                |
|   | 9      | 0.085         | 18    | 194            |                |               |                      |           |     |     |               |                |
|   | 51     | 0.052         | 18    | 330            |                |               |                      |           |     |     |               |                |
|   | 10     | 0.023         | 17    | 734            |                |               |                      |           |     |     |               |                |
|   | 5      | 0.008         | 17    | 2002           |                |               |                      |           |     |     |               |                |
| 6   | 41     | 0.027         | 14    | 514            |                |               |                      |           |     |     |               |                |
|   | 7      | 0.172         | 5     | 24             |                |               |                      |           |     |     |               |                |

*MaxQuant and Perseus*

# USER EXPERTISE IS KEY

|                           | Lab ID | Submission ID | Input | Identification | Quantification | Summarization | Statistical software | Inference | FDR | PPV   | True positive | False positive |
|---------------------------|--------|---------------|-------|----------------|----------------|---------------|----------------------|-----------|-----|-------|---------------|----------------|
| <b>PPV &gt; 0.7</b>       |        |               |       |                |                |               |                      |           |     |       |               |                |
| 3                         | 11     |               |       |                |                |               |                      |           |     | 0.857 | 36            | 6              |
| 3                         | 42     |               |       |                |                |               |                      |           |     | 0.972 | 35            | 1              |
| 20                        |        |               |       |                |                |               |                      |           |     | 0.821 | 32            | 7              |
| 48                        |        |               |       |                |                |               |                      |           |     | 0.744 | 32            | 11             |
| 2                         | 25     |               |       |                |                |               |                      |           |     | 1.000 | 30            | 0              |
| 13                        |        |               |       |                |                |               |                      |           |     | 1.000 | 30            | 0              |
| 12                        |        |               |       |                |                |               |                      |           |     | 1.000 | 27            | 0              |
| 1                         | 44     |               |       |                |                |               |                      |           |     | 1.000 | 26            | 0              |
| 2                         | 49     |               |       |                |                |               |                      |           |     | 0.963 | 26            | 1              |
| 36                        |        |               |       |                |                |               |                      |           |     | 0.897 | 26            | 3              |
| 2                         | 26     |               |       |                |                |               |                      |           |     | 1.000 | 24            | 0              |
| 4                         | 17     |               |       |                |                |               |                      |           |     | 0.920 | 23            | 2              |
| 4                         | 4      |               |       |                |                |               |                      |           |     | 0.821 | 23            | 5              |
| 40                        |        |               |       |                |                |               |                      |           |     | 0.767 | 23            | 7              |
| 1                         | 6      |               |       |                |                |               |                      |           |     | 0.955 | 21            | 1              |
| 5                         | 3      |               |       |                |                |               |                      |           |     | 1.000 | 19            | 0              |
| 6                         | 45     |               |       |                |                |               |                      |           |     | 0.708 | 17            | 7              |
| 6                         | 46     |               |       |                |                |               |                      |           |     | 0.867 | 13            | 2              |
| 6                         | 37     |               |       |                |                |               |                      |           |     | 1.000 | 10            | 0              |
| <b>0.2 ≤ PPV &lt; 0.7</b> |        |               |       |                |                |               |                      |           |     |       |               |                |
| 8                         | 8      |               |       |                |                |               |                      |           |     | 0.559 | 33            | 26             |
|                           | 1      |               |       |                |                |               |                      |           |     | 0.580 | 29            | 21             |
| 5                         | 14     |               |       |                |                |               |                      |           |     | 0.359 | 28            | 50             |
| 7                         | 28     |               |       |                |                |               |                      |           |     | 0.295 | 28            | 67             |
|                           | 21     |               |       |                |                |               |                      |           |     | 0.294 | 25            | 60             |
| 7                         | 43     |               |       |                |                |               |                      |           |     | 0.294 | 25            | 60             |
| 7                         | 19     |               |       |                |                |               |                      |           |     | 0.511 | 24            | 23             |
| 7                         | 31     |               |       |                |                |               |                      |           |     | 0.575 | 23            | 17             |
| 7                         | 23     |               |       |                |                |               |                      |           |     | 0.500 | 23            | 23             |
|                           | 15     |               |       |                |                |               |                      |           |     | 0.267 | 20            | 55             |
| <b>PPV &lt; 0.2</b>       |        |               |       |                |                |               |                      |           |     |       |               |                |
| 8                         | 30     |               |       |                |                |               |                      |           |     | 0.174 | 32            | 152            |
| 9                         | 16     |               |       |                |                |               |                      |           |     | 0.145 | 30            | 177            |
|                           | 27     |               |       |                |                |               |                      |           |     | 0.134 | 30            | 194            |
| 9                         | 24     |               |       |                |                |               |                      |           |     | 0.120 | 30            | 219            |
| 8                         | 38     |               |       |                |                |               |                      |           |     | 0.141 | 28            | 170            |
|                           | 47     |               |       |                |                |               |                      |           |     | 0.181 | 27            | 122            |
| 8                         | 29     |               |       |                |                |               |                      |           |     | 0.057 | 27            | 447            |
| 8                         | 35     |               |       |                |                |               |                      |           |     | 0.182 | 26            | 117            |
| 8                         | 22     |               |       |                |                |               |                      |           |     | 0.126 | 26            | 180            |
| 8                         | 34     |               |       |                |                |               |                      |           |     | 0.145 | 25            | 148            |
|                           | 2      |               |       |                |                |               |                      |           |     | 0.132 | 24            | 158            |
| 4                         | 39     |               |       |                |                |               |                      |           |     | 0.091 | 24            | 240            |
| 9                         | 33     |               |       |                |                |               |                      |           |     | 0.074 | 21            | 264            |
|                           | 9      |               |       |                |                |               |                      |           |     | 0.085 | 18            | 194            |
| 51                        |        |               |       |                |                |               |                      |           |     | 0.052 | 18            | 330            |
| 10                        |        |               |       |                |                |               |                      |           |     | 0.023 | 17            | 734            |
| 5                         |        |               |       |                |                |               |                      |           |     | 0.008 | 17            | 2002           |
| 6                         | 41     |               |       |                |                |               |                      |           |     | 0.027 | 14            | 514            |
| 6                         | 7      |               |       |                |                |               |                      |           |     | 0.172 | 5             | 24             |

*Skyline and linear modeling in R*



# USER EXPERTISE IS KEY

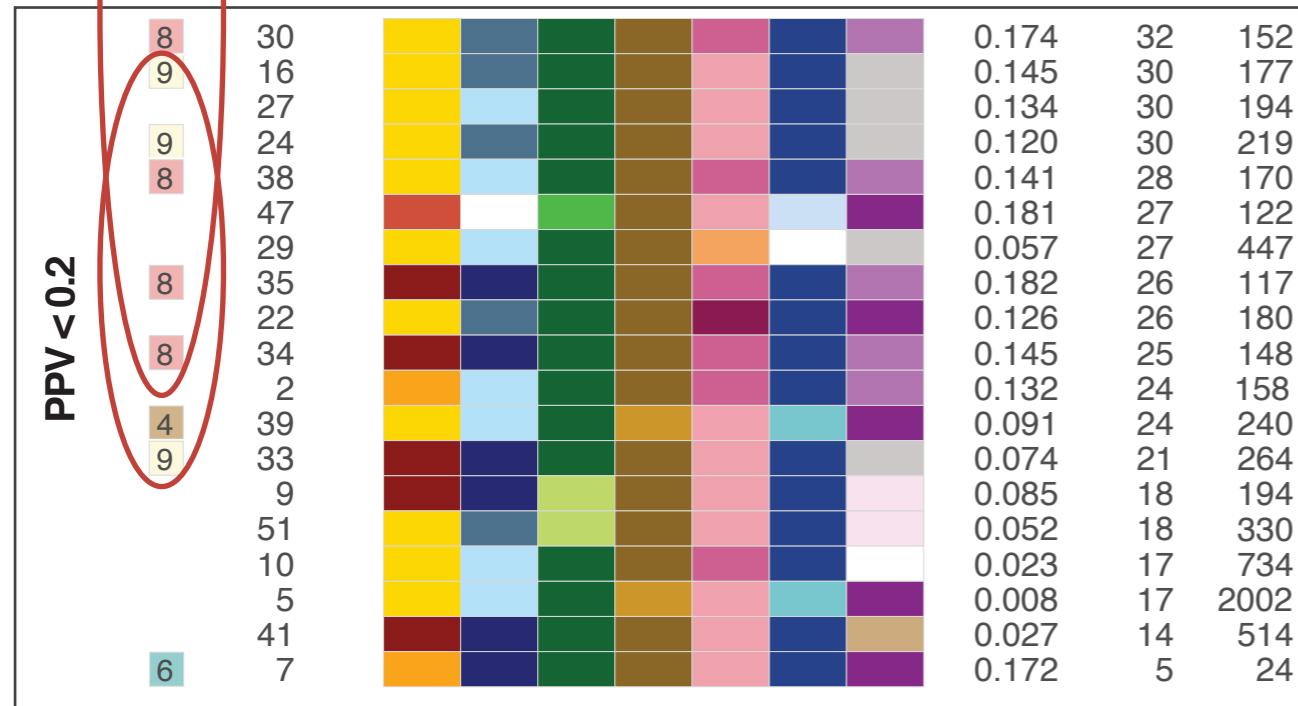
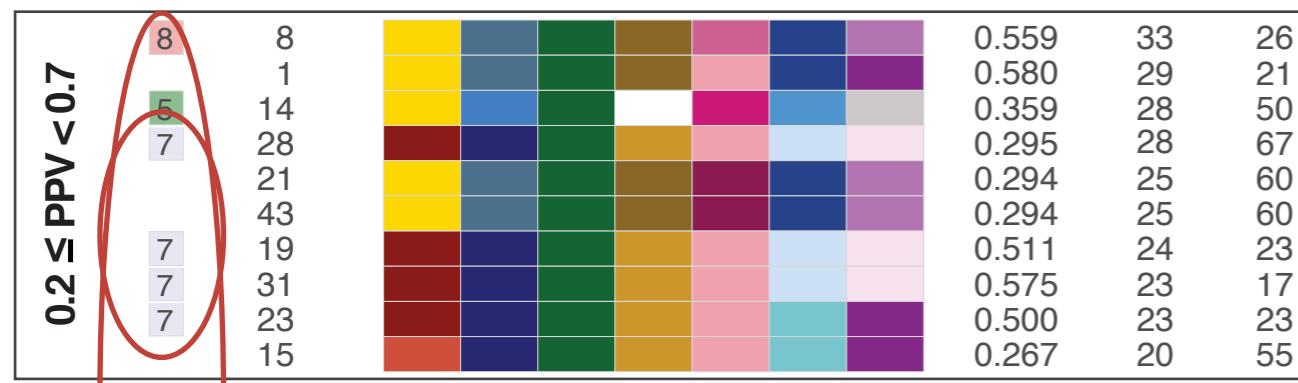
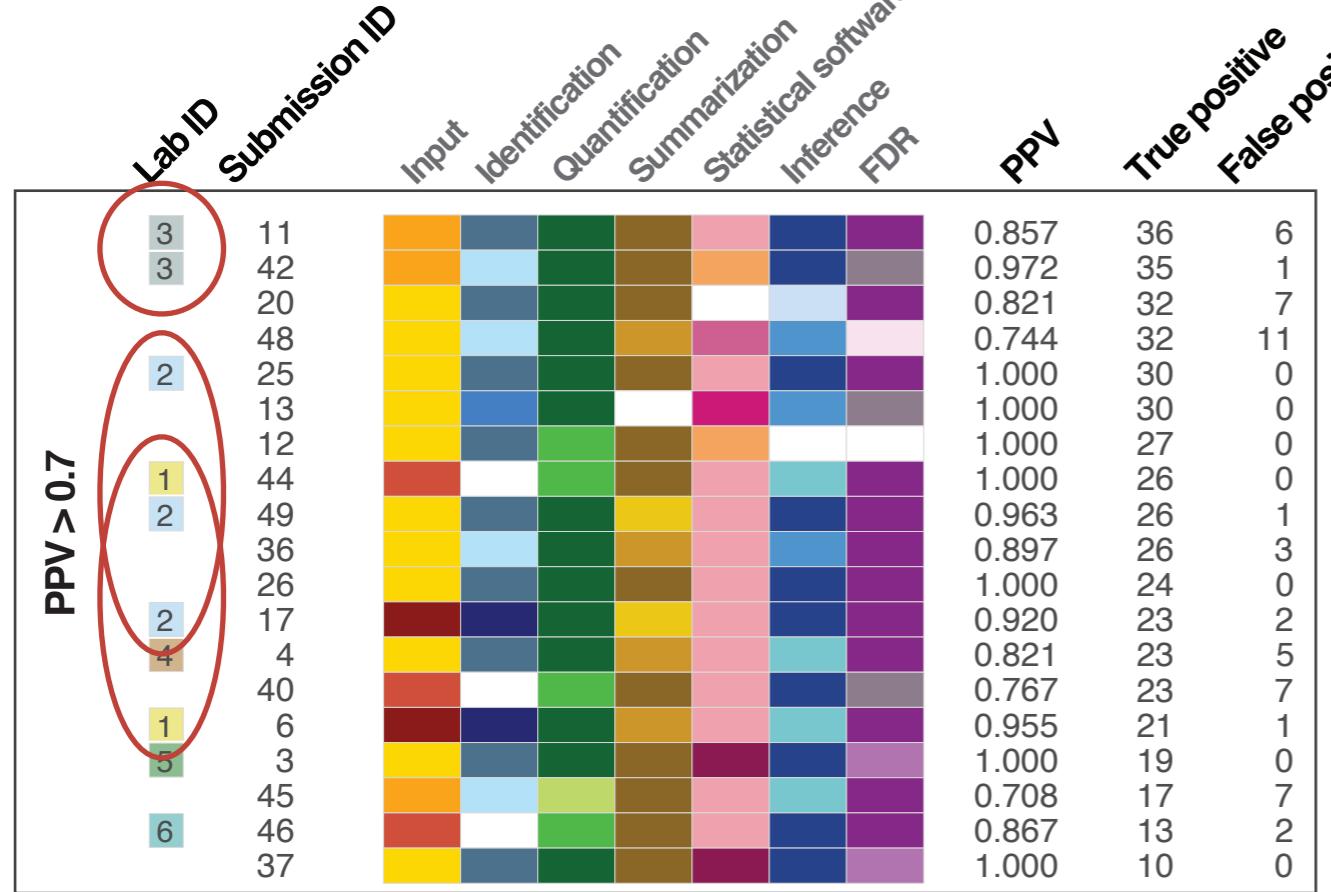
| Lab ID | Submission ID | Input | Identification | Quantification | Summarization | Statistical software | Inference | FDR | PPV   | True positive | False positive |
|--------|---------------|-------|----------------|----------------|---------------|----------------------|-----------|-----|-------|---------------|----------------|
| 3      | 11            |       |                |                |               |                      |           |     | 0.857 | 36            | 6              |
| 3      | 42            |       |                |                |               |                      |           |     | 0.972 | 35            | 1              |
| 20     | 20            |       |                |                |               |                      |           |     | 0.821 | 32            | 7              |
| 48     | 48            |       |                |                |               |                      |           |     | 0.744 | 32            | 11             |
| 2      | 25            |       |                |                |               |                      |           |     | 1.000 | 30            | 0              |
| 13     | 13            |       |                |                |               |                      |           |     | 1.000 | 30            | 0              |
| 12     | 12            |       |                |                |               |                      |           |     | 1.000 | 27            | 0              |
| 1      | 44            |       |                |                |               |                      |           |     | 1.000 | 26            | 0              |
| 2      | 49            |       |                |                |               |                      |           |     | 0.963 | 26            | 1              |
| 36     | 36            |       |                |                |               |                      |           |     | 0.897 | 26            | 3              |
| 2      | 26            |       |                |                |               |                      |           |     | 1.000 | 24            | 0              |
| 4      | 17            |       |                |                |               |                      |           |     | 0.920 | 23            | 2              |
| 4      | 4             |       |                |                |               |                      |           |     | 0.821 | 23            | 5              |
| 40     | 40            |       |                |                |               |                      |           |     | 0.767 | 23            | 7              |
| 1      | 6             |       |                |                |               |                      |           |     | 0.955 | 21            | 1              |
| 5      | 3             |       |                |                |               |                      |           |     | 1.000 | 19            | 0              |
| 45     | 45            |       |                |                |               |                      |           |     | 0.708 | 17            | 7              |
| 6      | 46            |       |                |                |               |                      |           |     | 0.867 | 13            | 2              |
|        | 37            |       |                |                |               |                      |           |     | 1.000 | 10            | 0              |

| PPV > 0.7 | PPV ≤ 0.7 | PPV < 0.2 |
|-----------|-----------|-----------|
| 8         | 8         | 30        |
| 1         | 1         | 16        |
| 5         | 14        | 27        |
| 7         | 28        | 24        |
|           | 21        | 38        |
|           | 43        | 47        |
| 7         | 19        | 29        |
| 7         | 31        | 35        |
| 7         | 23        | 35        |
|           | 15        | 22        |

| PPV < 0.2 | PPV > 0.7 | PPV ≤ 0.7 |
|-----------|-----------|-----------|
| 8         | 30        | 8         |
| 9         | 16        | 16        |
|           | 27        | 27        |
| 9         | 24        | 24        |
| 8         | 38        | 38        |
|           | 47        | 47        |
| 8         | 29        | 29        |
| 8         | 35        | 35        |
| 8         | 35        | 35        |
| 8         | 22        | 22        |
| 8         | 34        | 34        |
|           | 2         | 2         |
| 4         | 39        | 39        |
| 9         | 33        | 33        |
|           | 9         | 9         |
| 51        | 51        | 51        |
| 10        | 10        | 10        |
| 5         | 5         | 5         |
| 41        | 41        | 41        |
| 6         | 7         | 7         |

*Compared peak intensity vs spectral counts*

# USER EXPERTISE IS KEY



## Input

- Peaks
- Peptide ids
- Raw+check
- Raw

## Identification

- Skyline
- MaxQuant
- Progenesis
- Others

## Quantification

- Feature intensity
- Spectral counting
- Hybrid

## Summarization

- Protein summarization / Protein-level inference
- Peptide summarization / Protein-level inference
- Peptide summarization / Peptide-level inference

## Statistical software

- Persus
- Progenesis QI
- Others
- R, Excel, MatLab, Python
- In-house scripts

## Inference

- t-test / SAM's t test
- ANOVA
- Linear (mixed-effects) model
- Others

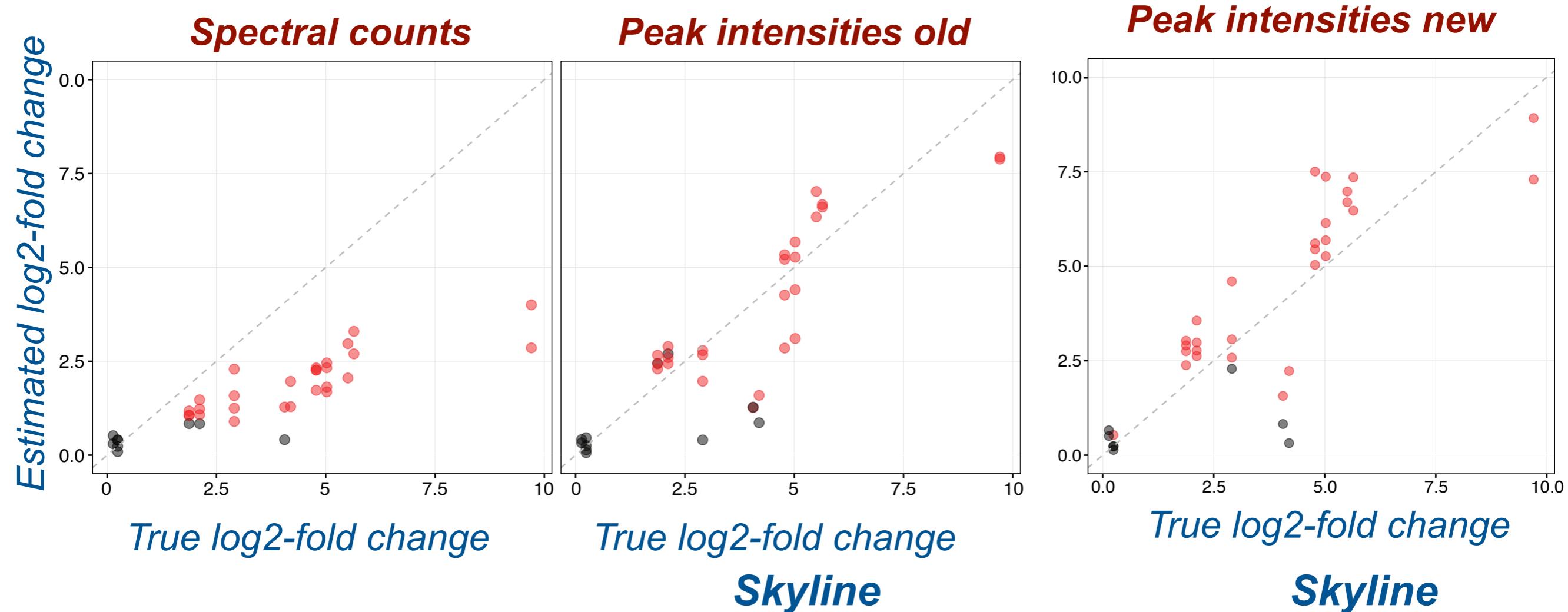
## FDR

- Benjamini Hochberg
- Permutation FDR
- Others
- Manual validation
- FC cutoff
- No adjustment

No information

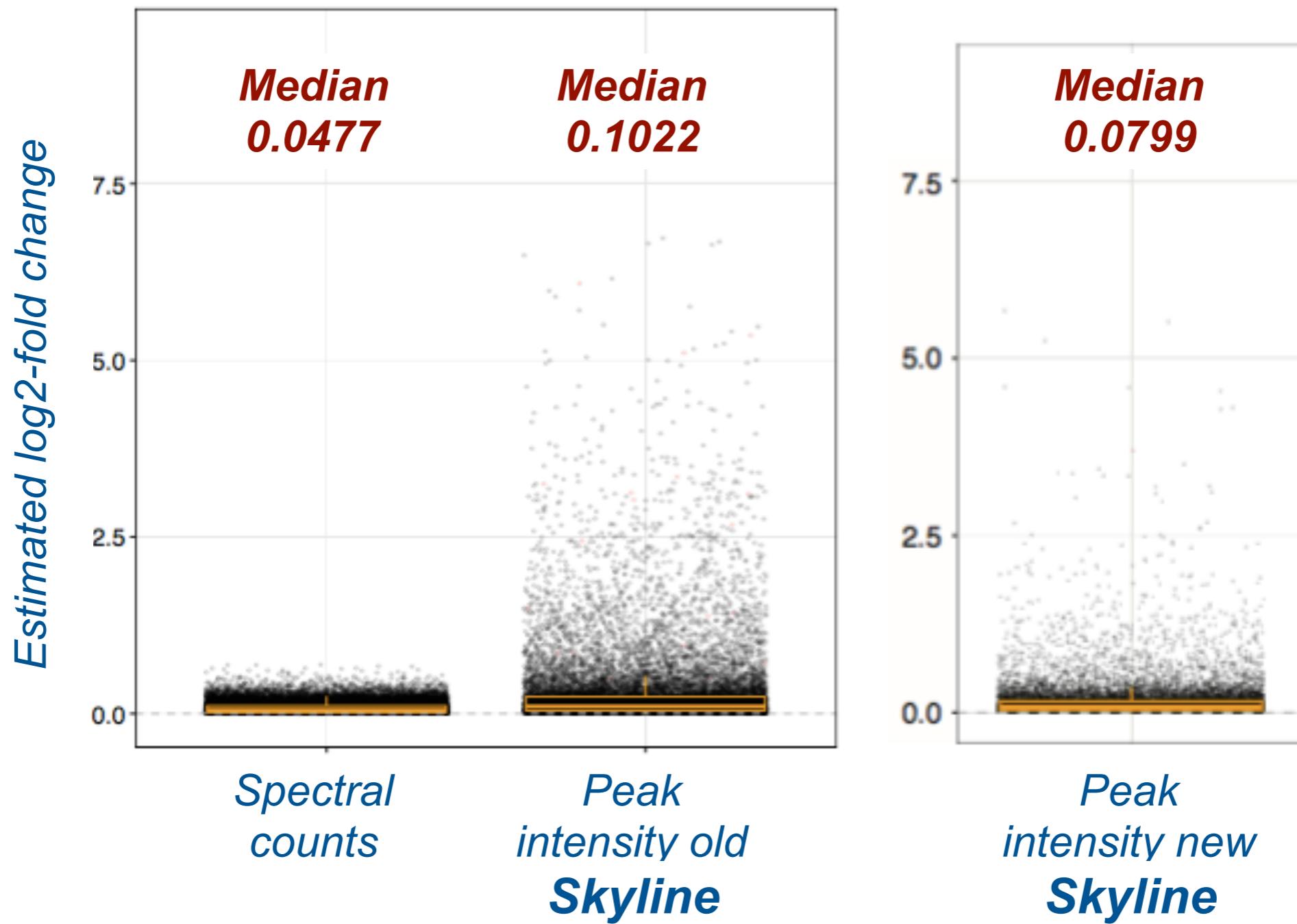
# DATA PROCESSING MATTERS

## IMPROVES ACCURACY OF QUANTIFICATION

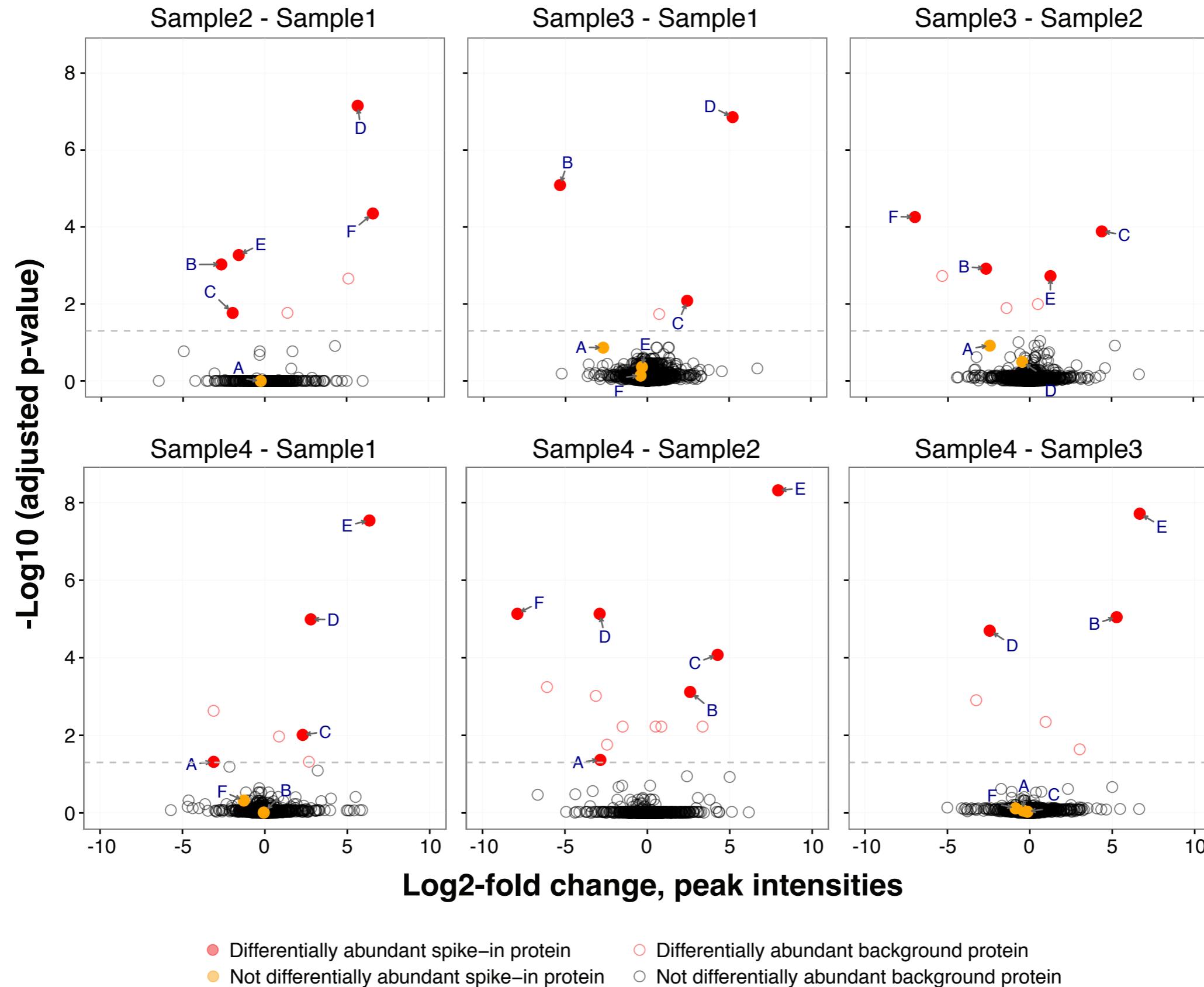


# DATA PROCESSING MATTERS

## IMPROVES ACCURACY OF QUANTIFICATION



# DETECTION OF DIFFERENTIALLY ABUNDANT PROTEINS



# CHOICES OF SIGNAL PROCESSING AFFECT THE RESULTS

## Statistical analysis with MSstats, using varying inputs

- No single-hit proteins
- Sum 3 isotopic peaks

|                           | ID QUALITY (Q-VALUE CUTOFF) |             |             |             |
|---------------------------|-----------------------------|-------------|-------------|-------------|
|                           | 0.05                        | 0.15        | 0.5         | 0.95        |
| True positives            | 29                          | 28          | 28          | 28          |
| Total positives           | 34                          | 29          | 31          | 29          |
| Positive predictive value | 0.852941176                 | 0.965517241 | 0.903225806 | 0.965517241 |

- No single-hit proteins
- Monoisotopic peak

|                           | ID QUALITY (Q-VALUE CUTOFF) |             |             |             |
|---------------------------|-----------------------------|-------------|-------------|-------------|
|                           | 0.05                        | 0.15        | 0.5         | 0.95        |
| True positives            | 28                          | 29          | 29          | 30          |
| Total positives           | 42                          | 34          | 37          | 35          |
| Positive predictive value | 0.666666667                 | 0.852941176 | 0.783783784 | 0.857142857 |

- Single-hit proteins
- Sum 3 isotopic peaks

|                           | ID QUALITY (Q-VALUE CUTOFF) |             |
|---------------------------|-----------------------------|-------------|
|                           | 0.15                        | 0.95        |
| True positives            | 29                          | 29          |
| Total positives           | 56                          | 53          |
| Positive predictive value | 0.517857143                 | 0.547169811 |

Positive predictive value = 
$$\frac{\text{# true differentially abundant proteins}}{\text{# claimed differentially abundant proteins}}$$

## Article

[!\[\]\(647e44ea77c89a016b9d36ad68afc84b\_img.jpg\) Previous Article](#)   [\*\*Next Article >\*\*](#)   [Table of Contents](#)

### ABRF Proteome Informatics Research Group (iPRG) 2015 Study: Detection of Differentially Abundant Proteins in Label-Free Quantitative LC–MS/MS Experiments



Meena Choi<sup>#†</sup> , Zeynep F. Eren-Dogru<sup>#‡</sup>, Christopher Colangelo<sup>§</sup>, John Cottrell<sup>¶</sup>, Michael R. Hoopmann<sup>¶</sup>, Eugene A. Kapp<sup>¶</sup>, Sangtae Kim<sup>§</sup>, Henry Lam<sup>¶</sup>, Thomas A. Neubert<sup>¶</sup>, Magnus Palmblad<sup>¶</sup>, Brett S. Phinney<sup>\*</sup>, Susan T. Weintraub<sup>△</sup>, Brendan MacLean<sup>▲</sup>, and Olga Vitek<sup>\*†</sup> 

<sup>#</sup> Northeastern University, Boston, Massachusetts 02115, United States

<sup>†</sup> Mugla Silki Kocman University, 48000 Mugla, Turkey

<sup>‡</sup> Primary Ion, LLC, Old Lyme, Connecticut 06371, United States

<sup>§</sup> Matrix Science Ltd., London W1U 7GB, U.K.

<sup>¶</sup> Institute for Systems Biology, Seattle, Washington 98109, United States

<sup>¶</sup> Walter and Eliza Hall Institute of Medical Research, Melbourne 3052, Australia

<sup>\*</sup> Pacific Northwest National Laboratory, Richland, Washington 99354, United States

<sup>▲</sup> Department of Chemical and Biomolecular Engineering and Division of Biomedical Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong

<sup>■</sup> Skirball Institute and Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, New York 10016, United States

<sup>○</sup> Center for Proteomics and Metabolomics, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands

<sup>▪</sup> University of California at Davis, Davis, California 95616, United States

<sup>△</sup> University of Texas Health Science Center at San Antonio, San Antonio, Texas 78229, United States

<sup>▲</sup> University of Washington, Seattle, Washington 98105, United States

*J. Proteome Res.*, 2017, 16 (2), pp 945–957

DOI: 10.1021/acs.jproteome.6b00881

Publication Date (Web): December 19, 2016

Copyright © 2016 American Chemical Society

\*E-mail: o.vitek@neu.edu. Tel: 617-370-2194.

## Article Options

 ACS ActiveView PDF

Hi Res Print, Annotate, Rotate/Zoom QuickView

[Abstract](#)

[Supporting Info](#)

 PDF (3135 KB)

[Figures](#)

 PDF w/ Links (885 KB)

[References](#)

 Full Text HTML

 Add to ACS ChemWorx

 Add to Favorites

 Download Citation

 Email a Colleague

 Order Reprints

 Rights & Permissions