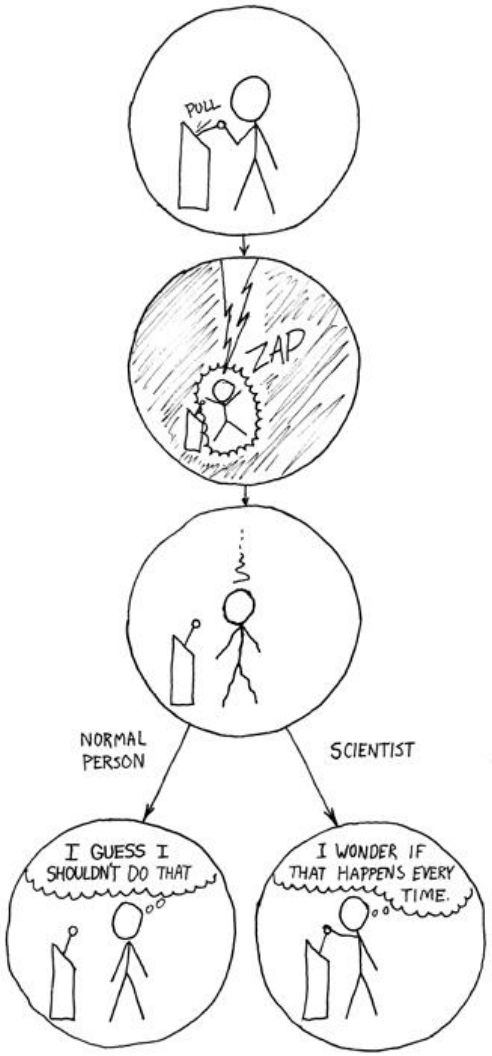
Study Design and Statistical Considerations

2019 Cape Town Targeted Quantitative Proteomics Course

Lindsay Pino

Lindsay.Pino@Pennmedicine.upenn.edu | www.lindsaykpino.com | @lkpino

Replication



xkcd.com/242

Objectives

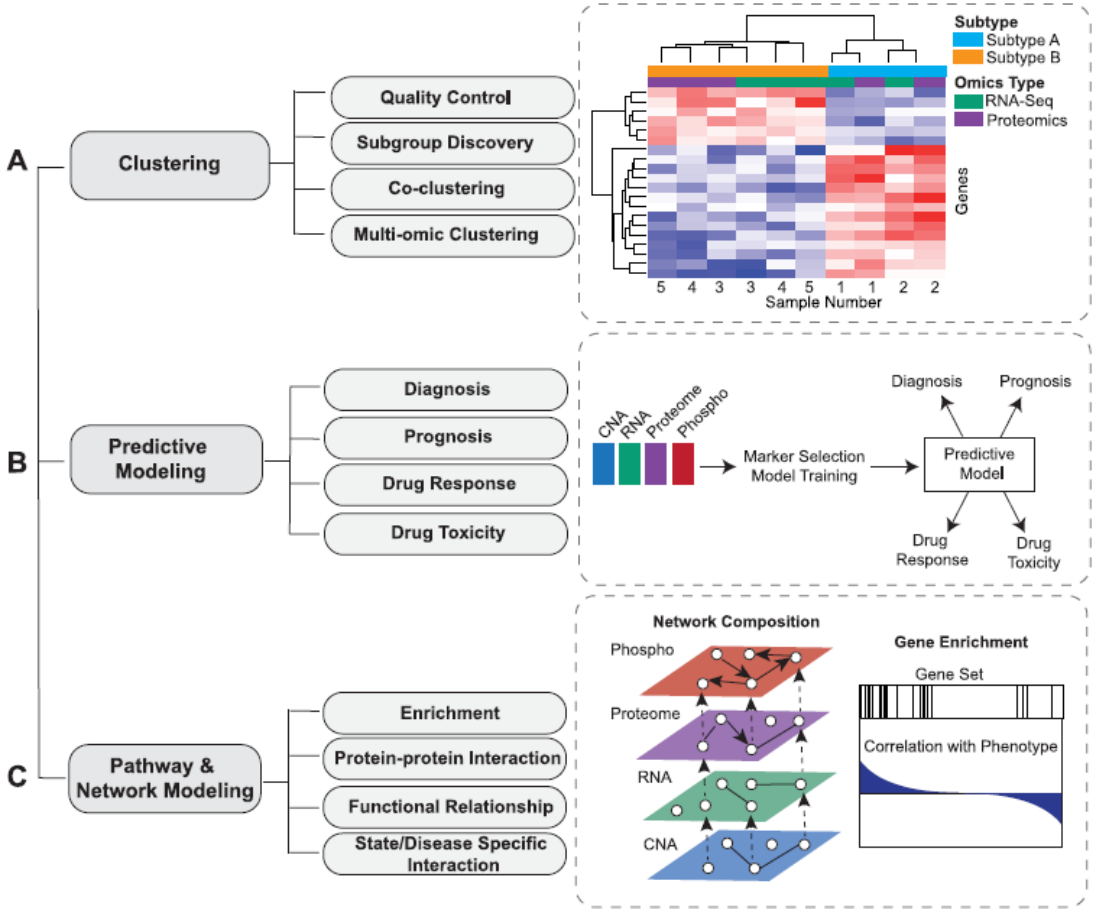
- Relate **experimental design methods** to LC-MS experiments
- Assess the **impact of variation on statistical analyses** and interpretations
- Build a **conceptual foundation** for the data processing and statistical tools contained in MSstats

Agenda

- Translating research experiments into statistical models
- Elements of good statistical/experimental design
  - Replication
  - Randomization
  - Blocking
  - Variation
- Differential abundance tests: the t-test and the ANOVA
- Statistical power: how many samples is enough and the problem with blanket fold change cut-offs

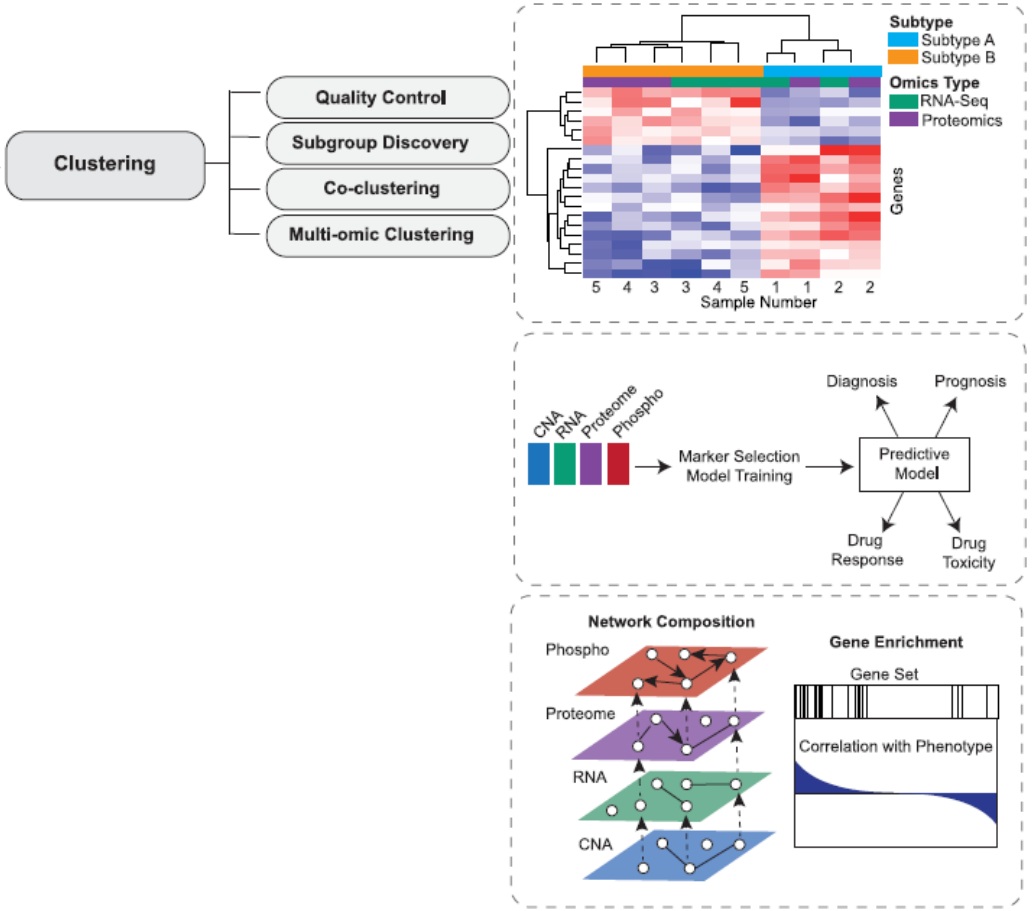
Yesterday: Group Comparisons natively within Skyline

General analysis approaches for proteomics experiments



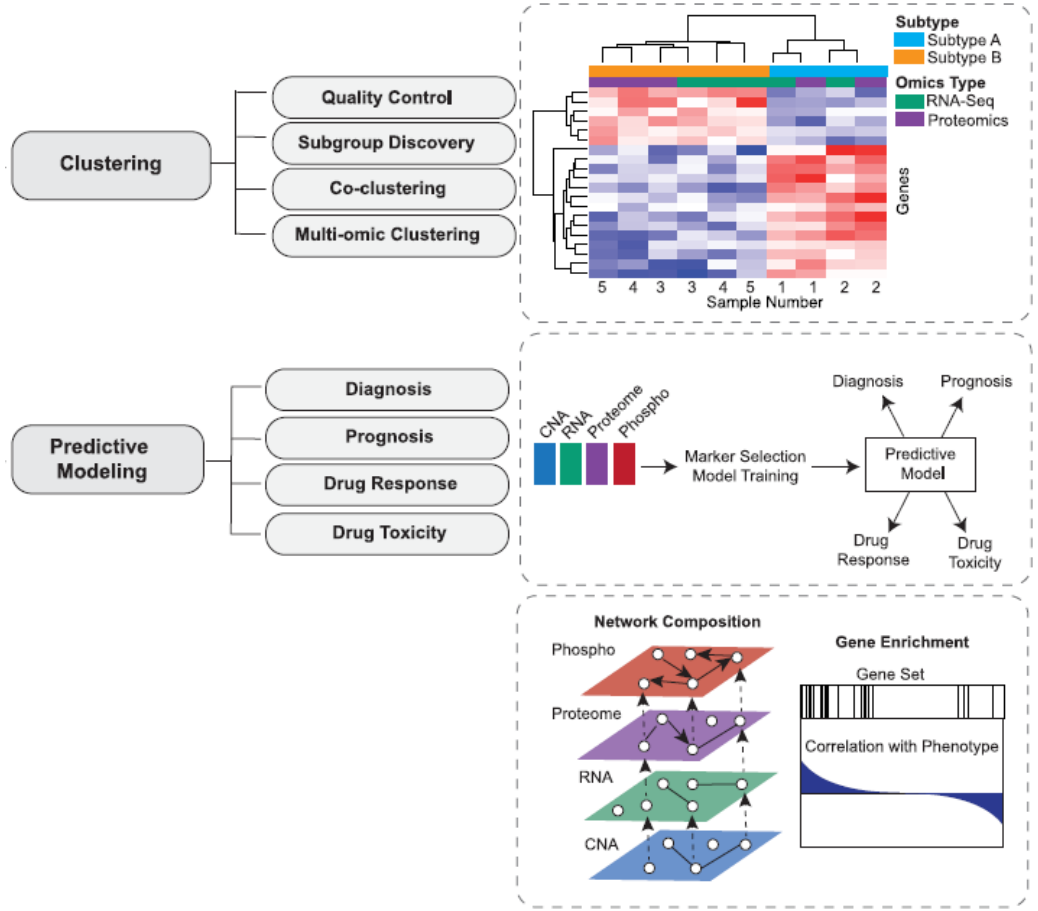
Ruggles 2017

General analysis approaches for proteomics experiments



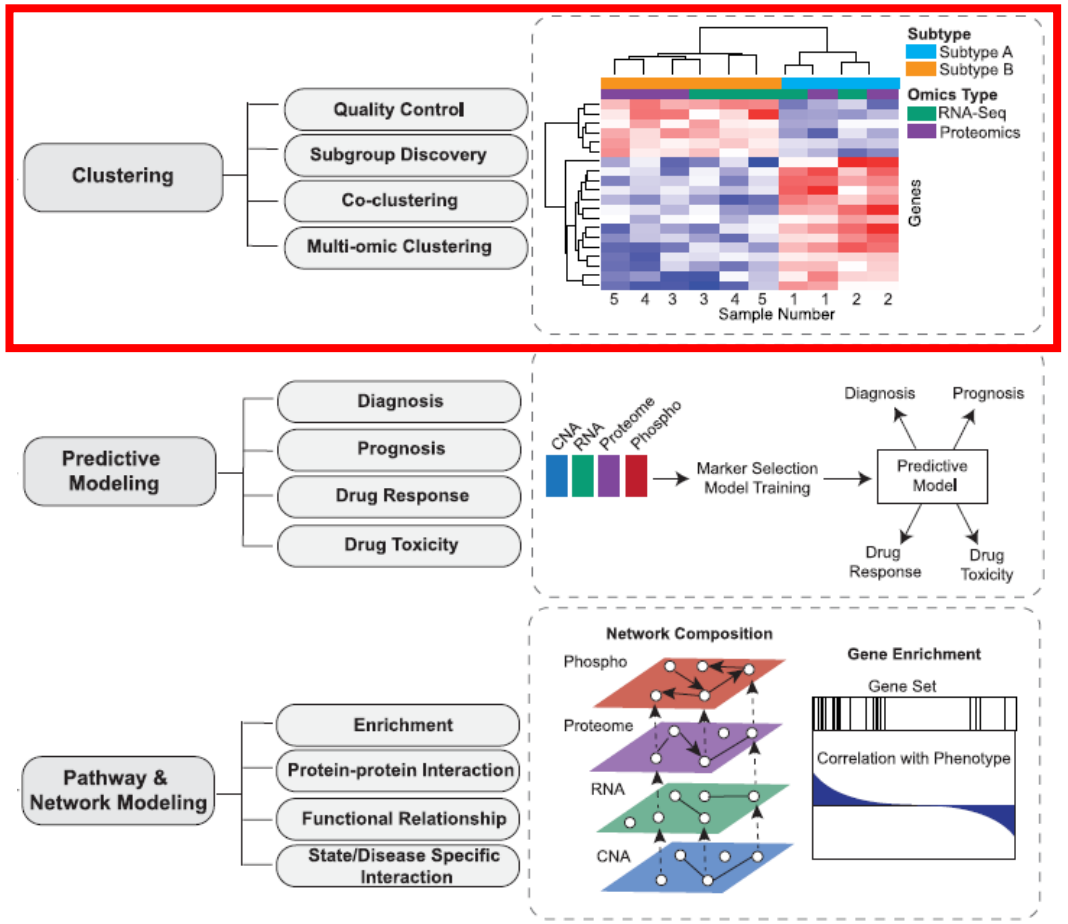
Ruggles 2017

General analysis approaches for proteomics experiments



Ruggles 2017

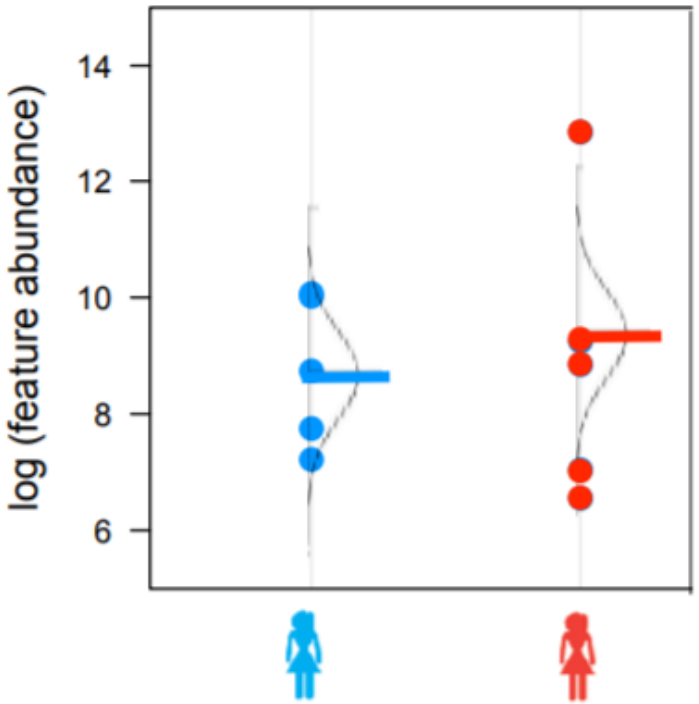
General analysis approaches for proteomics experiments



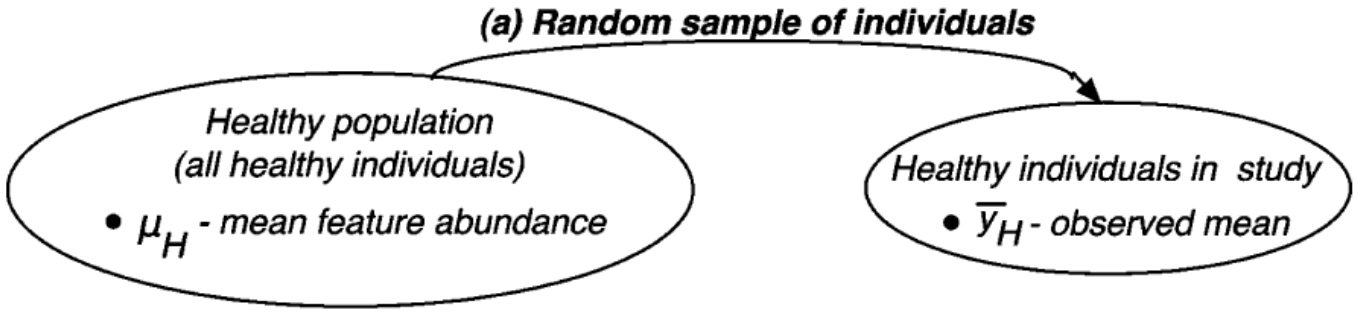
Ruggles 2017

Today's focus:
Class comparison, differential analysis

- 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

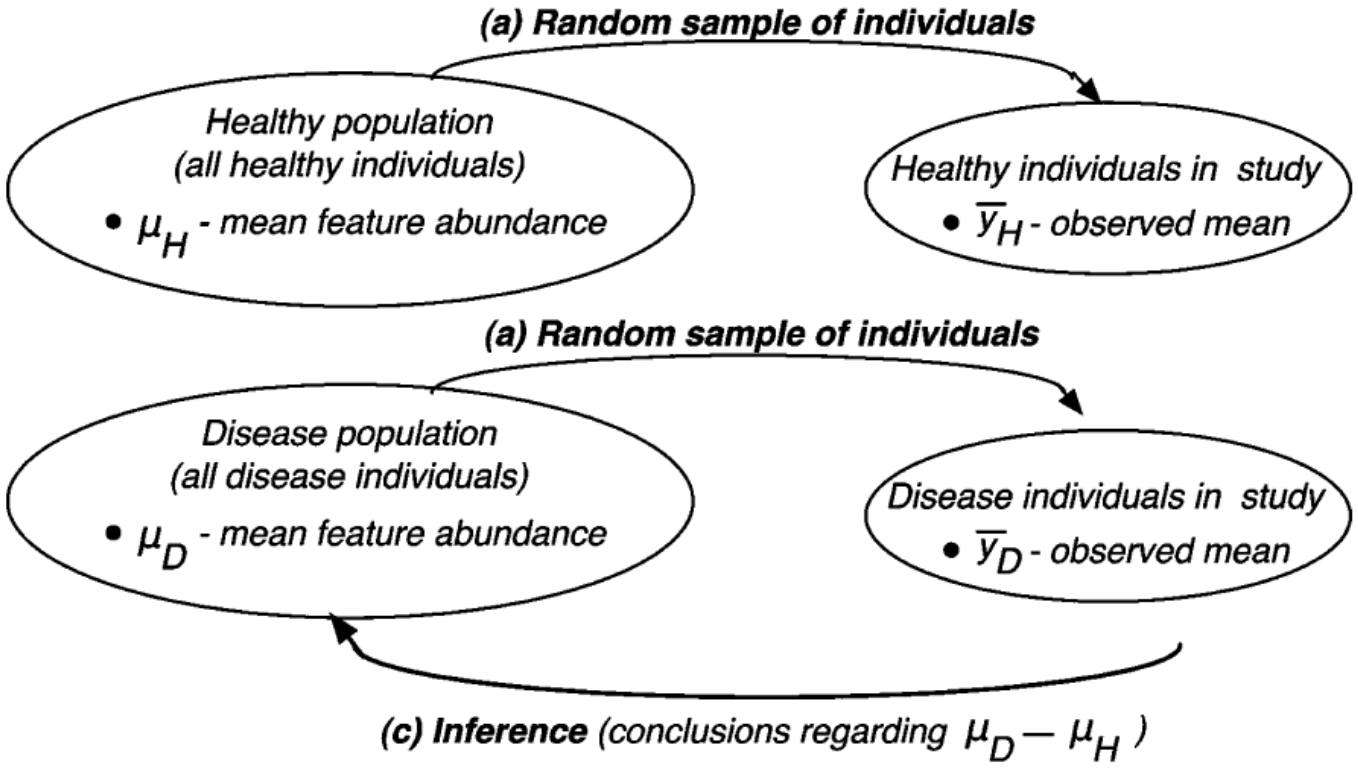


Statistical design of experiments: From data collection, to analysis, to interpretation



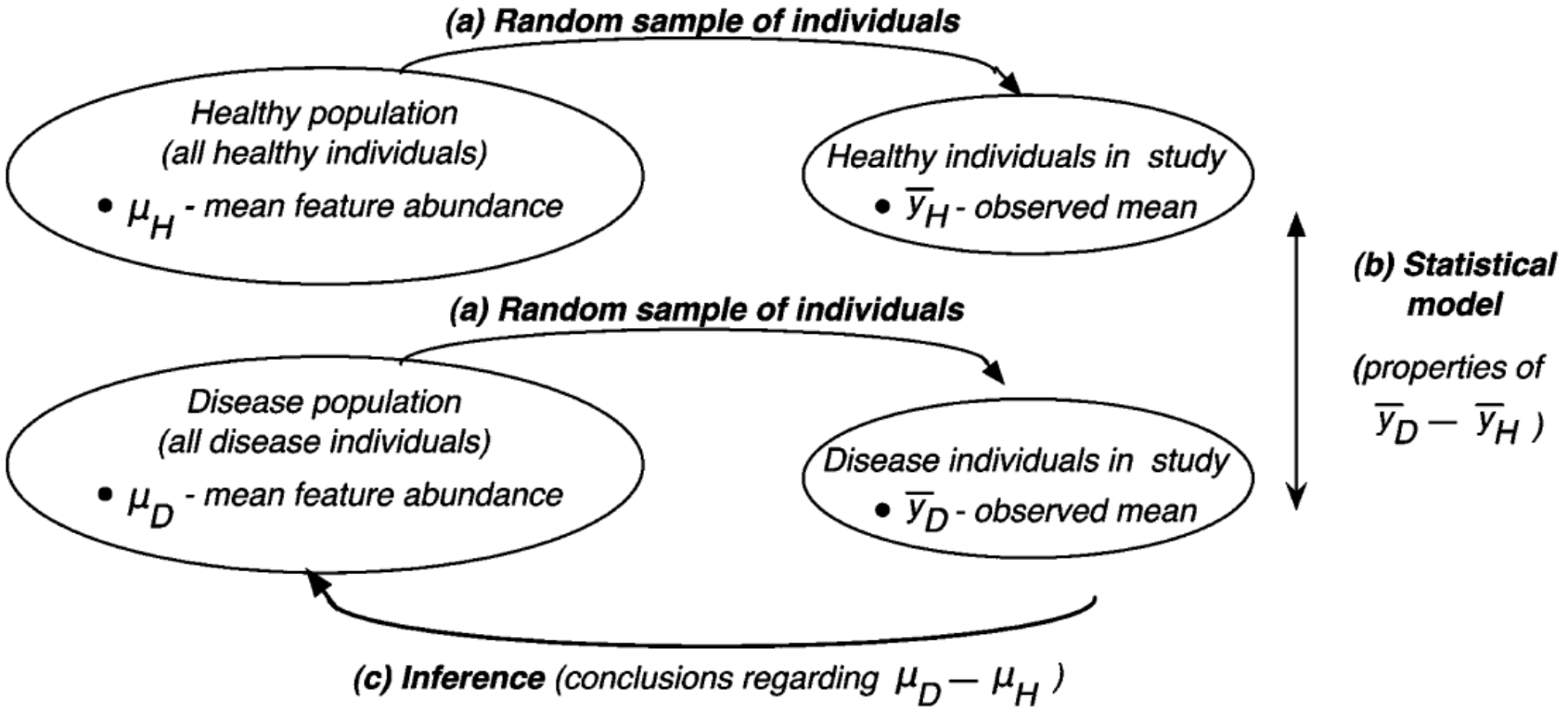
Oberg & Vitek 2008

Statistical design of experiments: From data collection, to analysis, to interpretation



Oberg & Vitek 2008

Statistical design of experiments: From data collection, to analysis, to interpretation



Oberg & Vitek 2008

Agenda

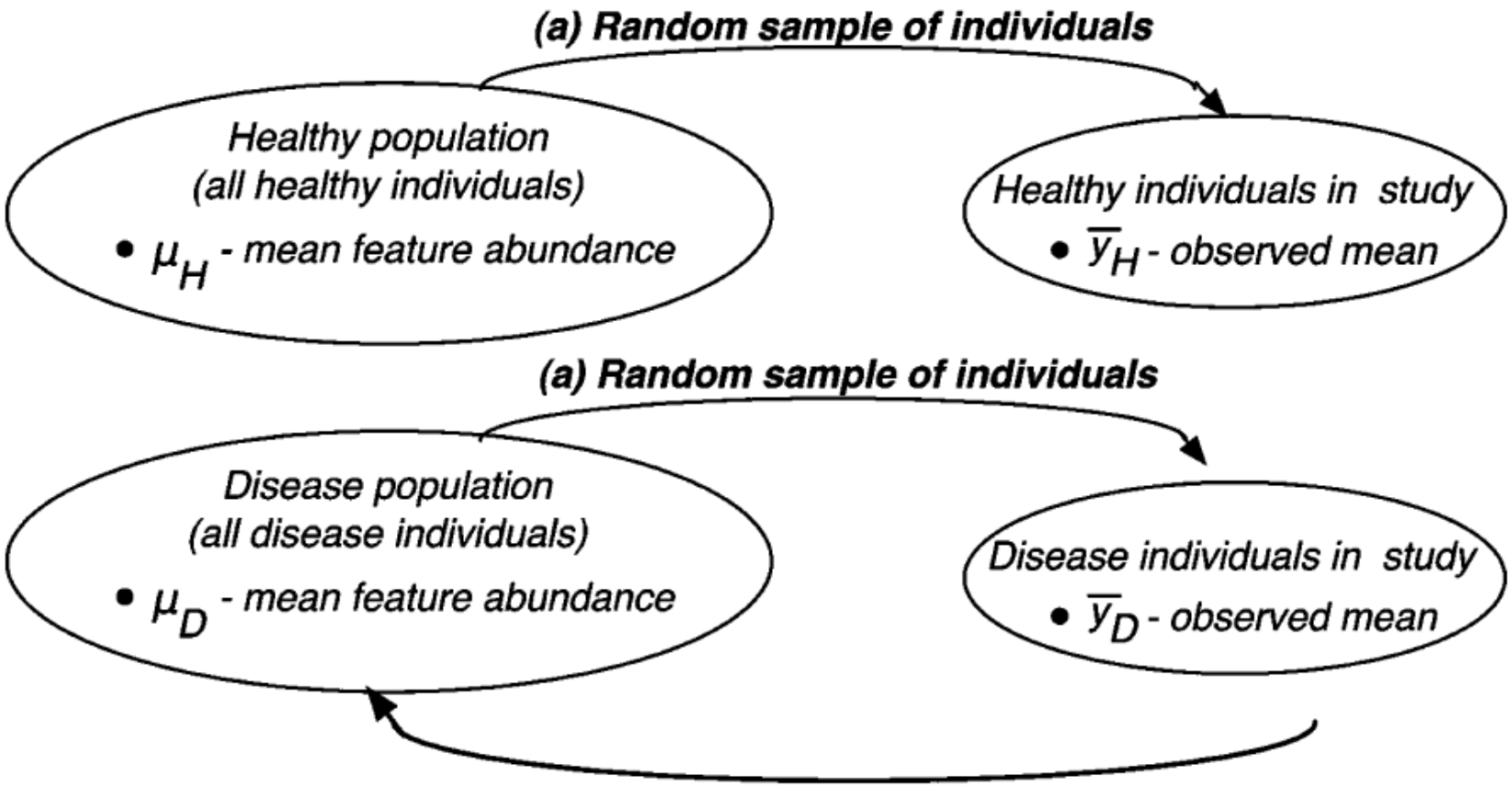
- Translating research experiments into statistical models
- Elements of good statistical/experimental design
  - Replication
  - Randomization
  - Blocking
  - Variation
- Differential abundance tests: the t-test and the ANOVA
- Statistical power: how many samples is enough and the problem with blanket fold change cut-offs

Earlier Today: Group Comparisons natively within Skyline

Take-Home Messages

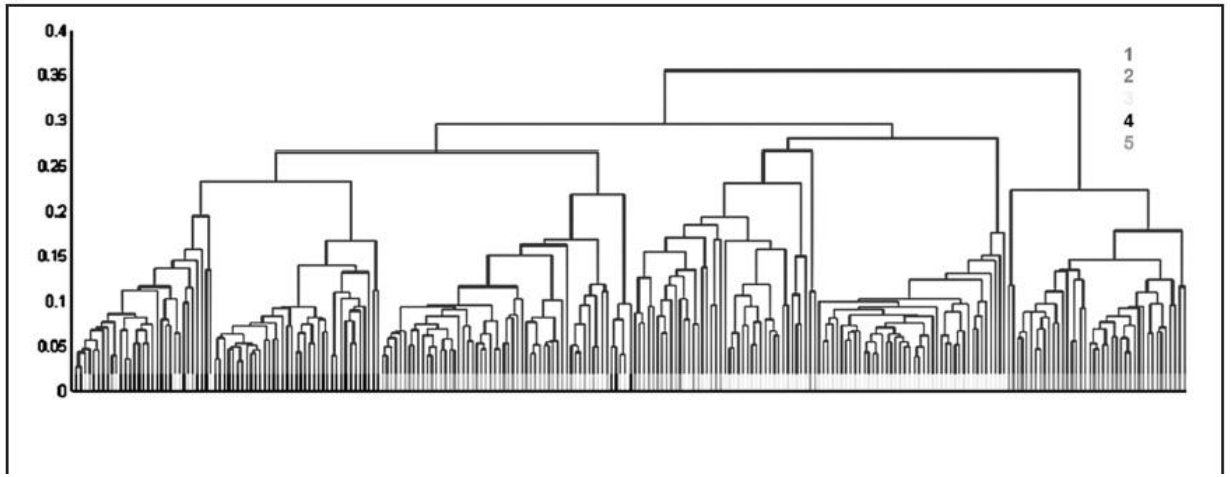
- **Replication** is best when biological
- **Block** what you can
- **Randomize** what you can't block
- **Include all sources of variation** in models

Randomization is easy for experiments, and less so for observational studies



Oberg & Vitek 2008

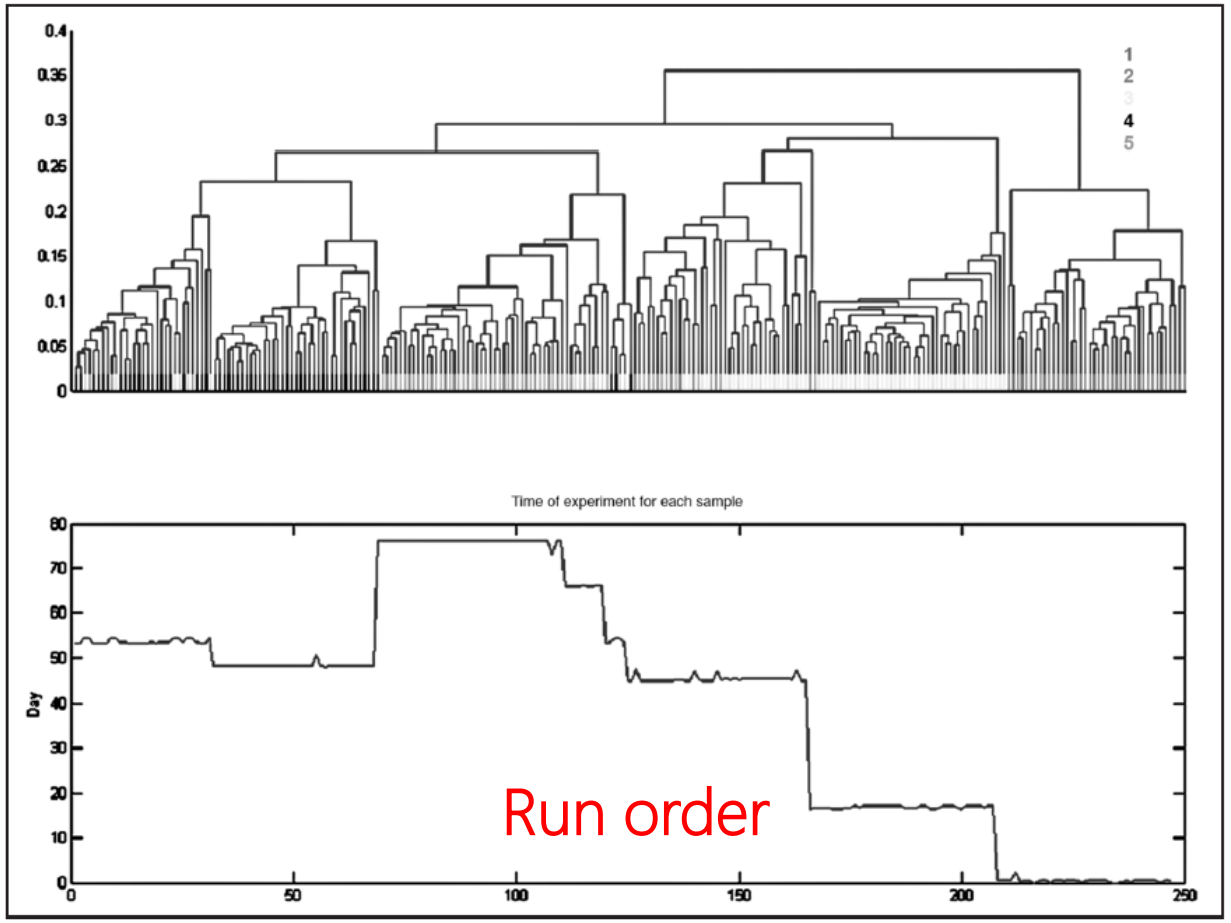
Example of randomization failure



Proteomic profiles of 5 cancer types… expected 5 groups, but analysis suggests six… what happened?

Hu 2005

Example of randomization failure



Hu 2005

Randomization addresses random-chance variance, blocking helps address planned, systematic variability

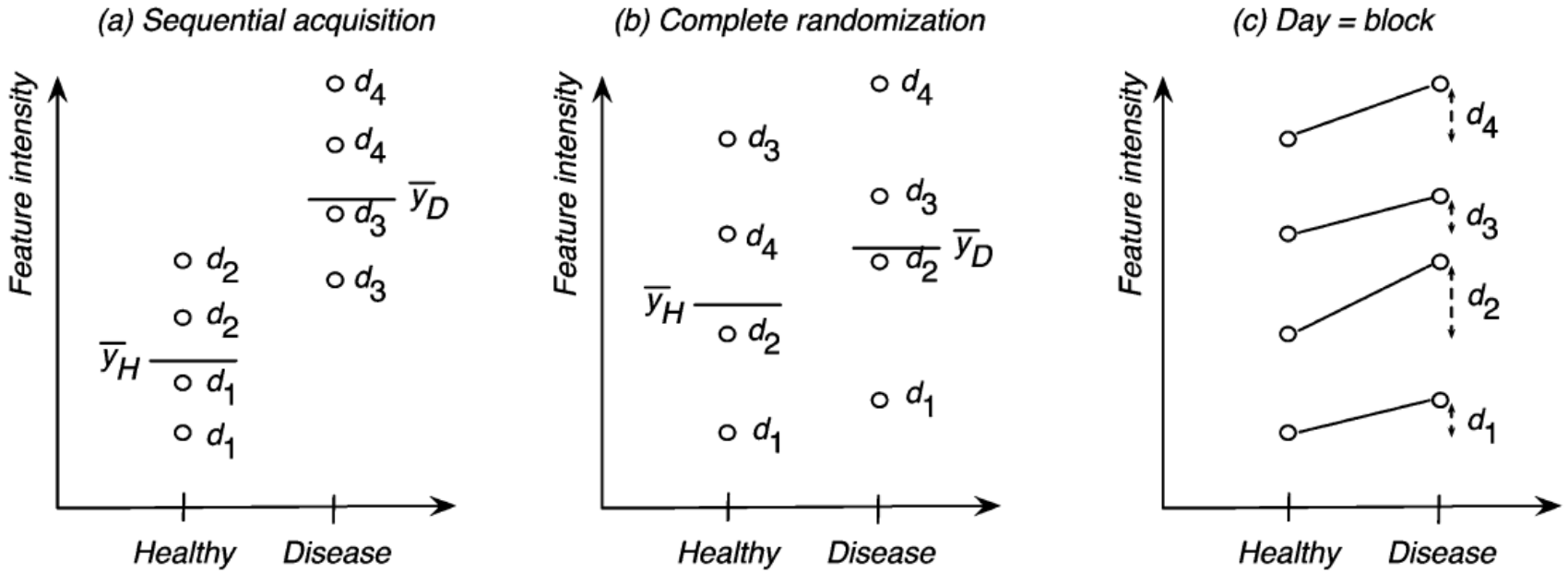
Example: blocking across time to control for variation due to instrument drift

Healthy (n=4)
Disease (n=4)

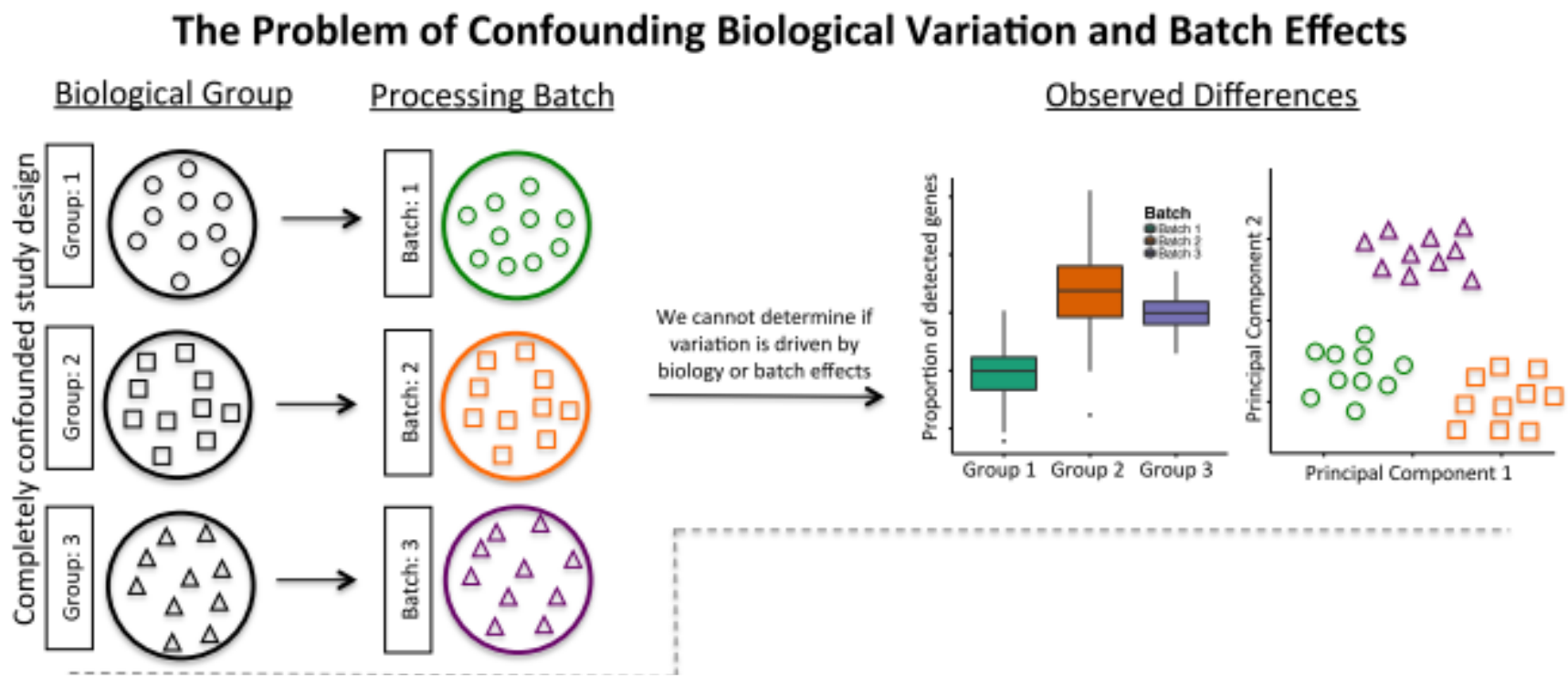
Oberg & Vitek 2008

Randomization addresses random-chance variance, blocking helps address planned, systematic variability

Example: blocking across time to control for variation due to instrument drift

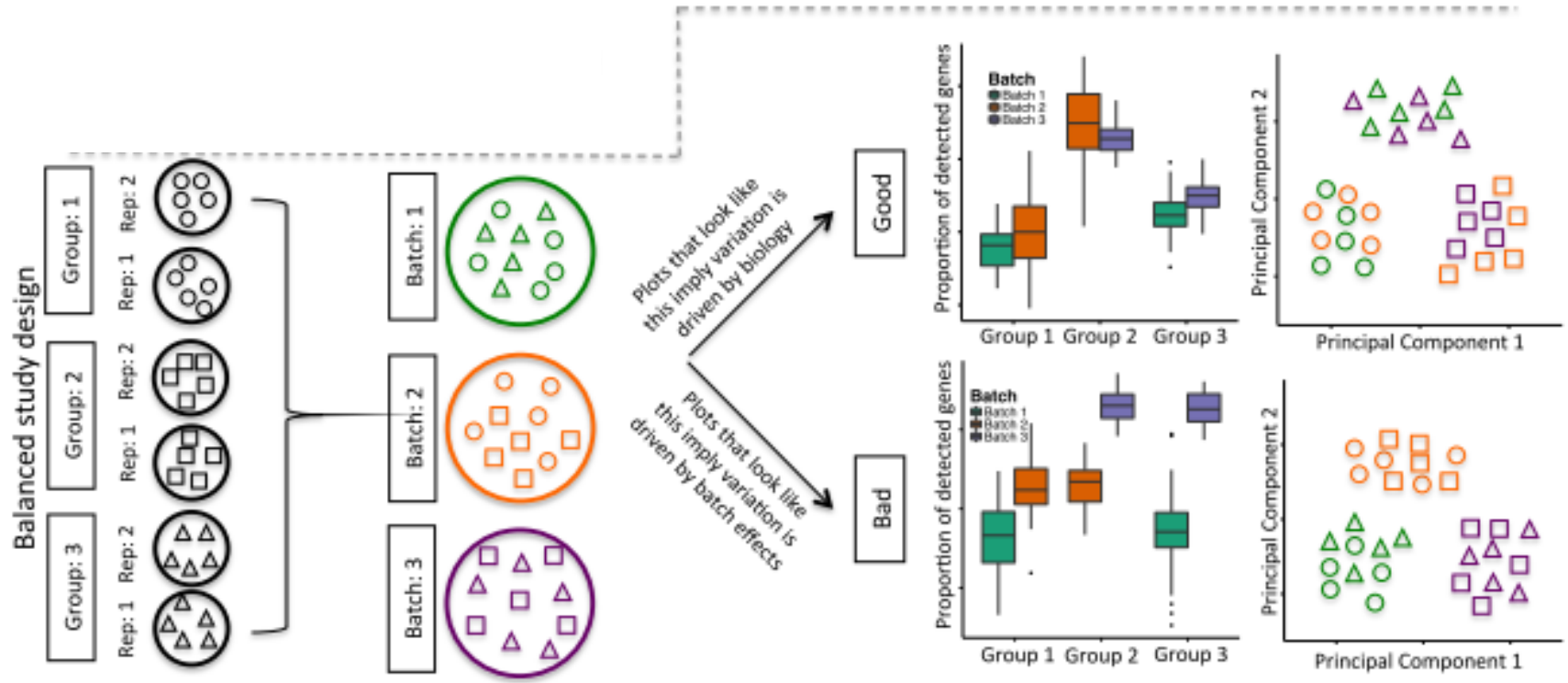


Oberg & Vitek 2008

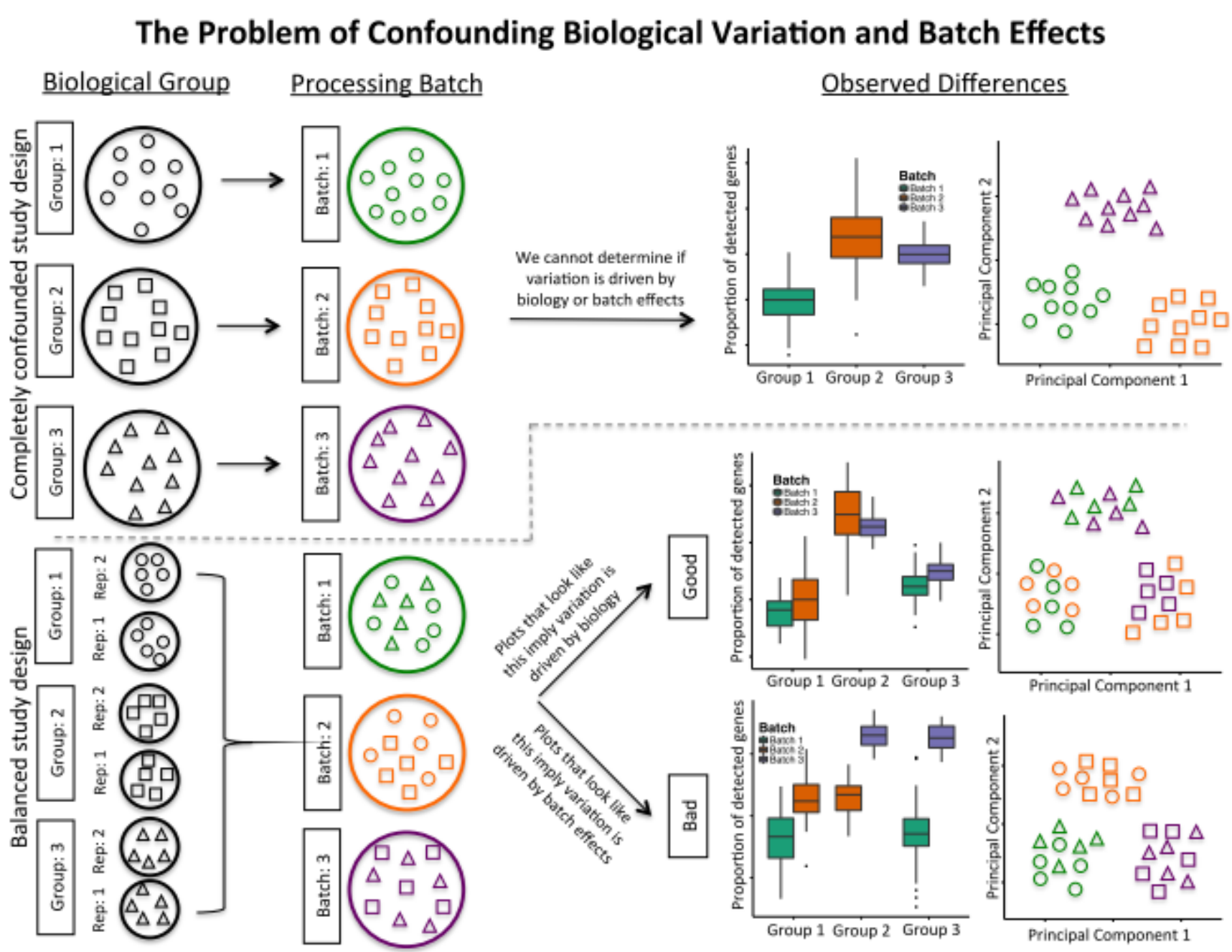


Hicks et al 2015

The Problem of Confounding Biological Variation and Batch Effects



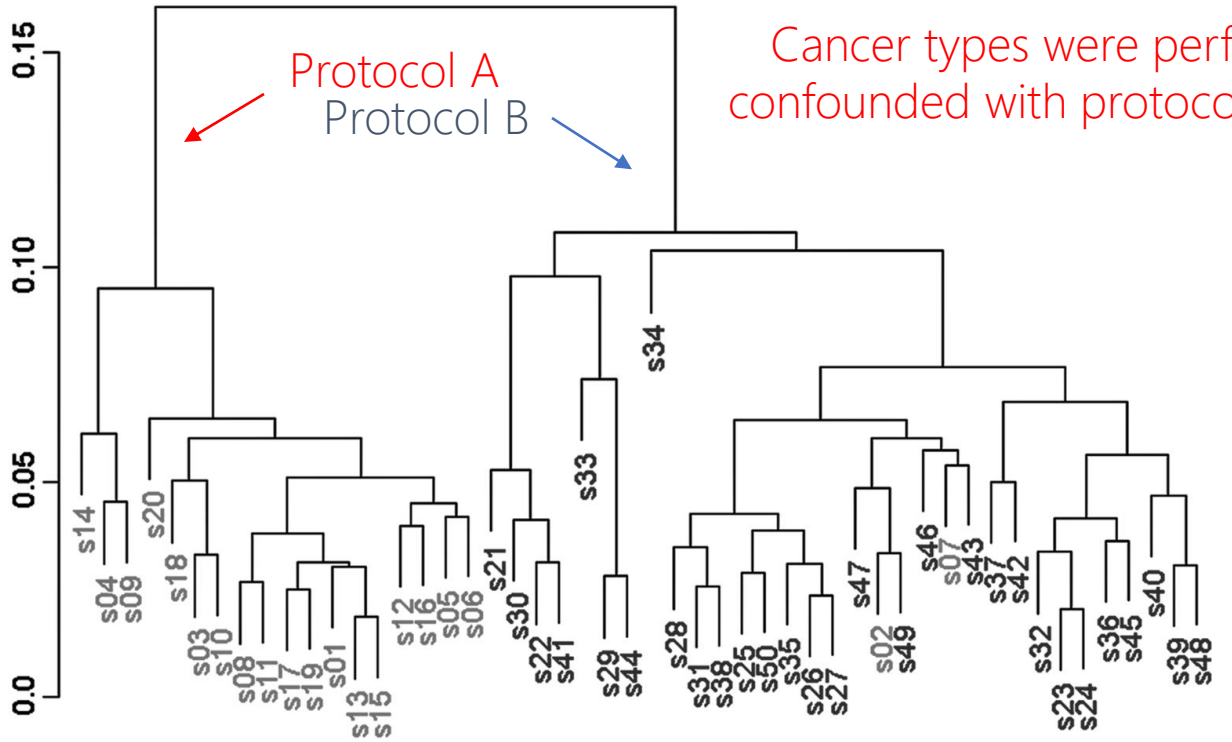
Hicks et al 2015



Hicks et al 2015

Example of blocking failure

Proteomic profiles of 2 cancer types…
so this clustering checks out?



Cancer types were perfectly confounded with protocol used.

The most basic experimental design: Completely randomized design with one replicate

Disease group	Replicate set 1
D_1	X
D_2	X
D_3	X
D_4	X

- **Advantages:** easy to set up, simple analysis, flexible
- **Disadvantages:** must be able to randomize all sources of variation (temporal or spacial)

Blocking designs that are balanced (all treatment comparisons are made with equal precision)

(a) Randomized Complete Block

Disease group	Replicate set 1 Block 1	Replicate set 2 Block 2	··· ···
D_1	X	X	···
D_2	X	X	···
D_3	X	X	···
D_4	X	X	···

- **Advantages:** includes blocking to deal with heterogeneity between experimental units, straightforward statistical analysis
- **Disadvantages:** limited by block size (number of experimental units in each block must be the same as the number of treatments)

Oberg & Vitek 2008

Blocking designs that are balanced (all treatment comparisons are made with equal precision)

(a) Randomized Complete Block

Disease group	Replicate set 1 Block 1	Replicate set 2 Block 2	...
D_1	X	X	...
D_2	X	X	...
D_3	X	X	...
D_4	X	X	...

(b) Balanced Incomplete Block

Disease group	Replicate set 1 Block 1	Block 2	Block 3	Block 4	Block 5	...
D_1	X	X	X	X		...
D_2	X	X	X		X	...
D_3	X	X		X	X	...
D_4	X		X	X	X	...
D_5		X	X	X	X	...

- **Advantages:** includes blocking to deal with heterogeneity between experimental units, straightforward statistical analysis
- **Disadvantages:** limited by block size (number of experimental units in each block must be the same as the number of treatments)

- **Advantages:** allows for smaller blocks than the RCB, still a balanced design
- **Disadvantages:** range of available designs is limited (treatments must all have equal replication, and each pair of treatments must occur together within a block exactly the same number of times over the whole experiment)

Oberg & Vitek 2008

Example constructions for Balanced Incomplete Block Designs

(a) Balanced Incomplete Block

Disease group	Replicate set 1										
	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8	Block 9	Block 10	⋯
D_1	X_{L_1}	X_{L_2}	X_{L_1}	X_{L_2}							⋯
D_2	X_{L_2}				X_{L_1}	X_{L_2}	X_{L_1}				⋯
D_3		X_{L_1}			X_{L_2}			X_{L_1}	X_{L_2}		⋯
D_4			X_{L_2}			X_{L_1}		X_{L_2}		X_{L_1}	⋯
D_5				X_{L_1}			X_{L_2}		X_{L_1}	X_{L_2}	⋯

Note: treatments must all have equal replication, and each pair of treatments must occur together within a block the same number of times over the whole experiment

(b) Reference

Disease group	Replicate set 1					
	Block 1	Block 2	Block 3	Block 4	Block 5	⋯
R	R_{L_1}	R_{L_1}	R_{L_1}	R_{L_1}	R_{L_1}	⋯
D_1	X_{L_2}					⋯
D_2		X_{L_2}				⋯
D_3			X_{L_2}			⋯
D_4				X_{L_2}		⋯
D_5					X_{L_2}	⋯

Here, the reference sample is of no scientific interest in the study, but adds experimental noise to the data to control the between-experiment variation

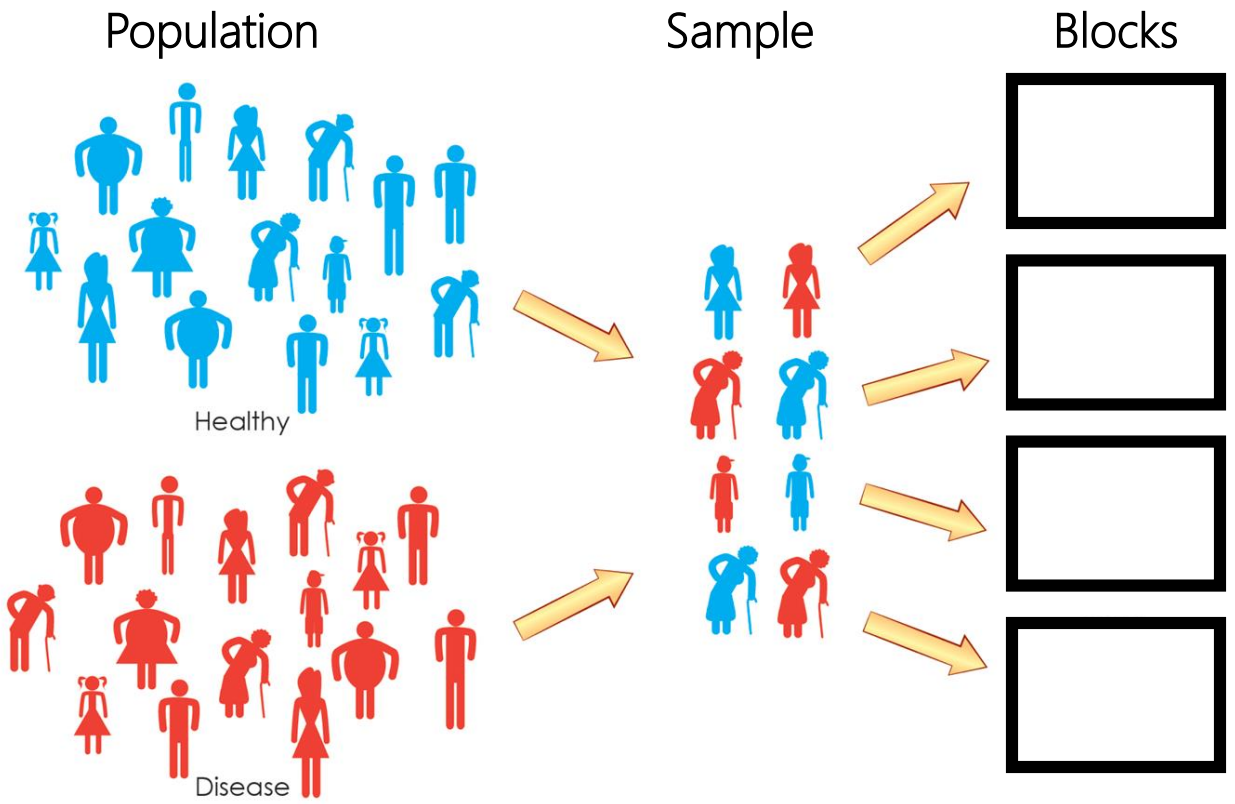
(c) Loop

Disease group	Replicate set 1					
	Block 1	Block 2	Block 3	Block 4	Block 5	⋯
D_1	X_{L_1}				X_{L_2}	⋯
D_2	X_{L_2}	X_{L_1}				⋯
D_3		X_{L_2}	X_{L_1}			⋯
D_4			X_{L_2}	X_{L_1}		⋯
D_5				X_{L_2}	X_{L_1}	⋯

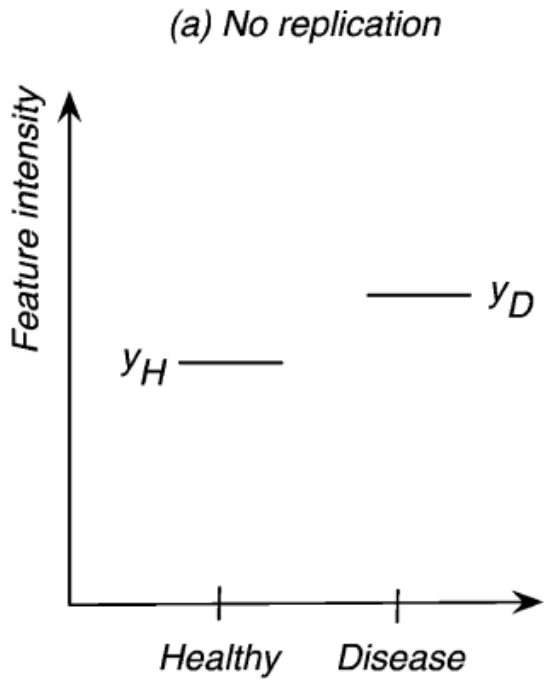
Here, the aim is to compare treatment conditions and there is no need for a control because each treatment is compared directly with another

Oberg & Vitek 2008

Thought exercise:
How should these eight samples be blocked?

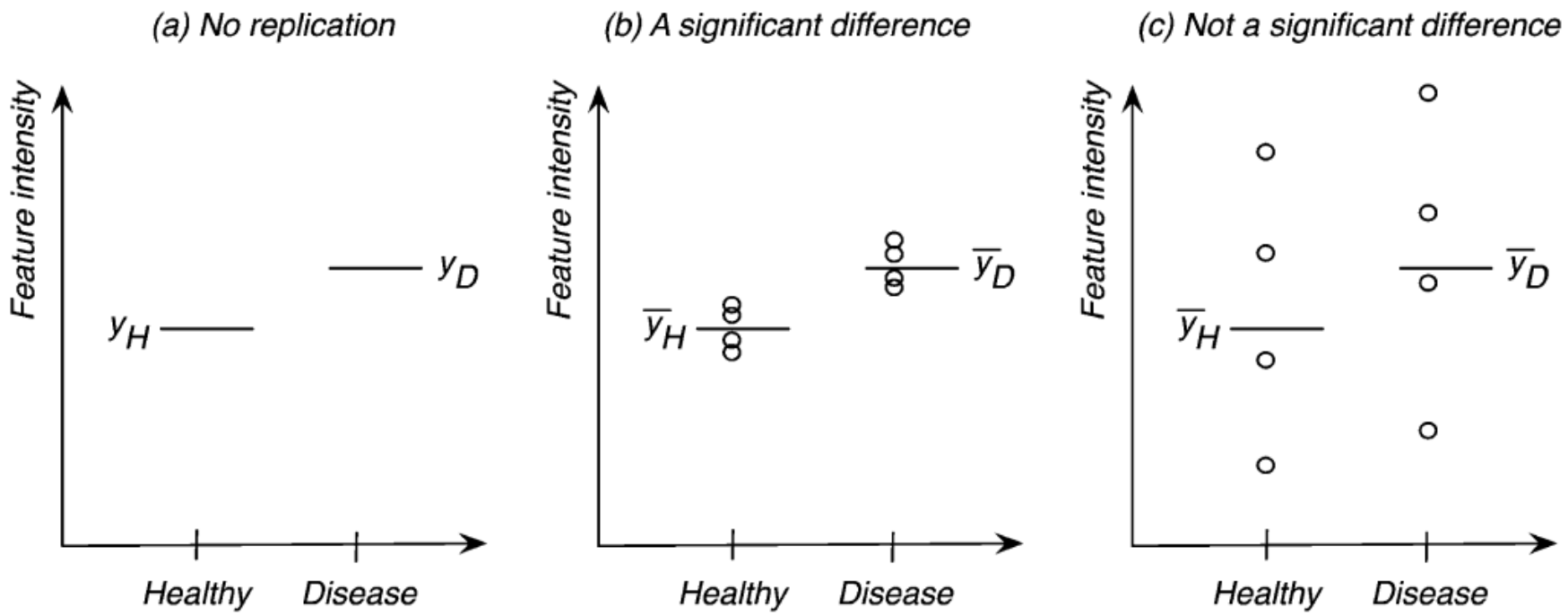


Replication assesses variation due to random chance, ensures reliability of conclusions



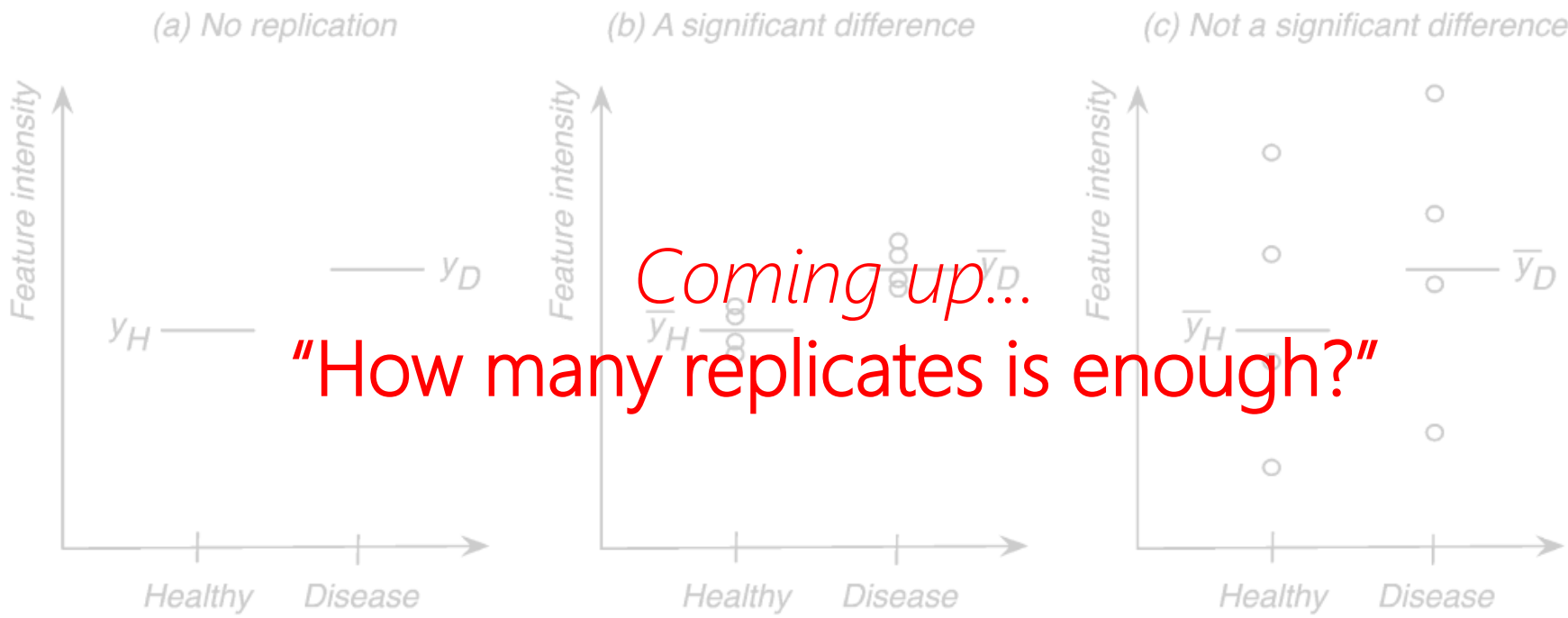
Oberg & Vitek 2008

Replication assesses variation due to random chance, ensures reliability of conclusions



Oberg & Vitek 2008

Replication assesses variation due to random chance, ensures reliability of conclusions

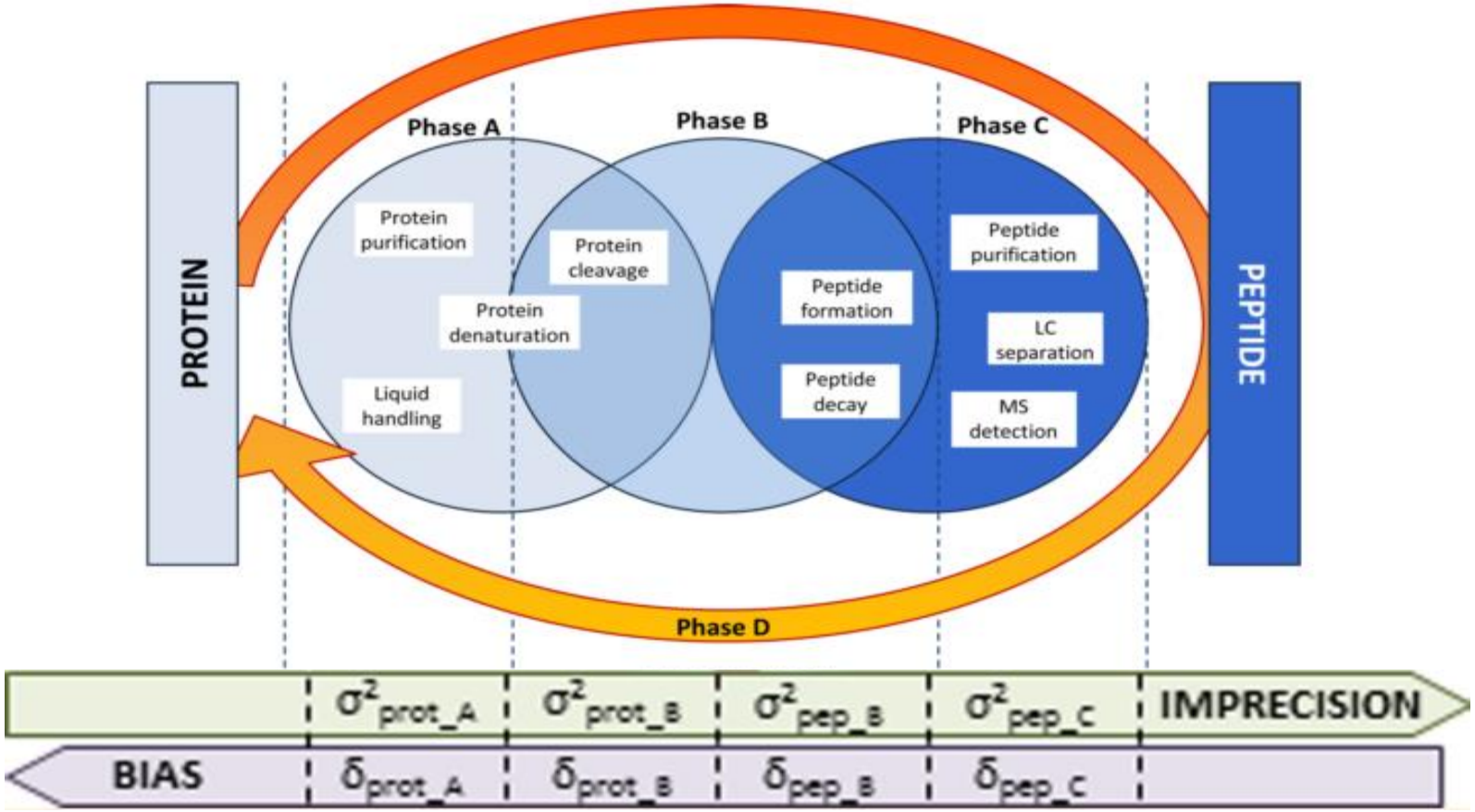


Coming up…

"How many replicates is enough?"

Oberg & Vitek 2008

What are common sources of variation in proteomics? How can they be avoided or at least accounted for?



Bias: systematic errors in our conclusions

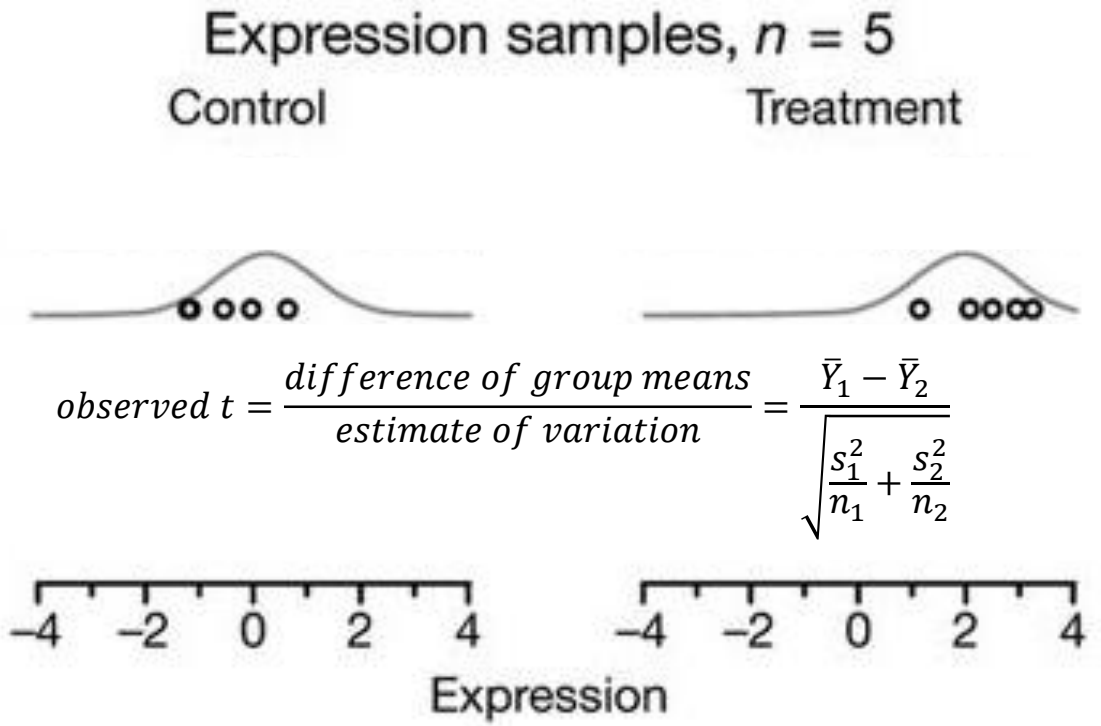
Van den Broek et al 2014

Agenda

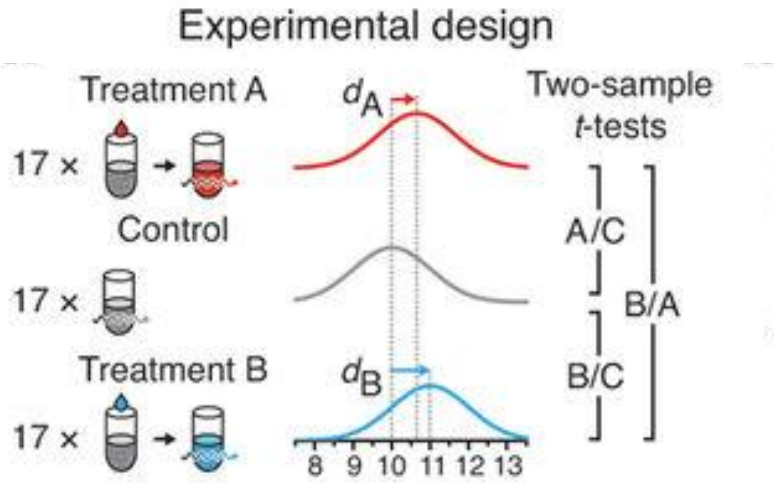
- Translating research experiments into statistical models
- Elements of good statistical/experimental design
  - Replication
  - Randomization
  - Blocking
  - Variation
- **Differential abundance tests: the t-test and the ANOVA**
- Statistical power: how many samples is enough and the problem with blanket fold change cut-offs

Earlier Today: Group Comparisons natively within Skyline

Differential abundance by the two-sample t-test

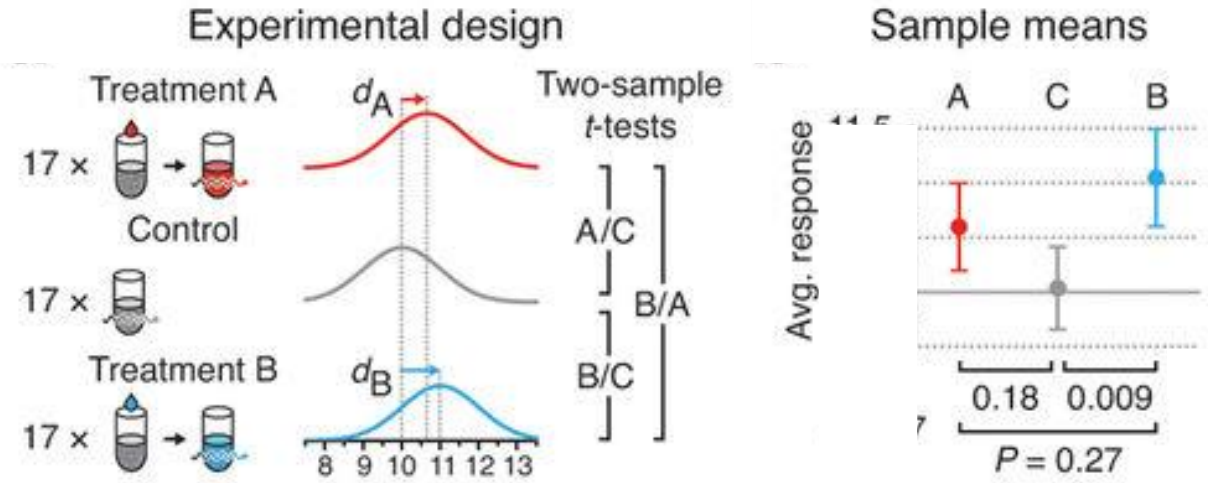


Multiple two-sample t-test for more complex experimental designs



https://www.nature.com/articles/nmeth.2974

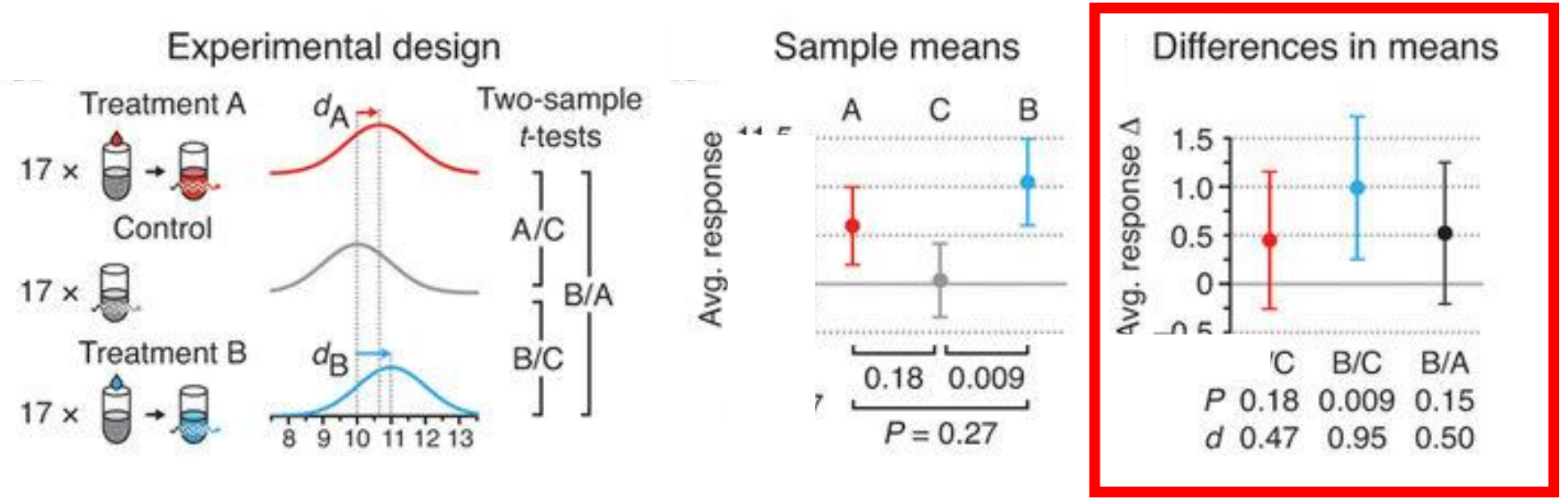
Multiple two-sample t-test for more complex experimental designs



https://www.nature.com/articles/nmeth.2974

Multiple two-sample t-test for more complex experimental designs

But these error bars overlap?

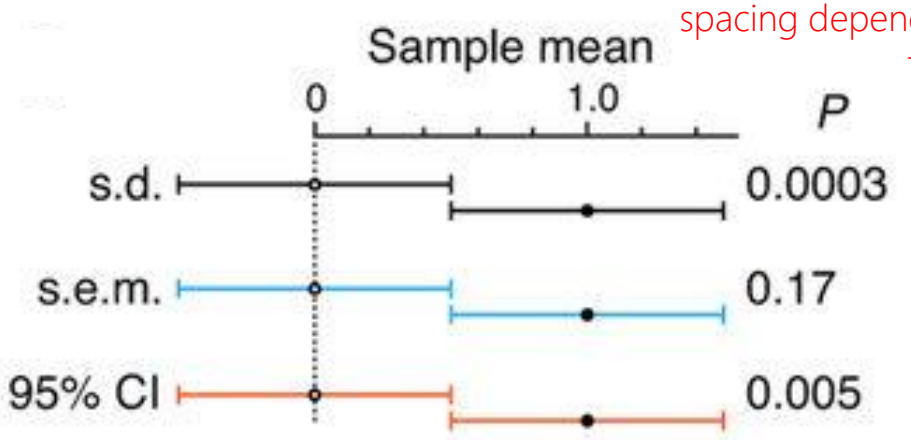


https://www.nature.com/articles/nmeth.2974

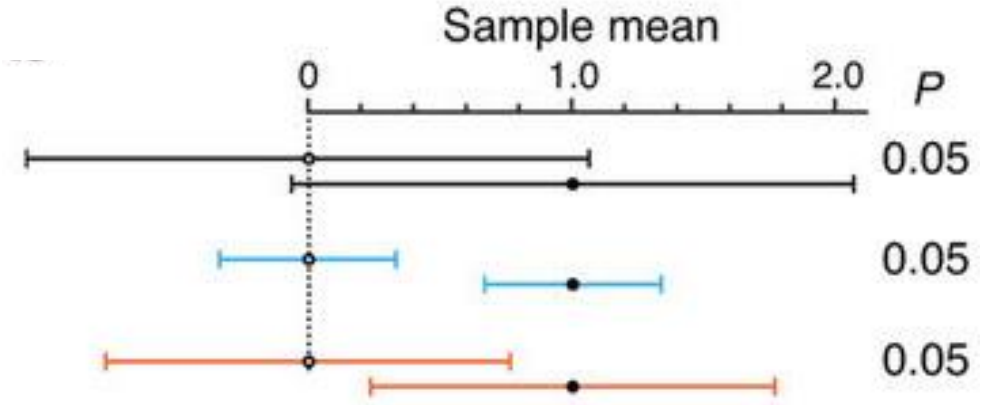
Absence (or presence) of error bar overlap does not always mean statistical significance

Error bar width and interpretation of spacing depends on the error bar type

Scaled error bars,
Unequal p-values



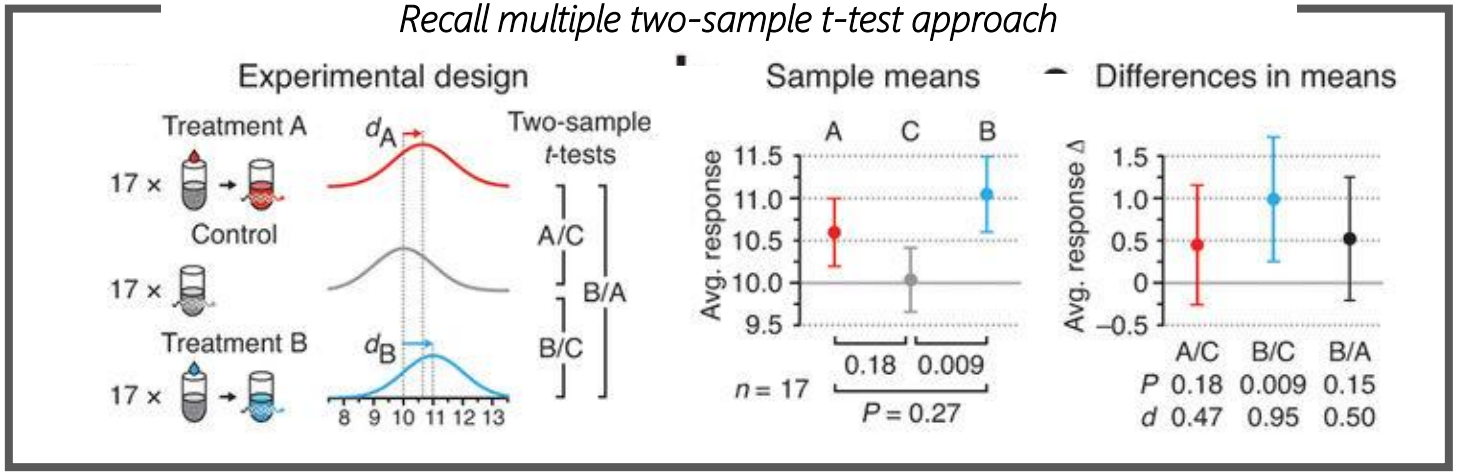
Unscaled error bars,
Equal p-values



https://www.nature.com/articles/nmeth.2659

One-way ANOVA as an alternative to multiple two-sample t-tests

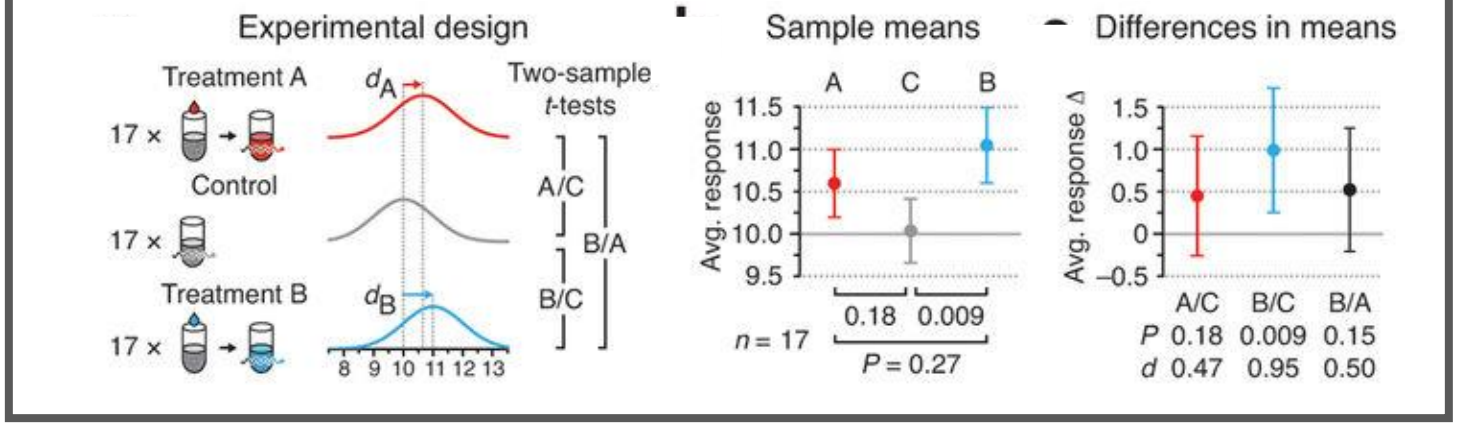
Recall multiple two-sample t-test approach



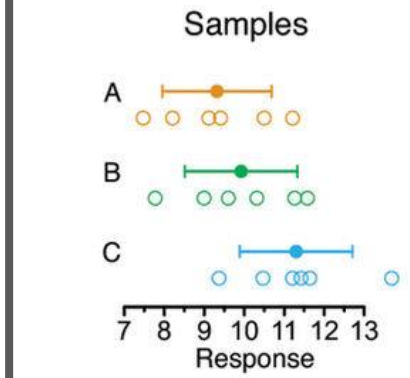
https://www.nature.com/articles/nmeth.2974

One-way ANOVA as an alternative to multiple two-sample t-tests

Recall multiple two-sample t-test approach



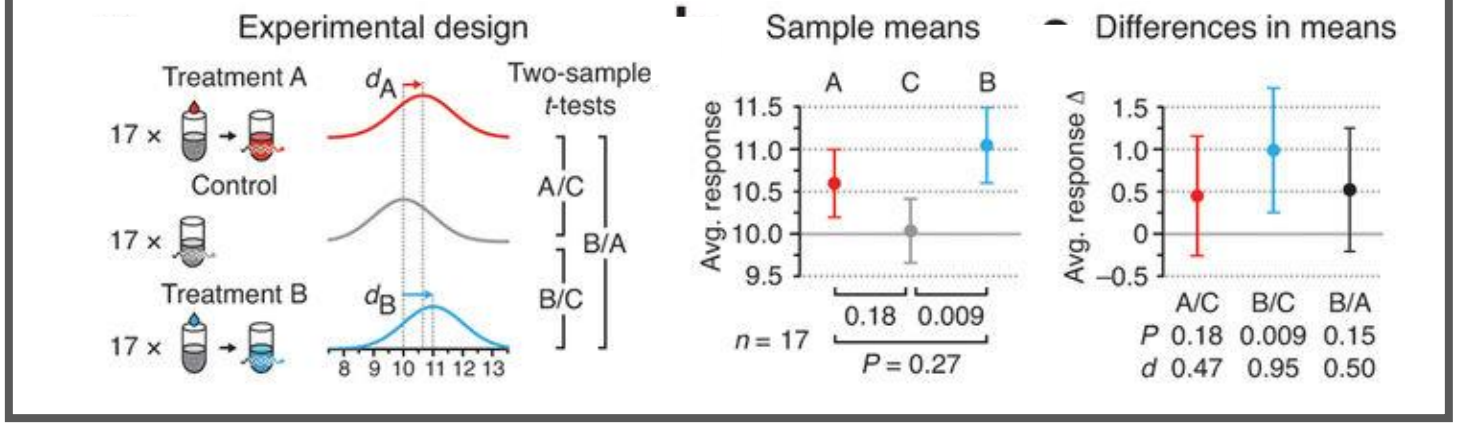
One-way ANOVA approach



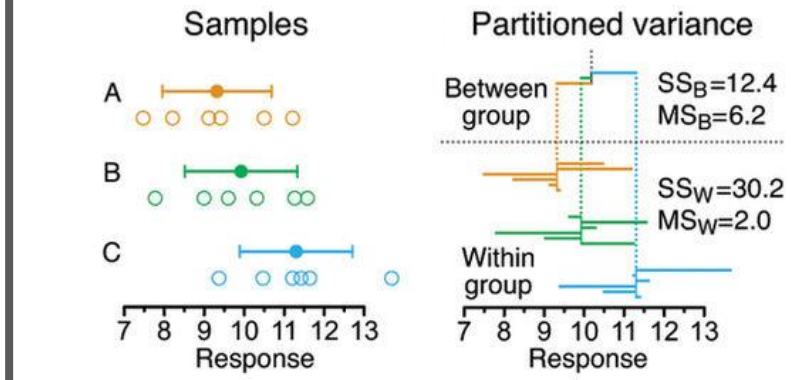
https://www.nature.com/articles/nmeth.3005

One-way ANOVA as an alternative to multiple two-sample t-tests

Recall multiple two-sample t-test approach



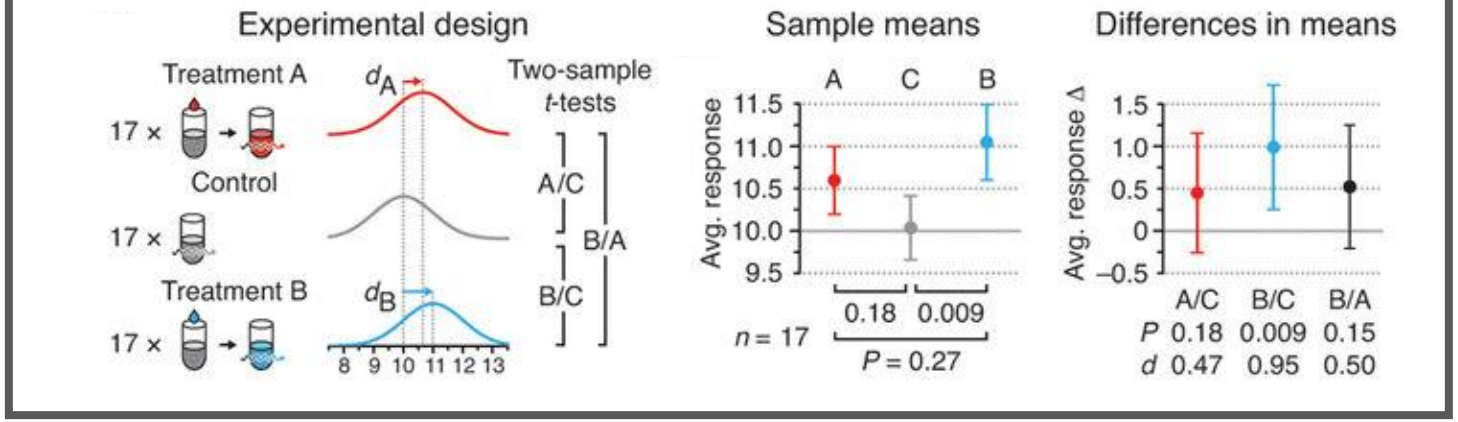
One-way ANOVA approach



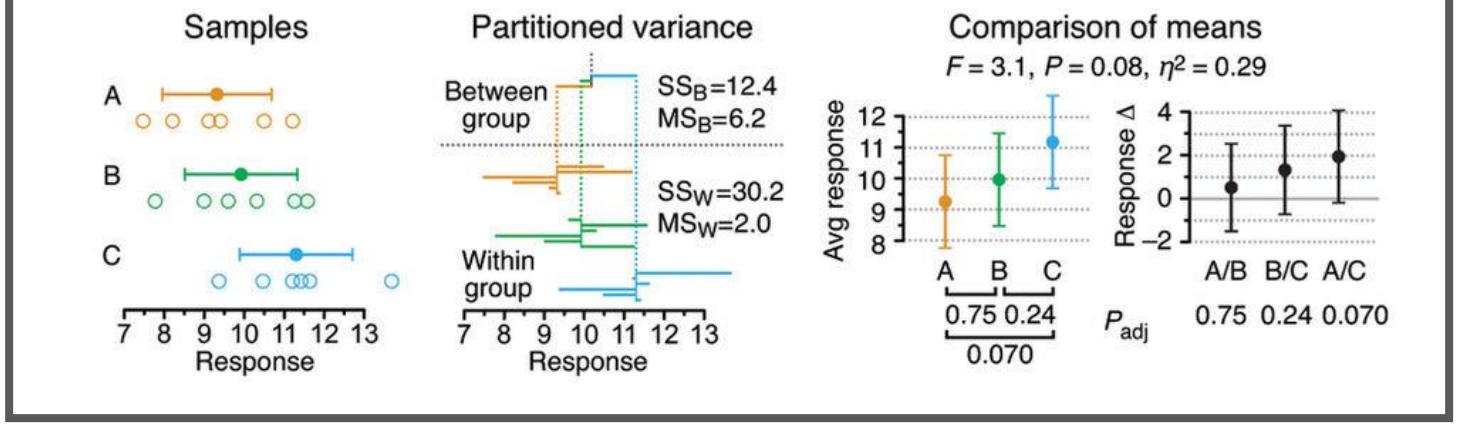
https://www.nature.com/articles/nmeth.3005

One-way ANOVA as an alternative to multiple two-sample t-tests

Recall multiple two-sample t-test approach



One-way ANOVA approach



https://www.nature.com/articles/nmeth.3005

ANOVA approaches are powerful because they include known sources of variation

Basic ANOVA, completely randomized design

Observed feature intensity	=	Systematic mean signal of disease group	+	Random deviation due to all sources of variation
y_{ij}	=	$Group\ mean_i$	+	$Error_{j(i)} \sim N\left(0, \sigma^2\right)$

Mixed effects ANOVA, with technical replicates

Observed feature intensity	=	Systematic mean signal of disease group	+	Random deviation due to individual	+	Random deviation due to sample preparation	+	Random deviation due to measurement error
y_{ijkl}	=	$Group\ mean_i$	+	$Indiv(Group)_{j(i)} \sim N\left(0, \sigma^2_{Indiv}\right)$	+	$Prep(Indiv)_{k(ij)} \sim N\left(0, \sigma^2_{Prep}\right)$	+	$Error_{l(ijk)} \sim N\left(0, \sigma^2_{Error}\right)$

Mixed effects ANOVA, with blocking

Observed feature intensity	=	Systematic mean signal of disease group	+	Random deviation due to block (e.g. plate or day)	+	Random deviation due to individual	+	Random deviation due to measurement error
y_{ijkl}	=	$Group\ mean_i$	+	$Block_k \sim N\left(0, \sigma^2_{Block}\right)$	+	$Indiv(Group)_{j(i)} \sim N\left(0, \sigma^2_{Indiv}\right)$	+	$Error_{l(ijk)} \sim N\left(0, \sigma^2_{Error}\right)$

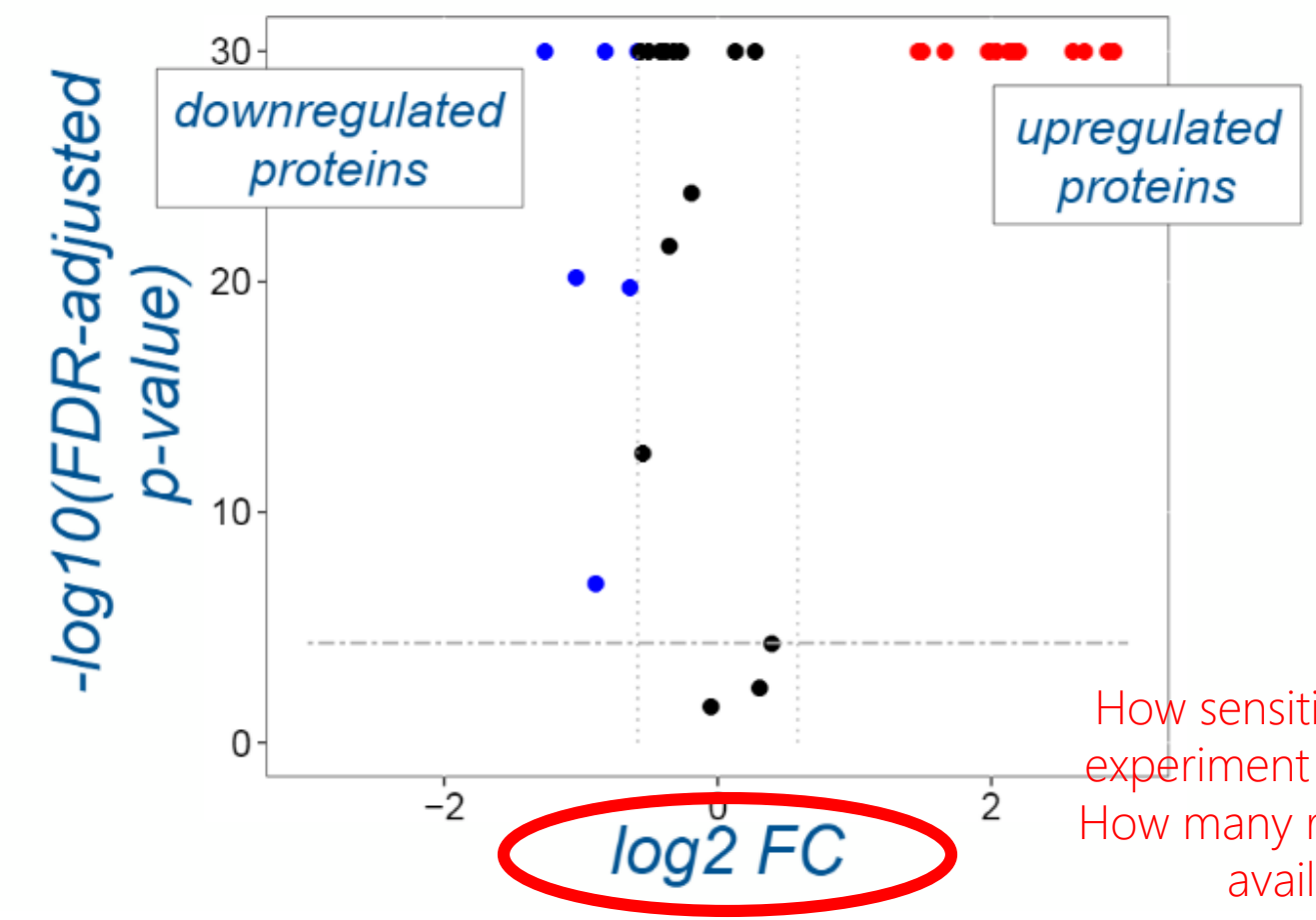
Oberg & Vitek 2008

Agenda

- Translating research experiments into statistical models
- Elements of good statistical/experimental design
  - Replication
  - Randomization
  - Blocking
  - Variation
- Differential abundance tests: the t-test and the ANOVA
- Statistical power: how many samples is enough and the problem with blanket fold change cut-offs

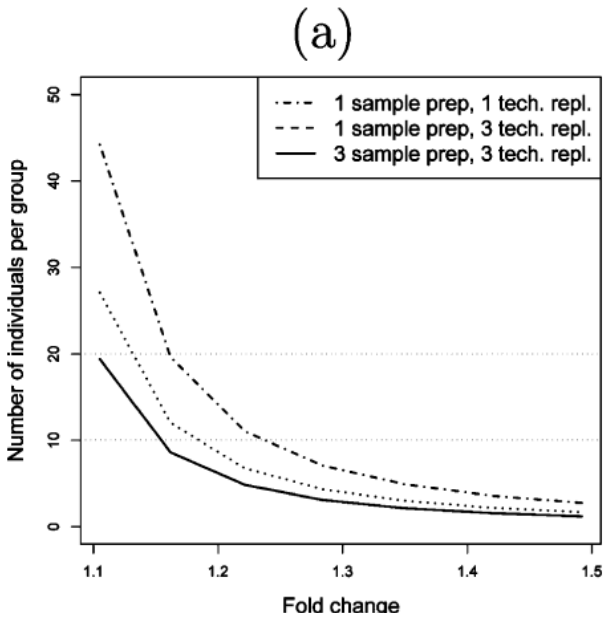
Earlier Today: Group Comparisons natively within Skyline

Statistical significance is only part of the data science story



How sensitive does the experiment need to be? How many replicates are available?

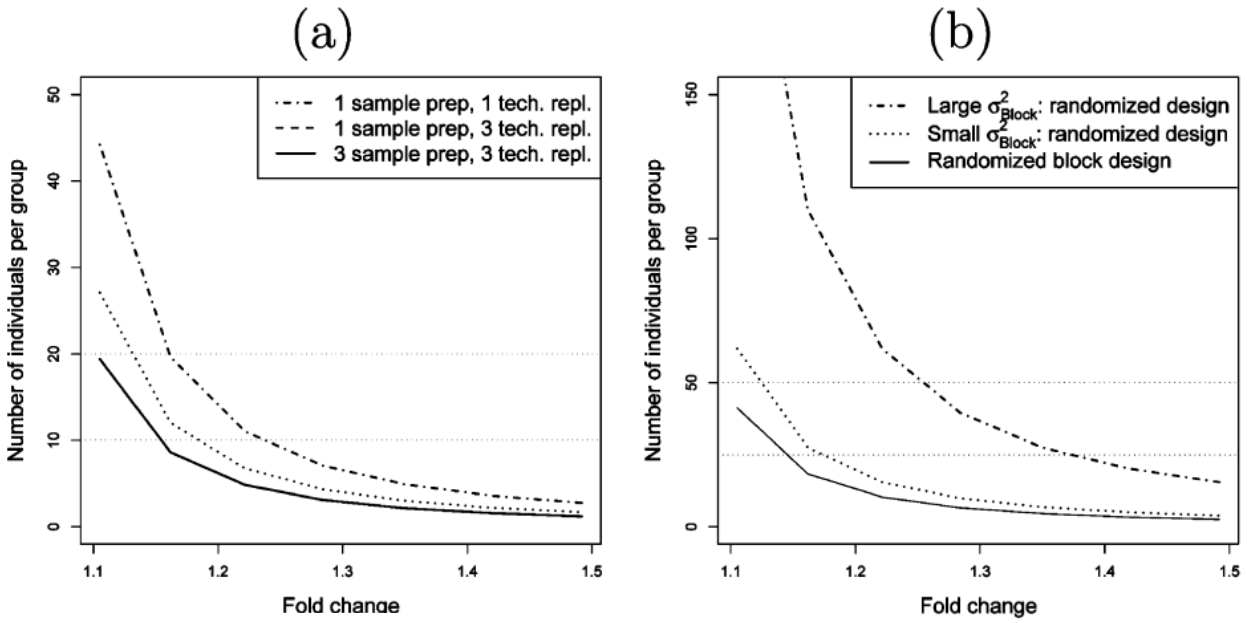
Power analysis informs how many replicates is "enough" and how sensitive the experiment is



Biological replicates are more statistically useful than technical replicates

Oberg & Vitek 2008

Power analysis informs how many replicates is "enough" and how sensitive the experiment is

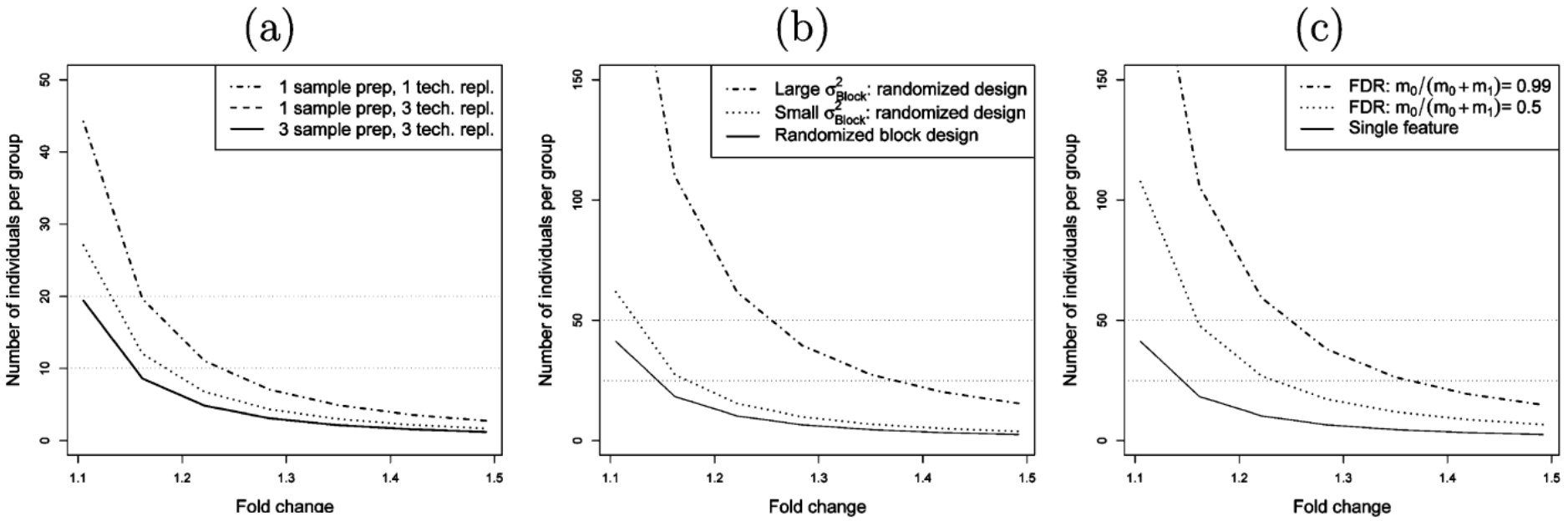


Biological replicates are more statistically useful than technical replicates

Block to minimize between-block variance

Oberg & Vitek 2008

Power analysis informs how many replicates is "enough" and how sensitive the experiment is



Biological replicates are more statistically useful than technical replicates

Block to minimize between-block variance

Fewer targets is better than more

Oberg & Vitek 2008

Example of power analysis: biospecimen cohort size in proteomics-based biomarker discovery and verification studies

Biospecimen Sample blood

100 candidates passing

Lower — *STATISTICAL POWER* — *Higher*

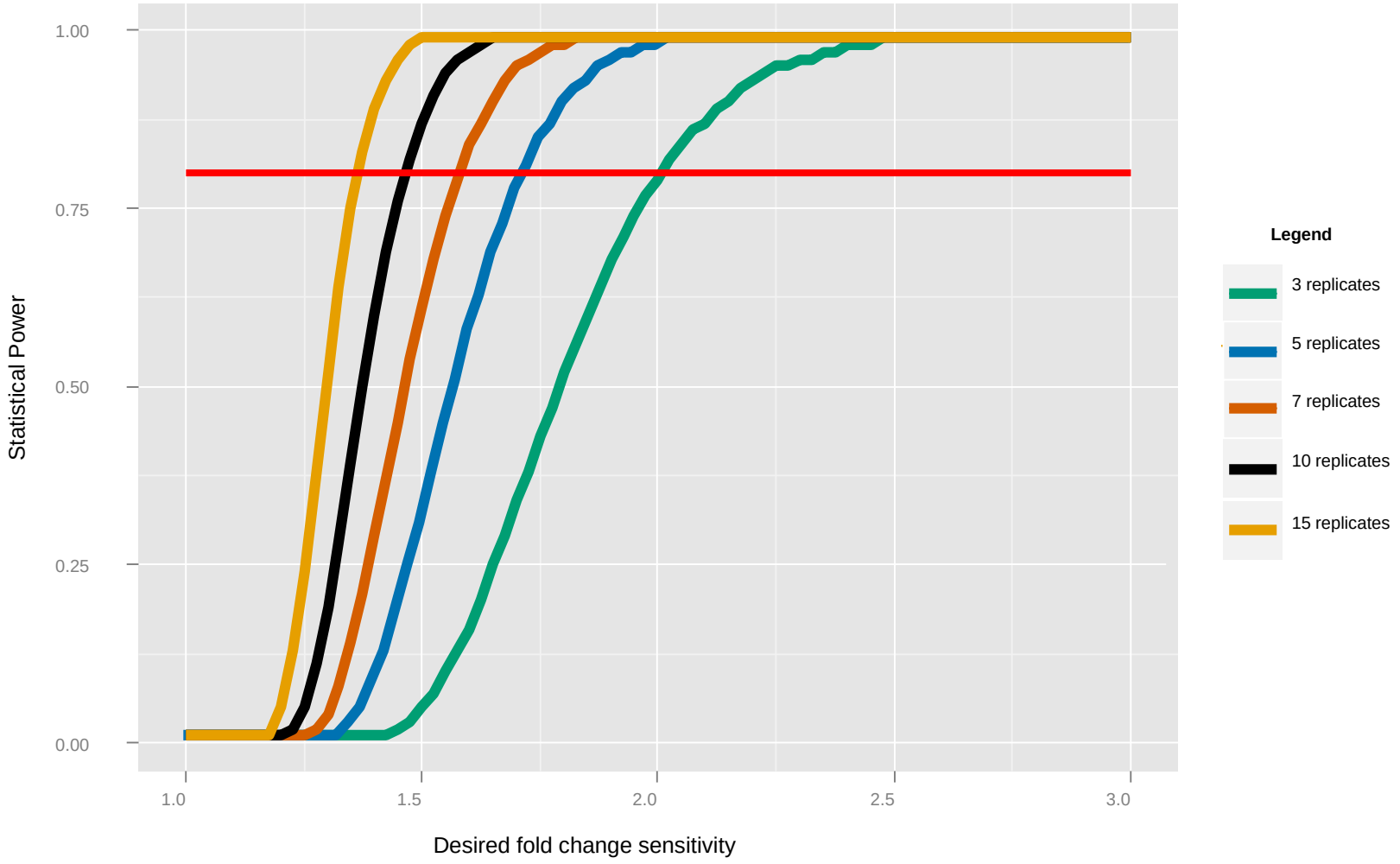
 # controls/# cases

25 **50** **100** **250**

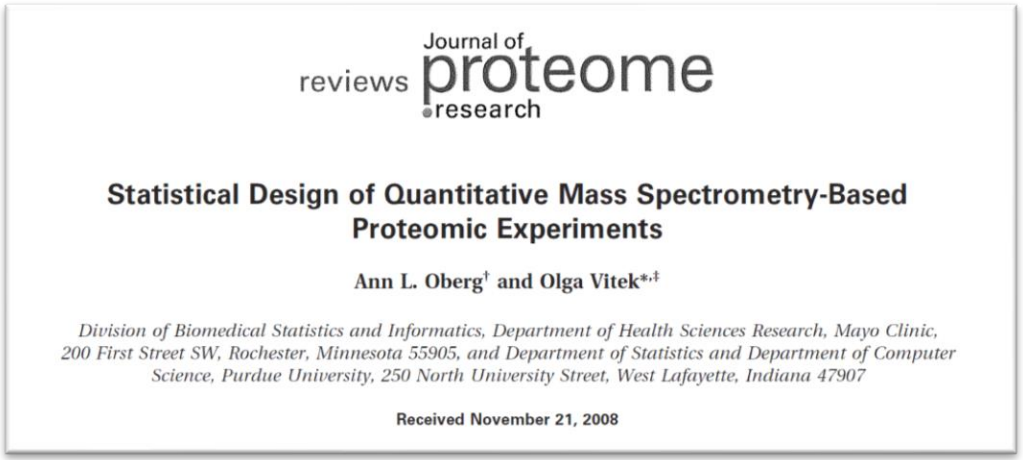
	Signal (Δ SD)	1.0	2.0	3.0	4.0	1.0	2.0	3.0	4.0	1.0	2.0	3.0	4.0	1.0	2.0	3.0	4.0
% Cases Producing Marker	10%	23.3%	34.1%	41.2%	49.2%	27.7%	39.0%	56.5%	63.6%	32.7%	55.6%	74.7%	83.1%	46.6%	79.3	94.8%	98.5%
	20%	33.5%	53.9%	67.7%	76.2%	43.0%	72.5%	85.6%	92.8%	58.2%	89.8%	97.9%	99.6%	82.3%	99.2%	100%	100%
	30%	45.9%	71.9%	85.8%	91.0%	57.8%	89.2%	97.2%	99.3%	78.5%	98.6%	100%	100%	97.1%	100%	100%	100%
	50%	69.2%	94.3%	99.1%	100%	87.6%	99.6%	100%	100%	97.9%	100%	100%	100%	100%	100%	100%	100%
	80%	90.7%	100%	100%	100%	99.5%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Skates et al 2013

Example of power analysis: yeast proteome response to perturbation examined by DIA-MS



For more information…



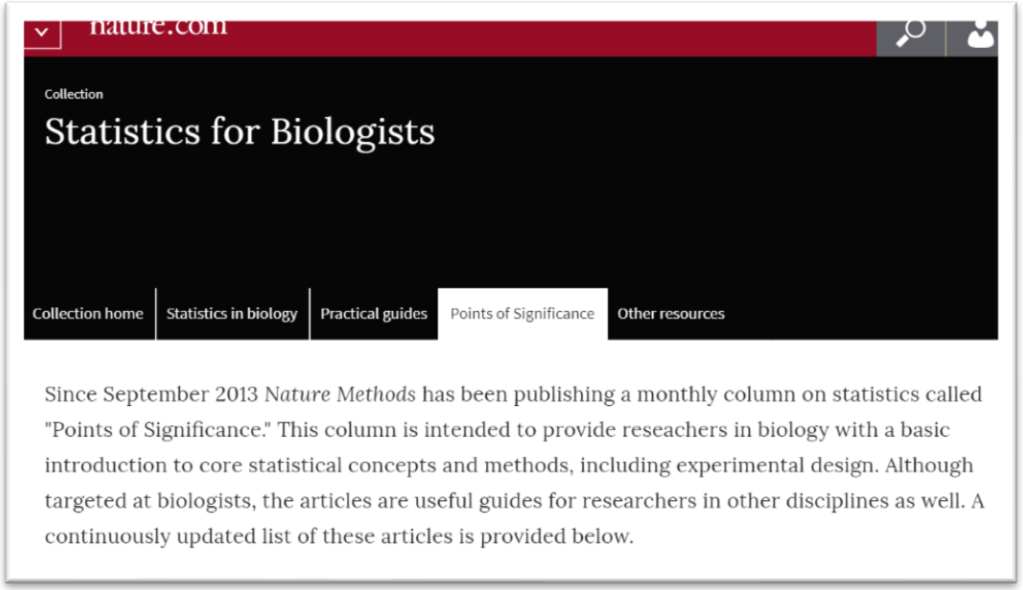
Journal of proteome reviews research

Statistical Design of Quantitative Mass Spectrometry-Based Proteomic Experiments

Ann L. Oberg† and Olga Vitek*,‡

Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, 200 First Street SW, Rochester, Minnesota 55905, and Department of Statistics and Department of Computer Science, Purdue University, 250 North University Street, West Lafayette, Indiana 47907

Received November 21, 2008



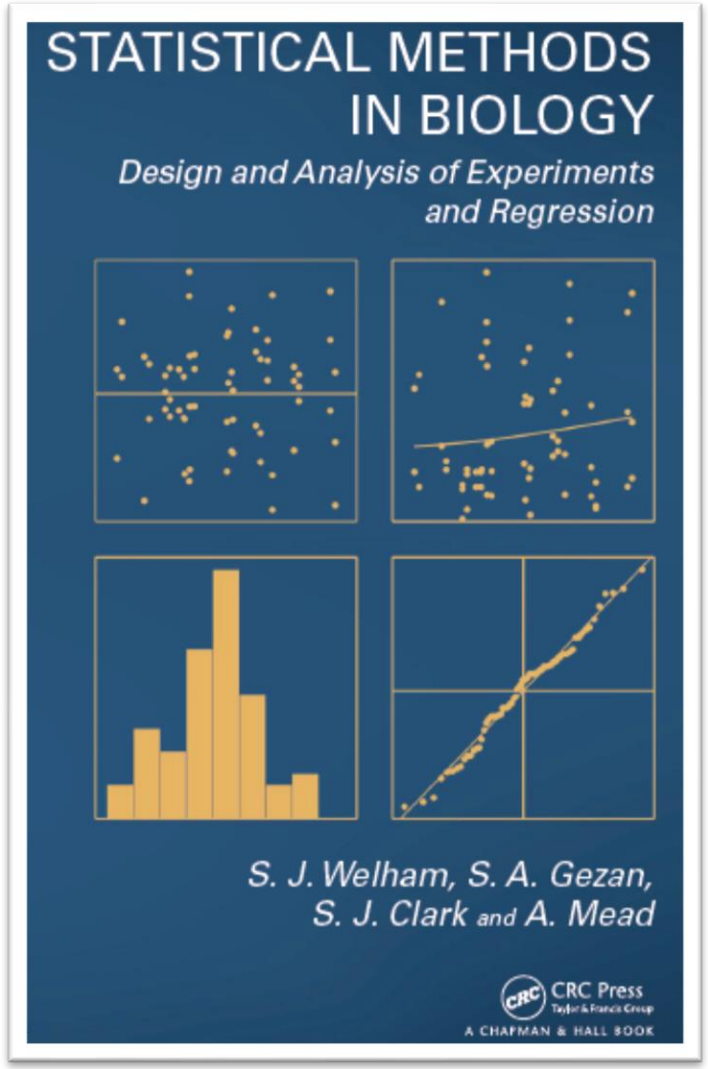
nature.com

Collection

Statistics for Biologists

Collection home | Statistics in biology | Practical guides | Points of Significance | Other resources

Since September 2013 *Nature Methods* has been publishing a monthly column on statistics called "Points of Significance." This column is intended to provide reseachers in biology with a basic introduction to core statistical concepts and methods, including experimental design. Although targeted at biologists, the articles are useful guides for researchers in other disciplines as well. A continuously updated list of these articles is provided below.



STATISTICAL METHODS IN BIOLOGY

Design and Analysis of Experiments and Regression

S. J. Welham, S. A. Gezan, S. J. Clark and A. Mead

CRC Press

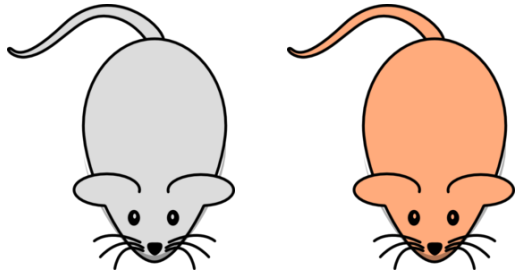
A CHAPMAN & HALL BOOK

Agenda

- Translating research experiments into statistical models
- Elements of good statistical/experimental design
  - Replication
  - Randomization
  - Blocking
  - Variation
- Differential abundance tests: the t-test and the ANOVA
- Statistical power: how many samples is enough and the problem with blanket fold change cut-offs

Earlier Today: Group Comparisons natively within Skyline

Example 1: plasma proteomics in healthy vs salt-sensitive rat



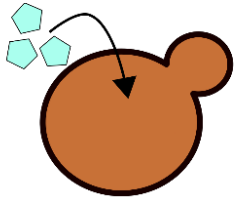
7 Healthy 7 Disease

3 technical replicates

- What statistical analysis is most appropriate?
- How to block the sample preparation?

Example 2: effect of genotype + treatment on cellular aging

5 genotypes

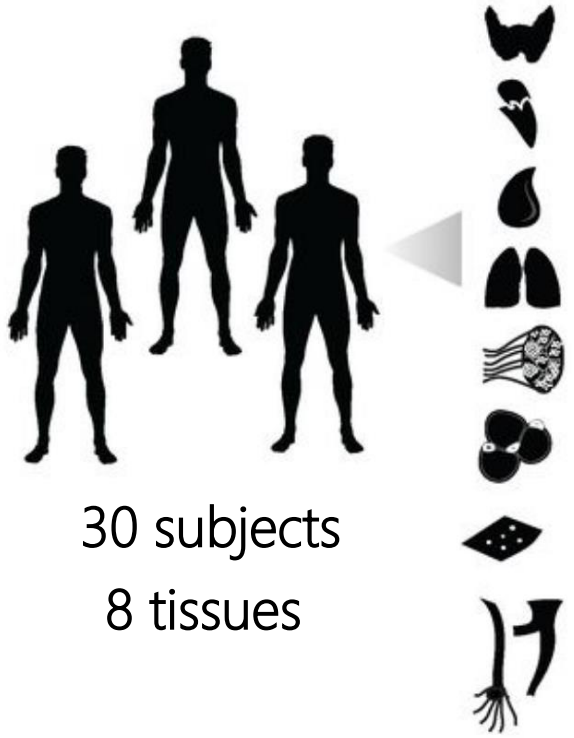


5 treatments

5 biological replicates

- What statistical analysis is most appropriate?
- How to block the sample preparation?

Example 3: tissue-specific proteomics atlas



30 subjects

8 tissues

- What statistical analysis is most appropriate?
- How to block the sample preparation?