Experimental design

Charlotte Soneson
University of Zurich
Brixen 2017

“To call in the statistician
after the experiment is done
may be no more than asking him
to perform a postmortem examination:
he may be able to say what the
experiment died of.”

Sir Ronald Fisher, Indian Statistical Congress, Sankhya, around 1938

Different types of experiments

Learning experiment questions	Confirming experiment questions
• Does the drug have toxic side effects (at what dose, given for how long, in which tissue)?	• Does 5 mg/kg of the drug given once a day for 5 days increase blood creatinine[a] concentration?
• Does stress affect rodent behaviour (what kind of stress, for how long, on what behavioural tasks)?	• Does fox urine odour (a stressor) affect the amount of food Wistar rats consume during the first 24 hours after exposure?
• How dose exercise affect cognitive functioning of older people (what type of exercise, how much, which aspect of cognition)?	• Does 30 min of aerobic activity (treadmill running) at 60% VO_2 max[b], 3 days a week for 6 weeks, in males between 55–70 years of age, improve performance on a mental rotation task?

[a] Increased creatinine indicates kidney damage.
[b] VO_2 max is the maximal oxygen uptake and is a measure of a person's aerobic fitness.

[Lazic, 2016]

What is experimental design?

- The organization of an experiment, to ensure that the **right type** of data, and **enough** of it, is available to answer the **questions of interest** as clearly and efficiently as possible.

http://www.stats.gla.ac.uk/steps/glossary/anova.html#expdes

What characterizes well-designed experiments?

- Effects can be estimated unambiguously and without bias.
- Estimates are precise.
- Protected from possible one-off events that might compromise the results.
- Easy to conduct.
- Easy to analyse and interpret.
- Maximum information obtained for fixed time, resources, and samples.
- Applicability of the findings to a wide variety of subjects, conditions, and situations.

[Lazic, 2016]

What affects the outcome of an experiment?

$$\text{Outcome} = \underbrace{\text{Treatment effects}}_{\substack{\text{Environment}\\ \text{Compound}\\ \text{Inhibitor}\\ \text{siRNA}\\ \text{Dose}\\ \text{Time}}} + \underbrace{\text{Biological effects}}_{\substack{\text{Sex}\\ \text{Age}\\ \text{Weight}\\ \text{Litter}\\ \text{Genotype}\\ \text{Species}\\ \text{Cell line}}} + \underbrace{\text{Technical effects}}_{\substack{\text{Technician}\\ \text{Batch}\\ \text{Plate}\\ \text{Cage}\\ \text{Array}\\ \text{Day}\\ \text{Order}\\ \text{Source}}} + \underbrace{\text{Error}}_{\substack{\text{Experimental}\\ \text{Treatment}\\ \text{Sampling}\\ \text{Measurement}}}$$

[Lazic, 2016]

What is **bad** experimental design?

Treatment I

M M M M M M M M

Treatment II

F F F F F F F F

What is bad experimental design?

Treatment 1

M M M M M M

Treatment

F F F F F F

Confounding!

What is bad experimental design?

Analysis batch I / Study center I / Processing protocol I ...

Tr Tr Tr Tr Tr Tr Tr Tr

Analysis batch II / Study center II / Processing protocol II ...

Ctl Ctl Ctl Ctl Ctl Ctl Ctl Ctl

What is bad experimental design?

Analysis batch I / Study center I / Pr… ing protocol I ...

Tr Tr Tr Tr Tr … Tr

Confounding!

Analysis batch II … er II / Processing protocol II ...

Ctl C… Ctl Ctl Ctl Ctl Ctl

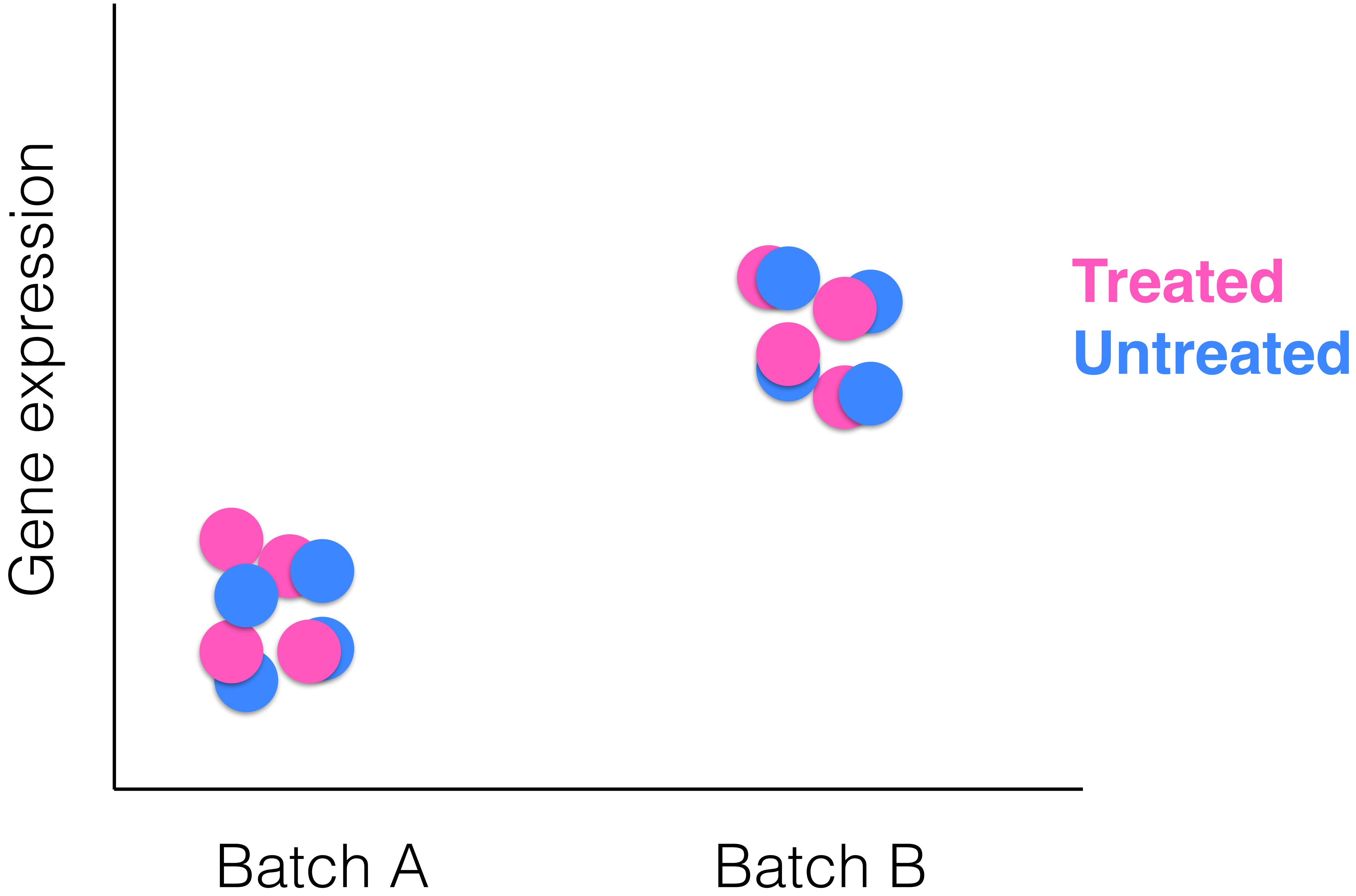
What would be a better experimental design?

- Process all samples at the same time/in one batch (not always feasible)

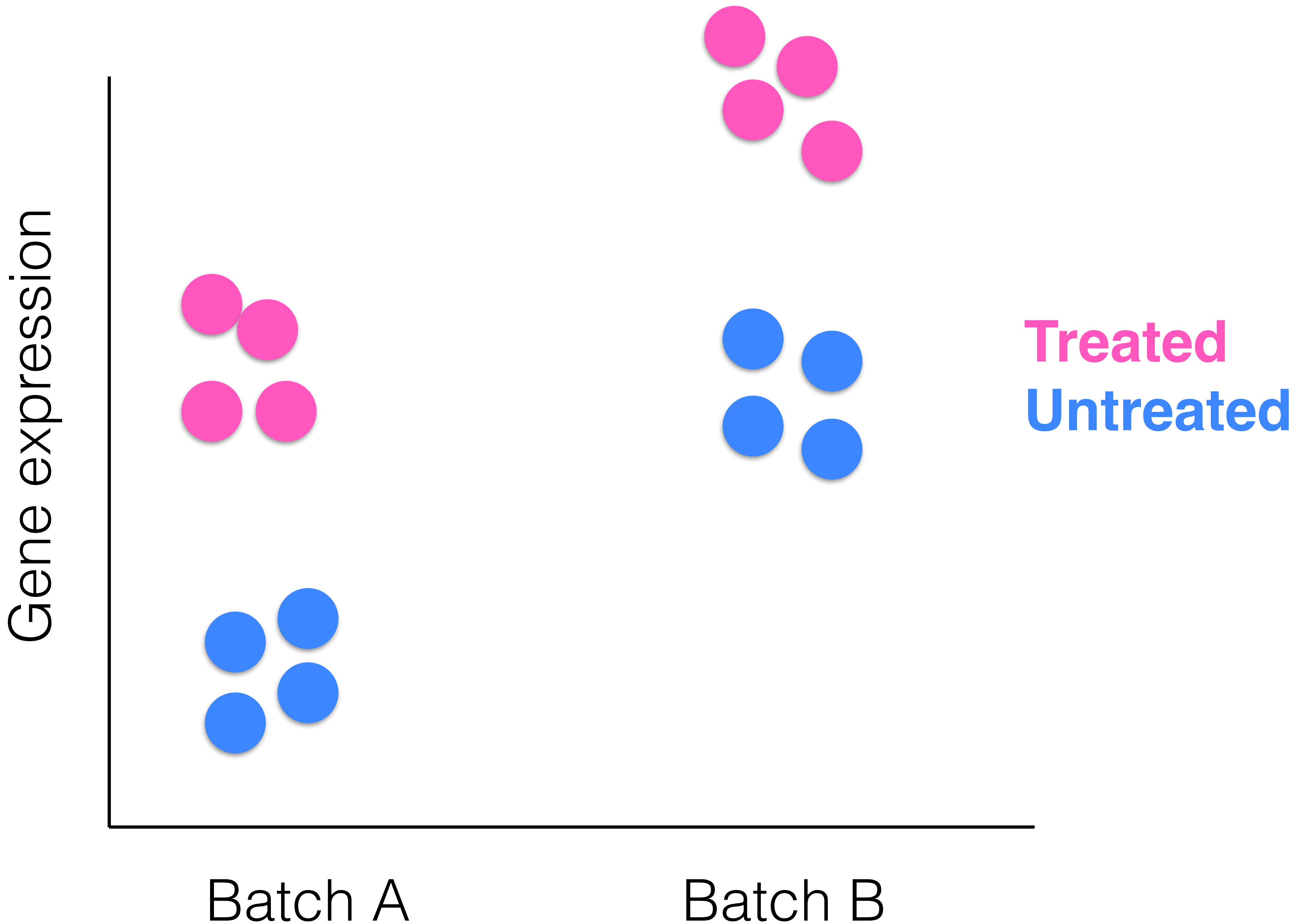
- Minimize confounding as much as possible through
  - blocking
  - randomization

- The batch effect will still be there, but with an appropriate design we can account for it

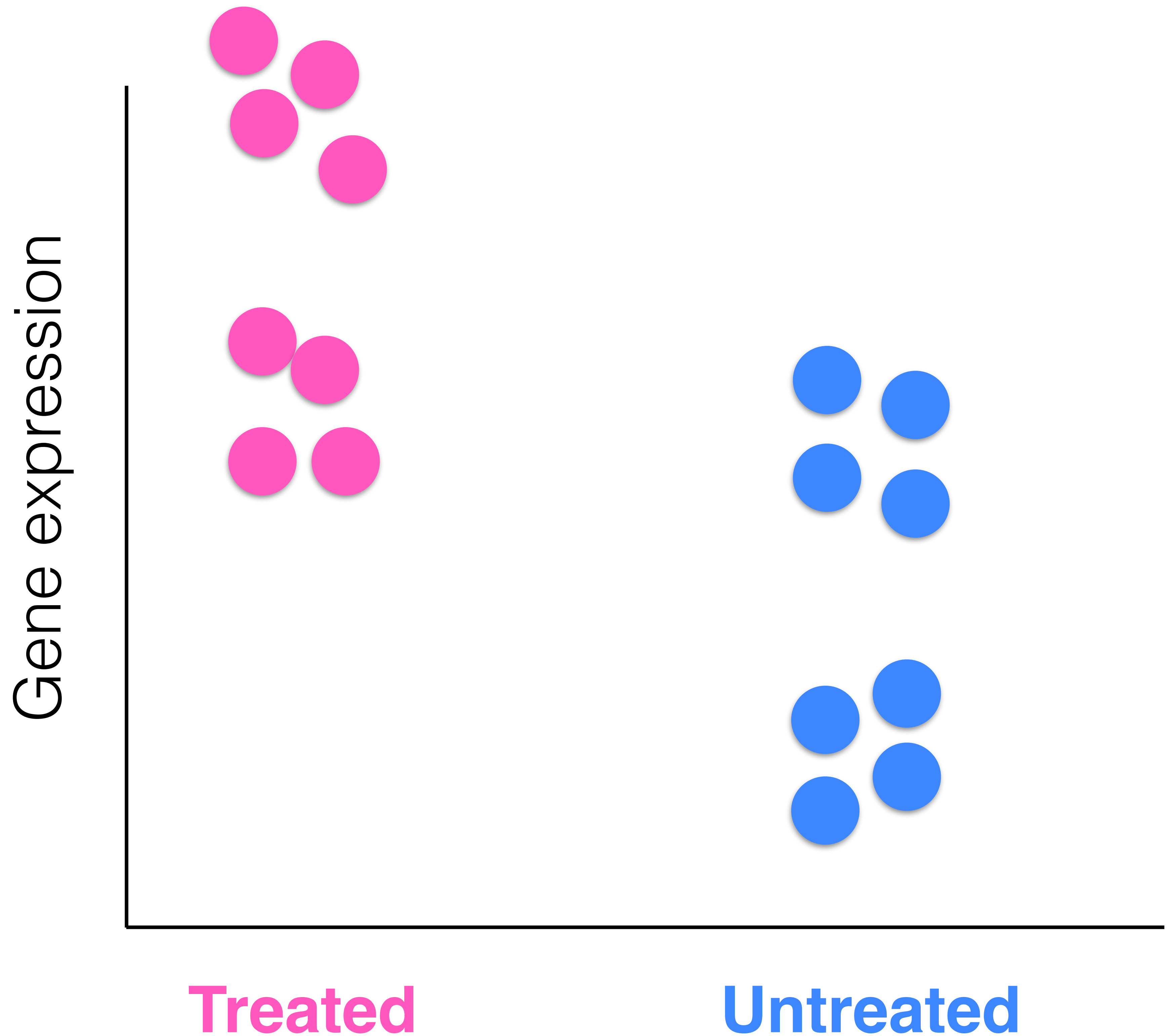
Non-confounded design



Non-confounded design



Non-confounded design



Accounting for batch effects

- In statistical modeling, batch effects can be included as **covariates** (additional predictors) in the model.

- For exploratory analysis, we often attempt to "eliminate" or "adjust for" such unwanted variation in advance, by subtracting the estimated effect from each variable.

- Even partial confounding between batch and signal of interest can lead to bias.

Nygaard et al., Biostatistics 2016

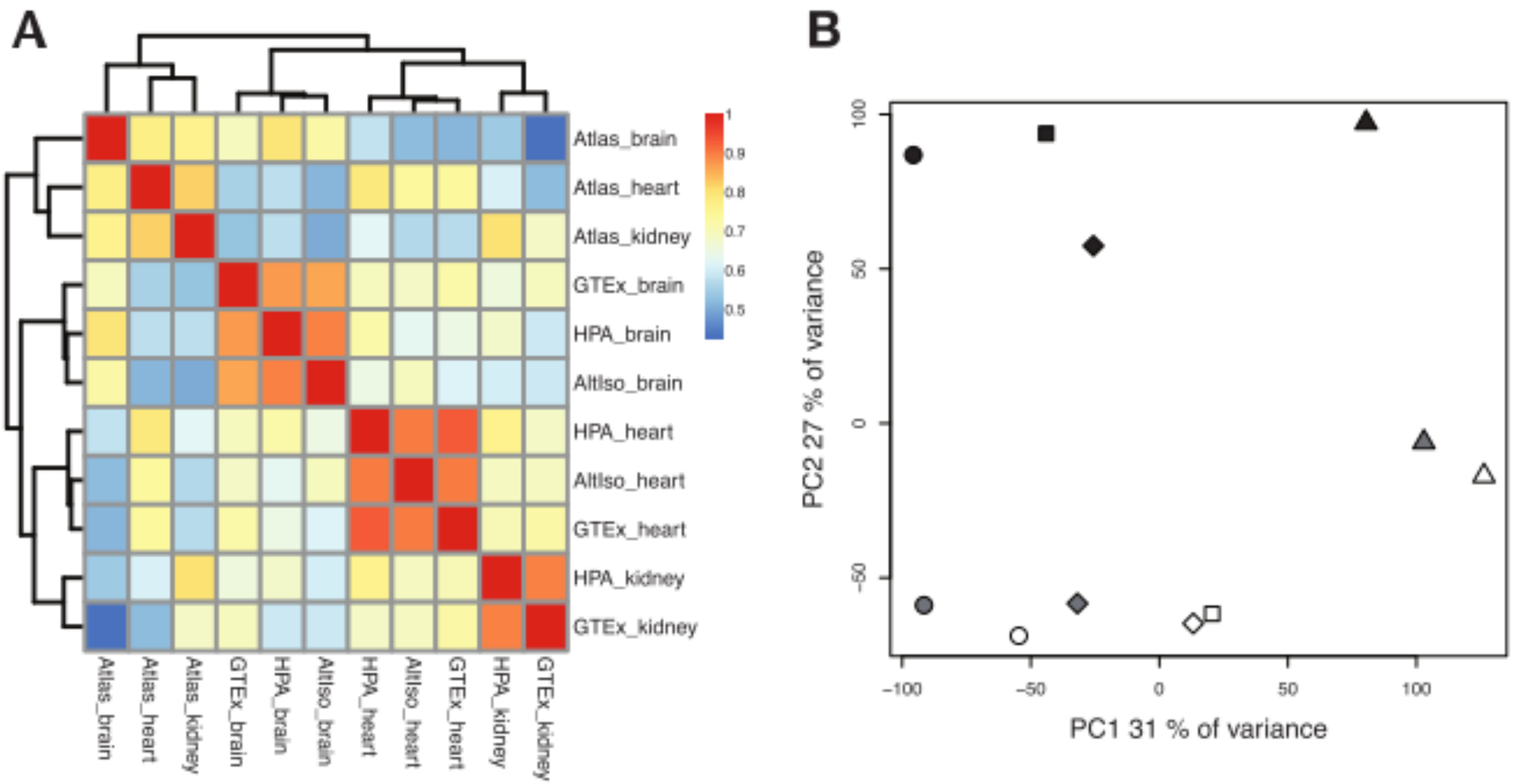
ComBat

- “Eliminate” the impact of the (known) batch variable on the observed values
- Can provide information about variables of interest, whose effect should be retained in the data
- Works best if batch and variable of interest are not confounded

Johnson et al., Biostatistics 2007

Accounting for batch effects in practice

- Public, processed RNA-seq data from 3 tissues, 4 studies show strong association with study

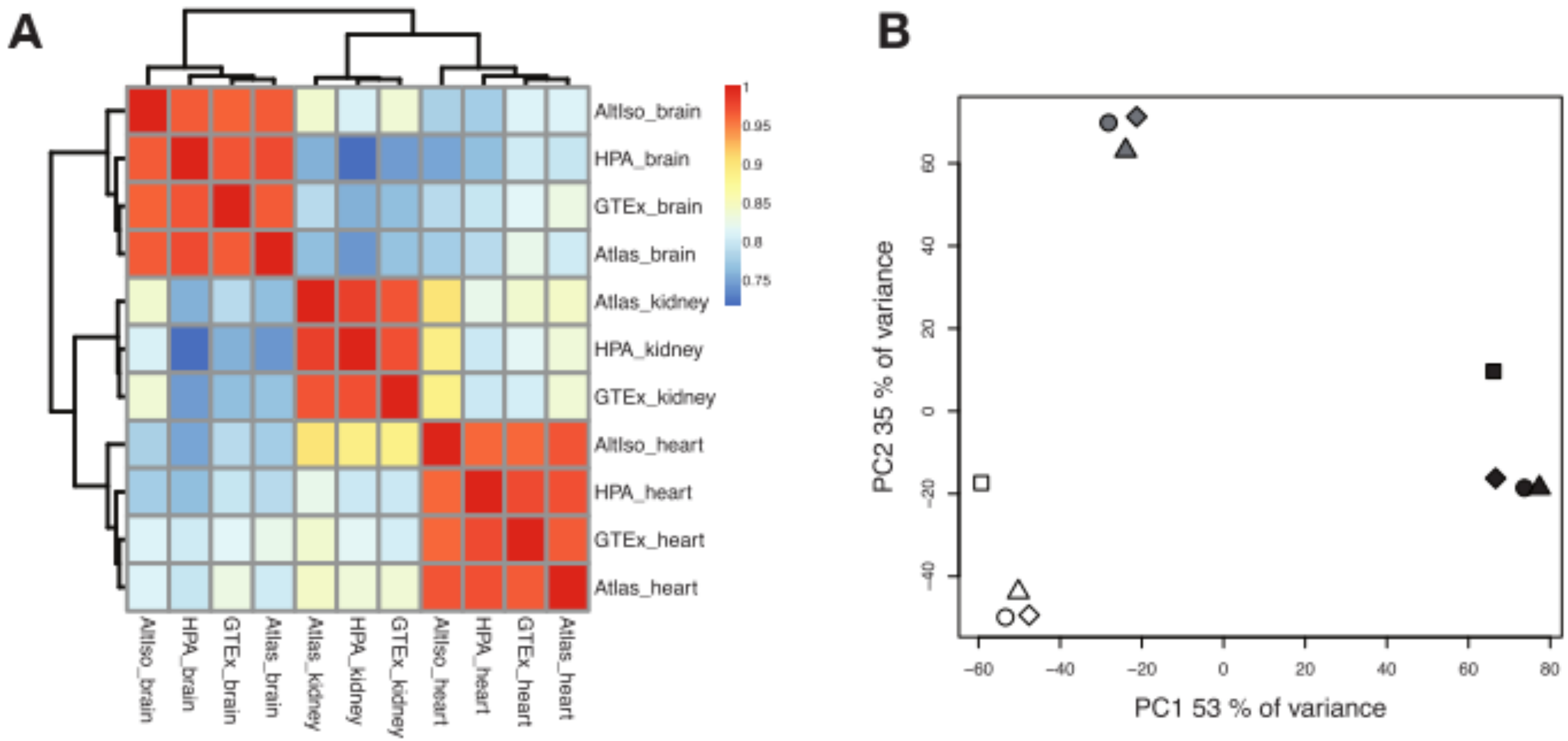


color = tissue; symbol = study (batch)

Danielsson et al., Briefings in Bioinformatics 2015

Accounting for batch effects in practice

- Accounting for the batch effect brings out signal of interest.

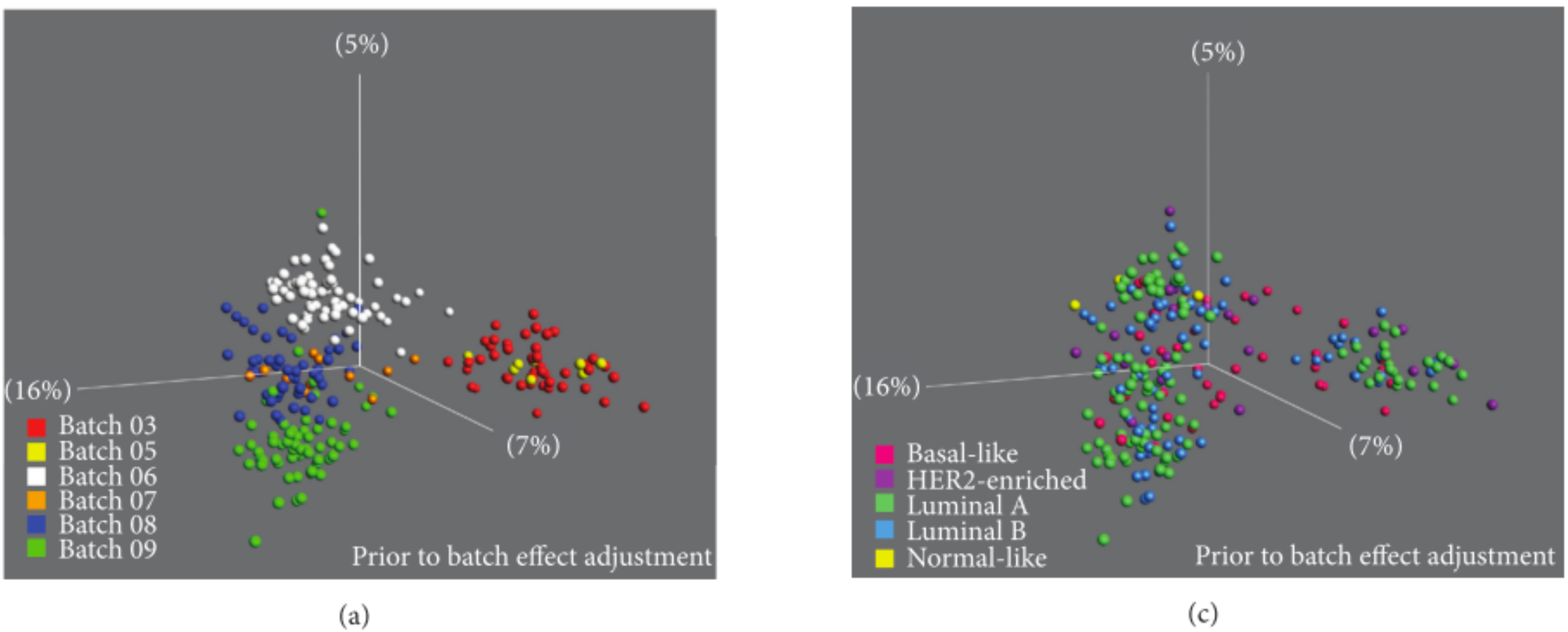


color = tissue; symbol = study (batch)

Danielsson et al., Briefings in Bioinformatics 2015

Accounting for batch effects in practice

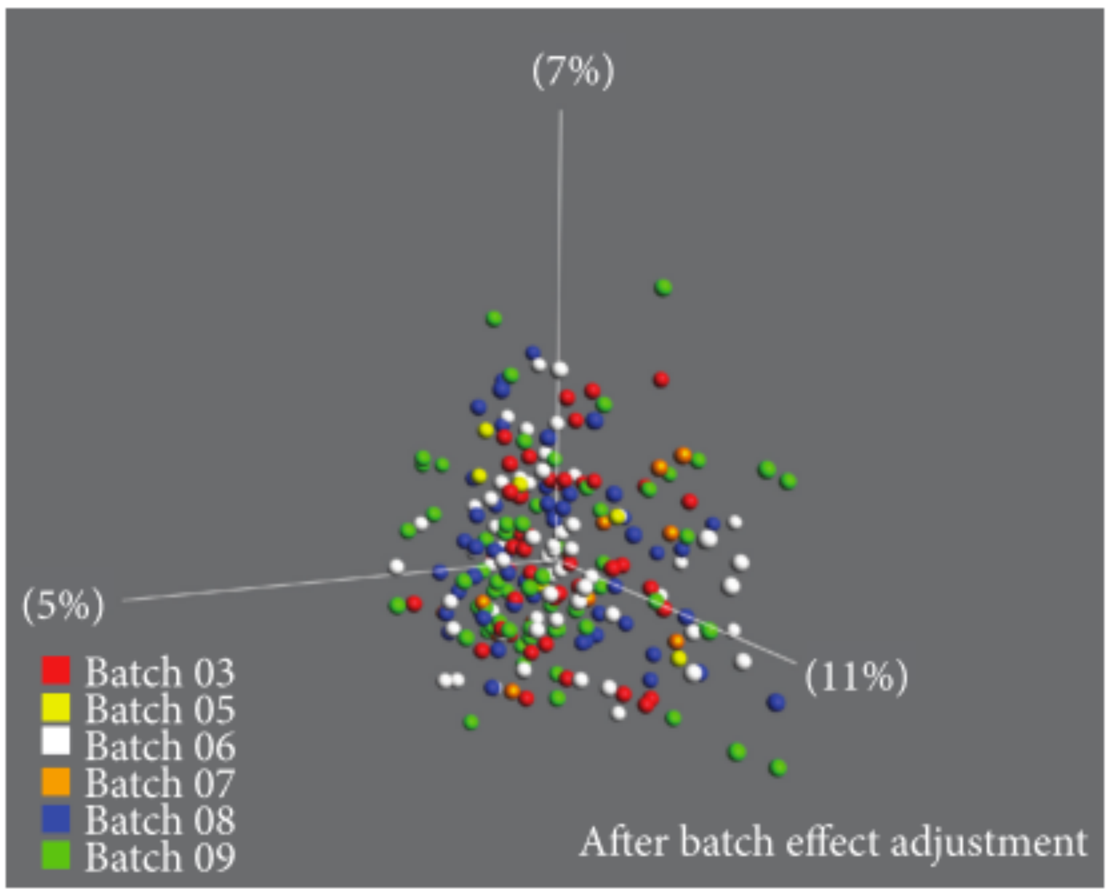
- 5-subtype breast cancer microarray data processed in six batches.



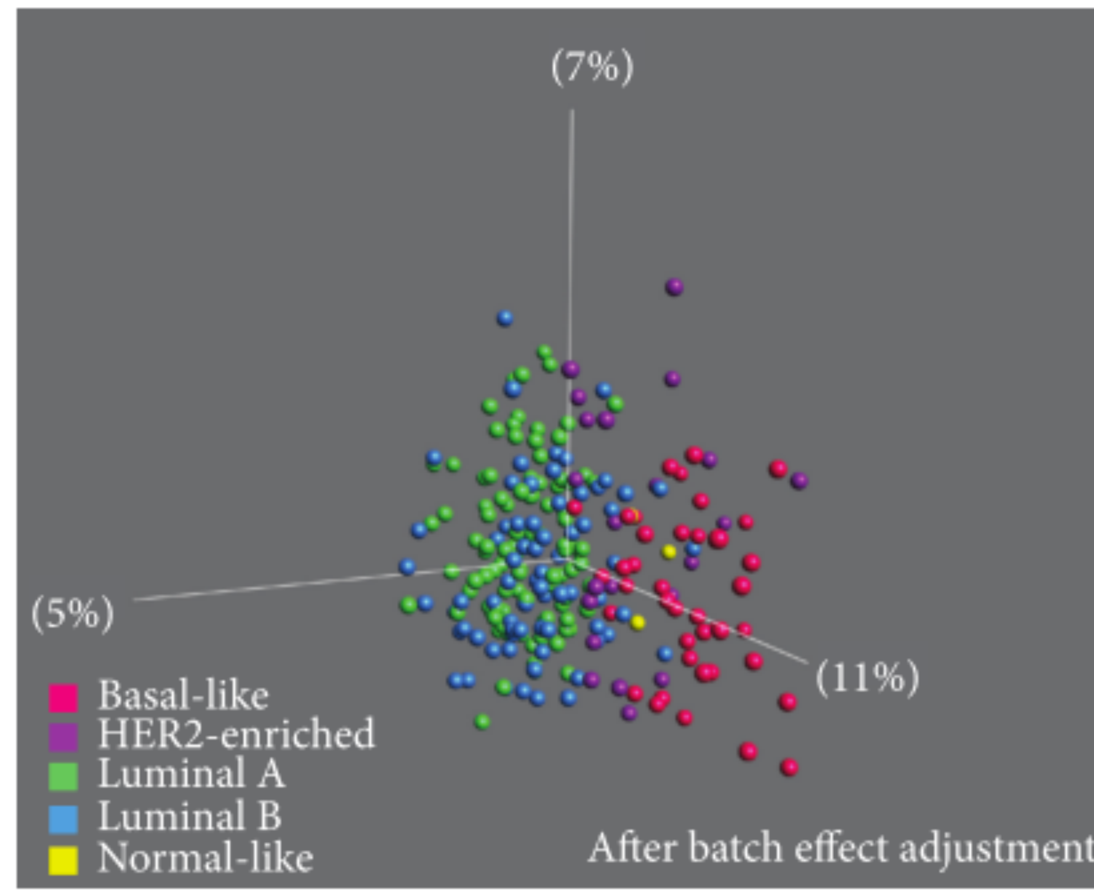
Larsen et al., BioMed Research International 2014

Accounting for batch effects in practice

- 5-subtype breast cancer microarray data processed in six batches.



(b)



(d)

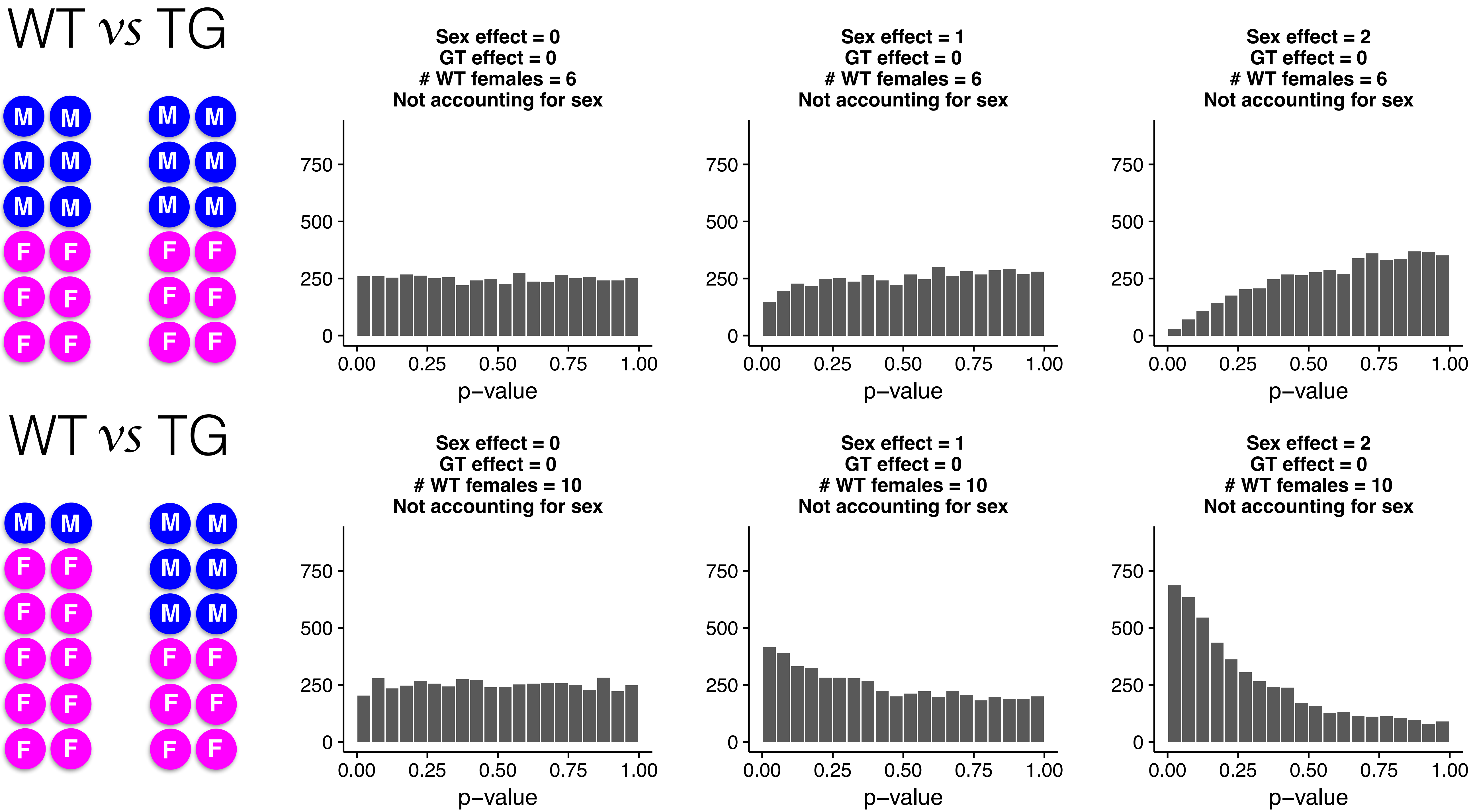
Larsen et al., BioMed Research International 2014

What if the batch variable is unknown?

- Manifests as systematic "unwanted variation" in data
- Identify using e.g.
  - control genes ("housekeeping" genes, spike-ins)
  - residuals after eliminating known signal
- Include estimated unwanted variation as covariate(s) in the statistical model
- **RUV**, **sva** packages commonly used in genomics

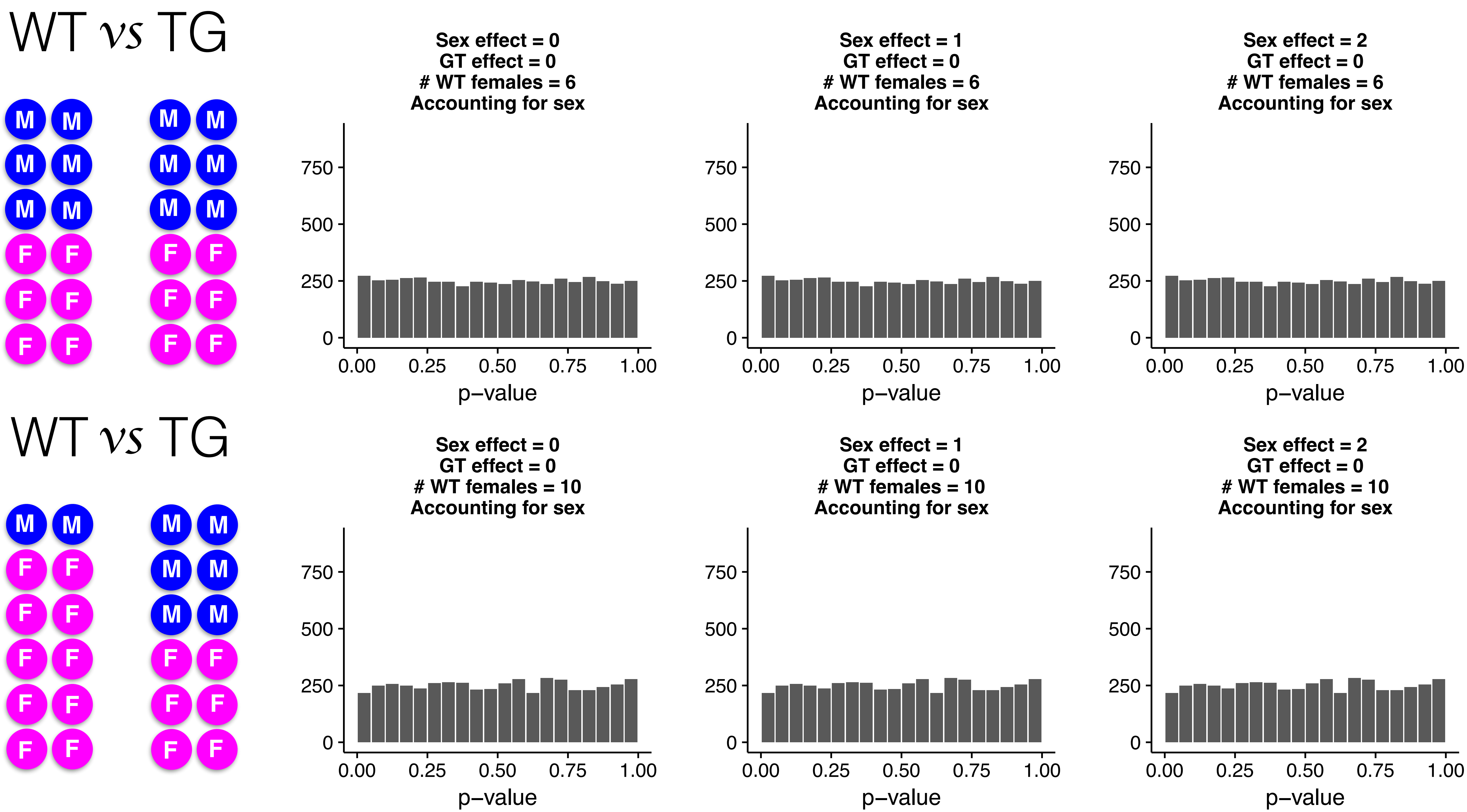
Leek & Storey (2007); Gagnon-Bartsch & Speed (2012); Leek (2014); Risso et al (2014)

Impact of batch effect on p-value histogram



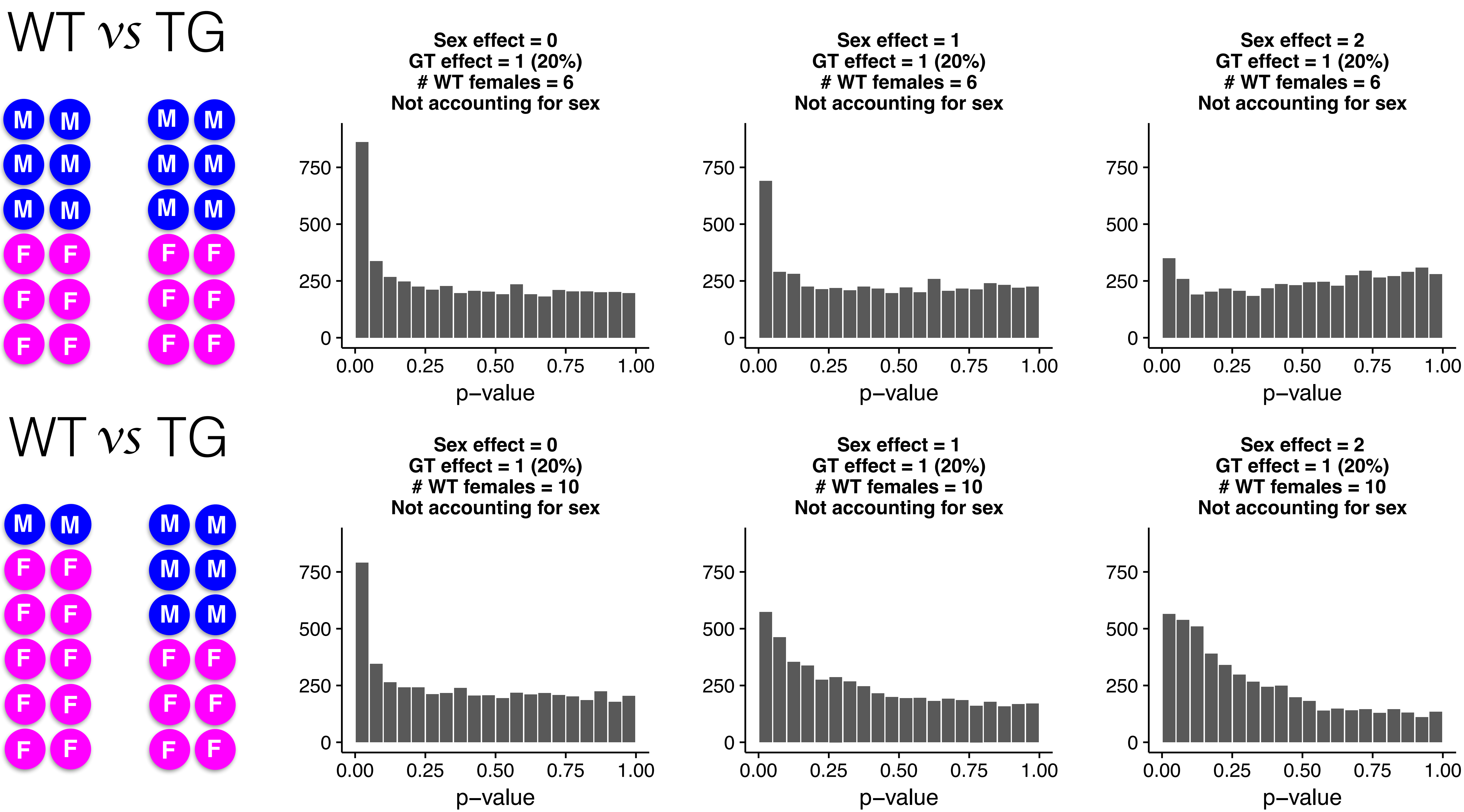
[adapted and extended from Lazic, 2016]

Impact of batch effect on p-value histogram



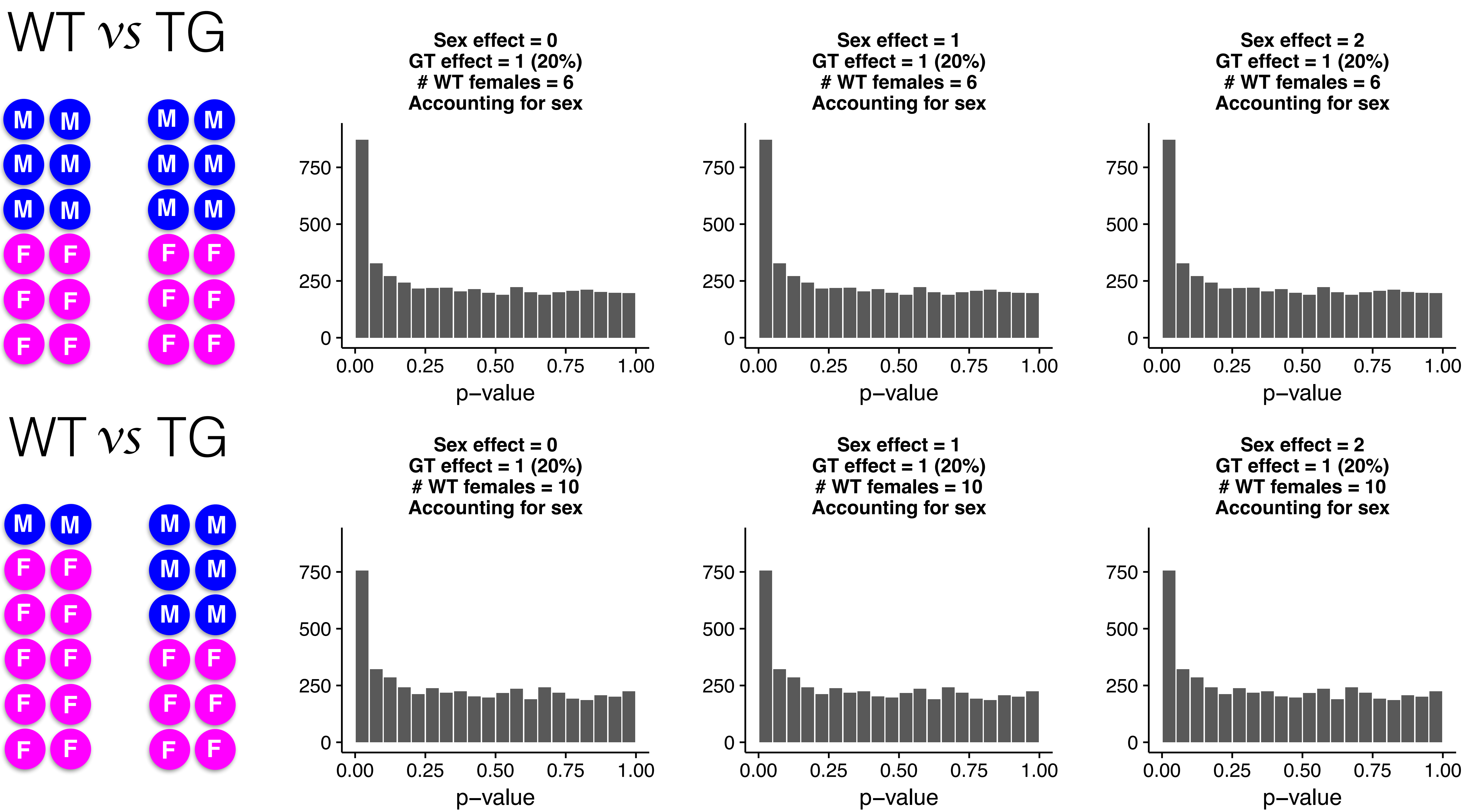
[adapted and extended from Lazic, 2016]

Impact of batch effect on p-value histogram



[adapted and extended from Lazic, 2016]

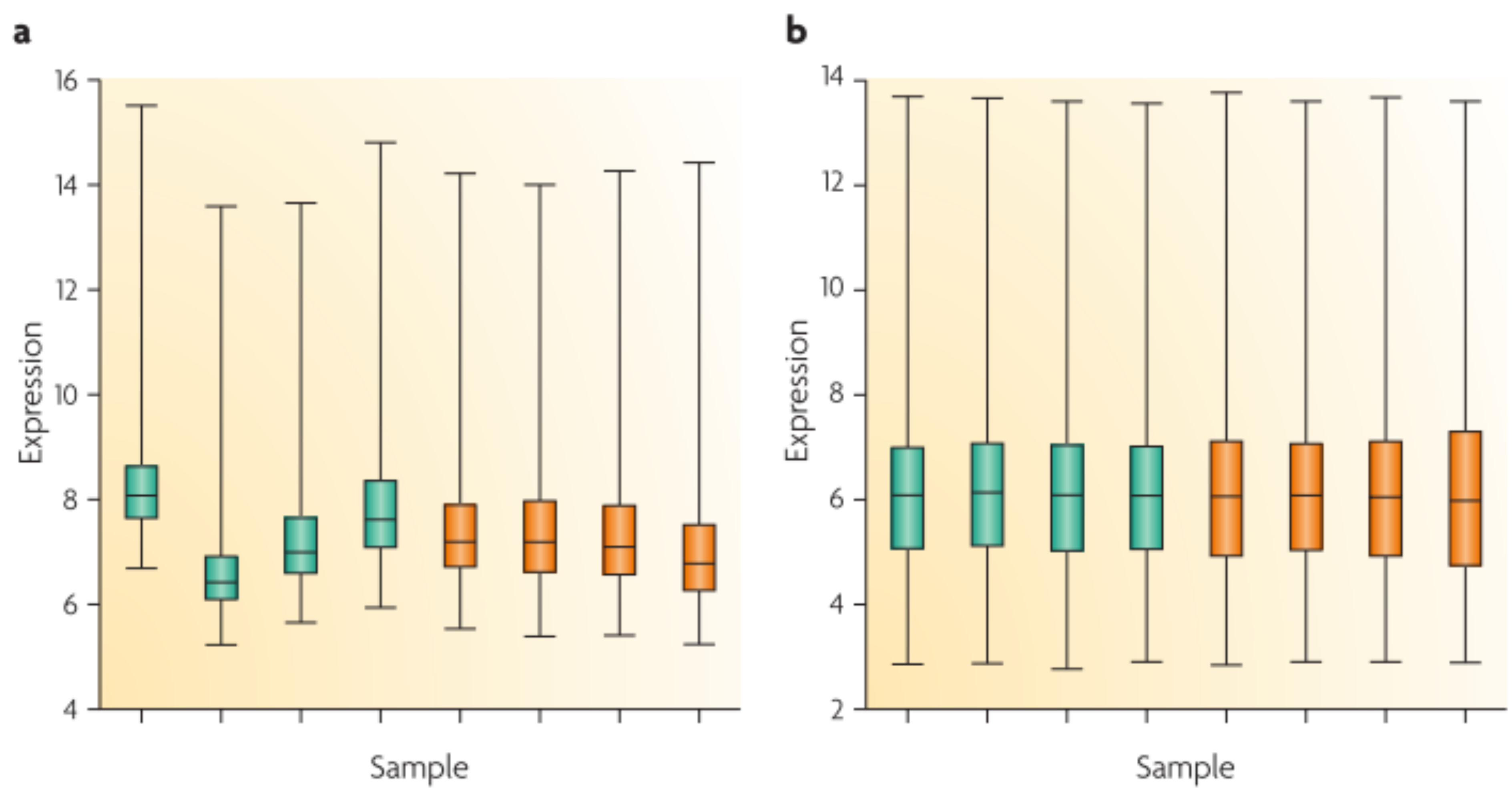
Impact of batch effect on p-value histogram



[adapted and extended from Lazic, 2016]

Batch effect adjustment vs normalization

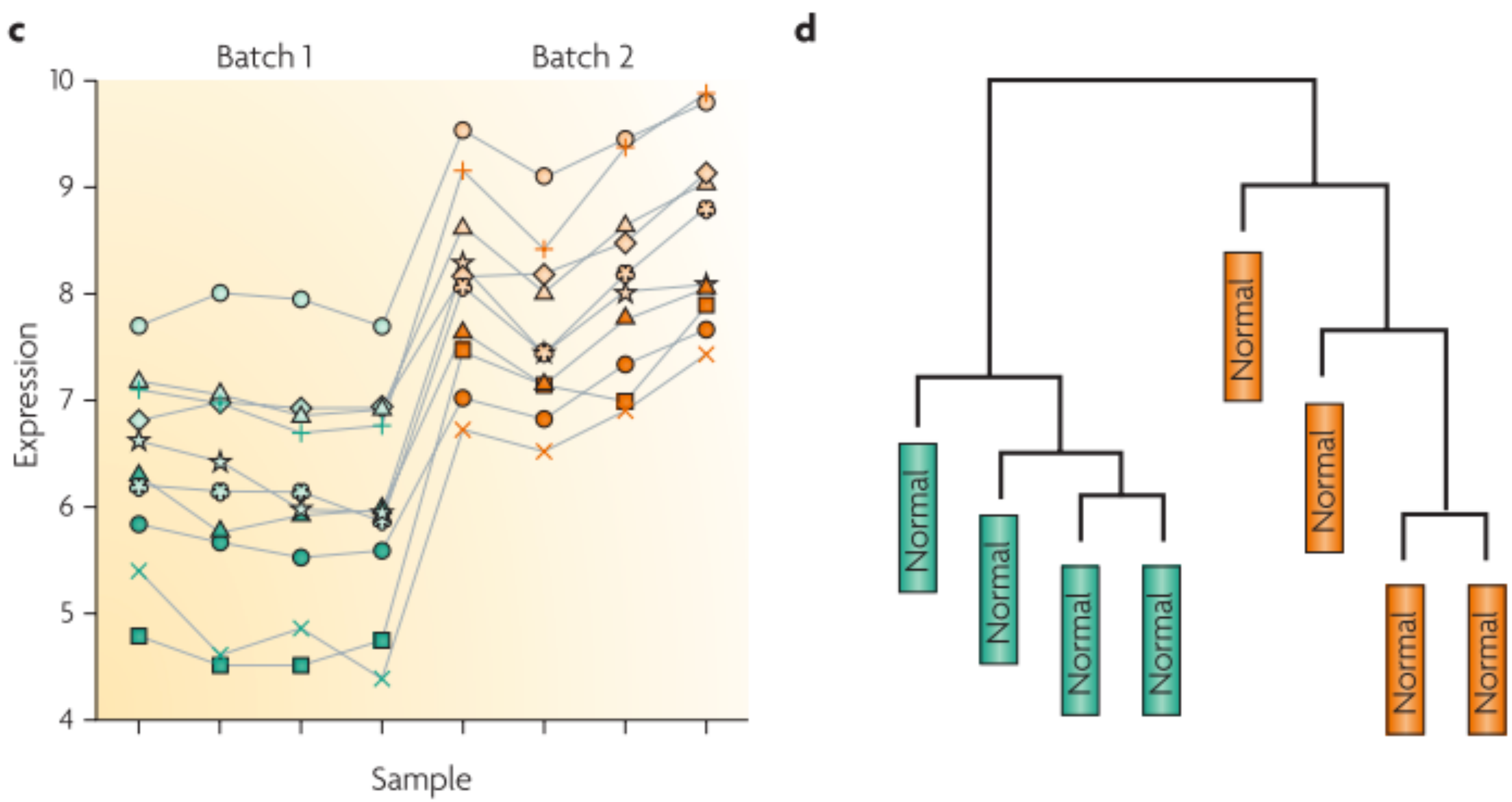
- Batch effect adjustment goes **beyond** the “global” between-sample normalization methods.



Leek et al., Nature Reviews Genetics 2010

Batch effect adjustment vs normalization

- Batch effect adjustment goes **beyond** the “global” between-sample normalization methods.

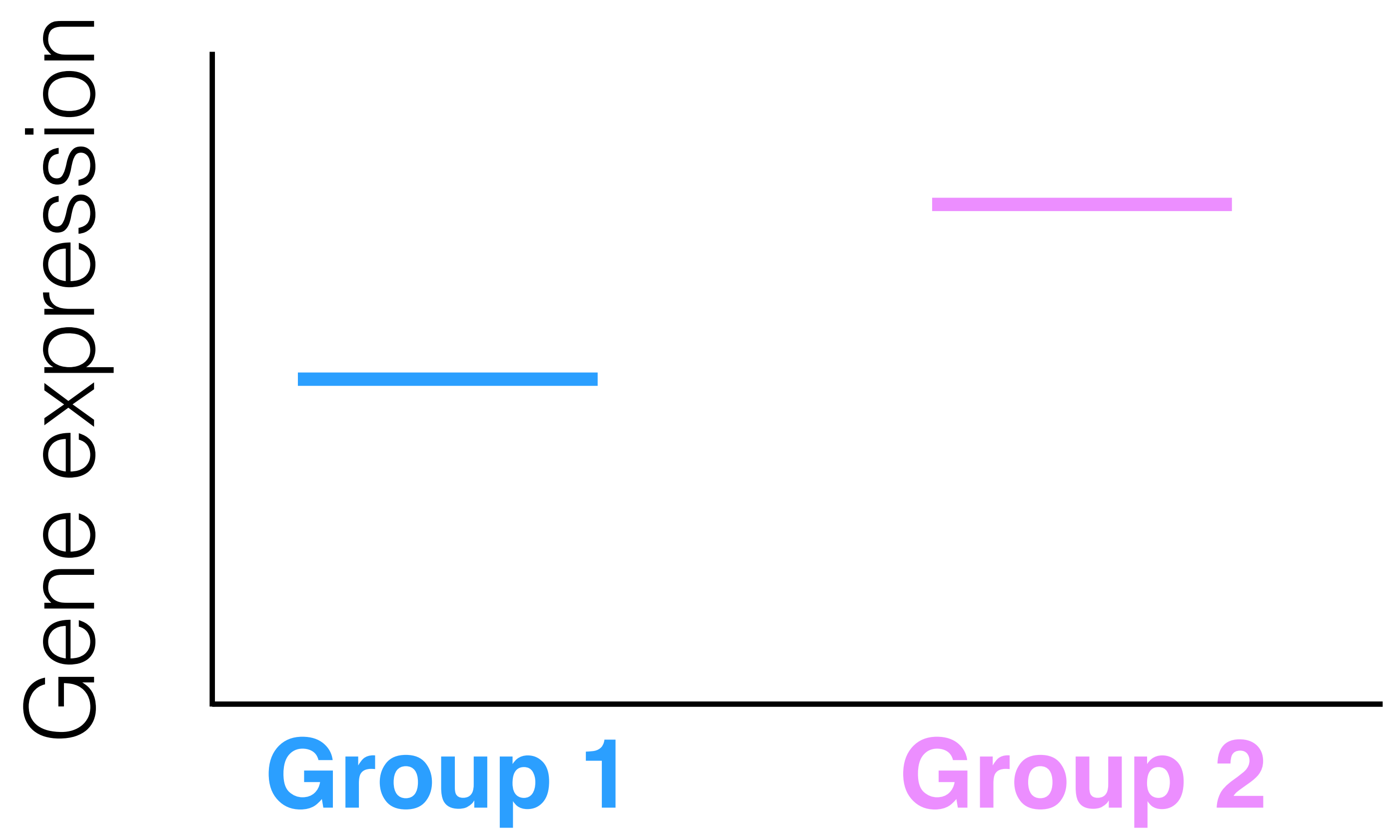


Leek et al., Nature Reviews Genetics 2010

Other design issues: replication

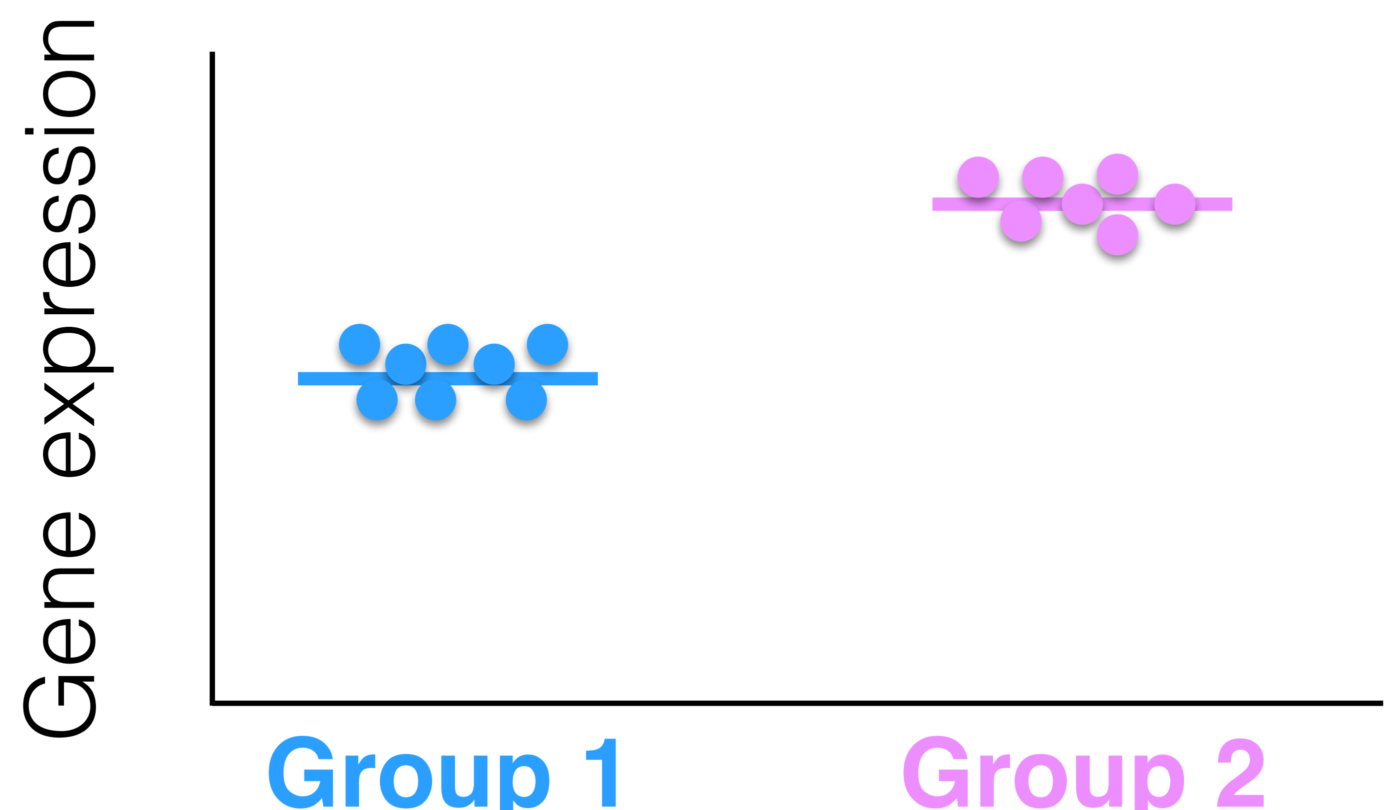
- Replicates are **necessary** to estimate within-condition variability.

- Variability estimates are, in turn, **vital** for statistical testing.



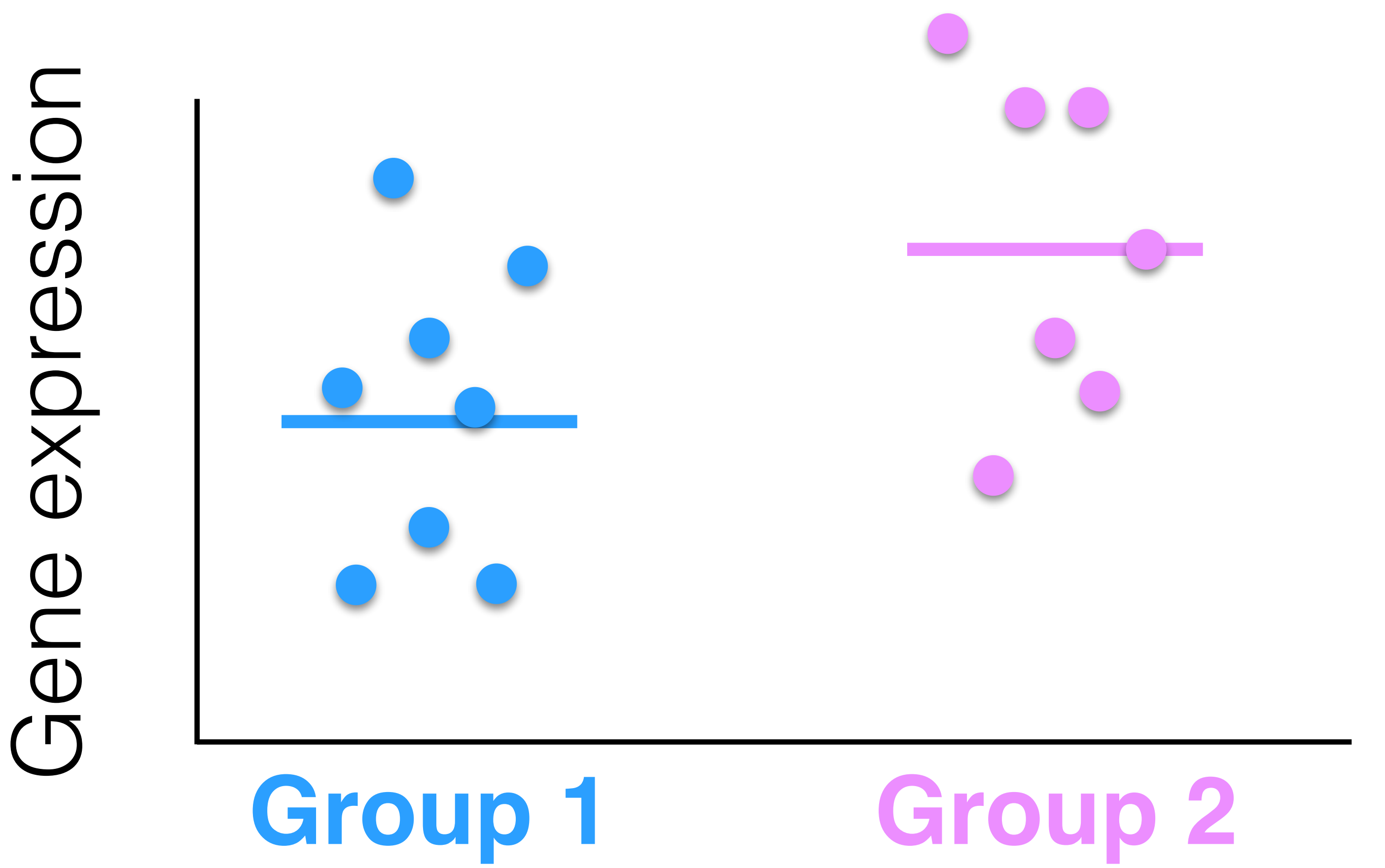
Other design issues: replication

- Replicates are **necessary** to estimate within-condition variability.
- Variability estimates are, in turn, **vital** for statistical testing.



Other design issues: replication

- Replicates are **necessary** to estimate within-condition variability.
- Variability estimates are, in turn, **vital** for statistical testing.



Different types of units

- Biological units (BU) - entities we want to make inferences about (e.g., animal, person)
- Experimental units (EU) - smallest entities that can be independently assigned to a treatment (e.g., animal, litter, cage, well)
- Observational units (OU) - entities at which measurements are made

[Lazic, 2016]

Pseudoreplication

- "**Artificial inflation** of the sample size, that usually occurs when the biological unit of interest differs from the experimental unit or observational unit."
- Only replication of experimental units is true replication

[Lazic, 2016]

Other design issues: sample size

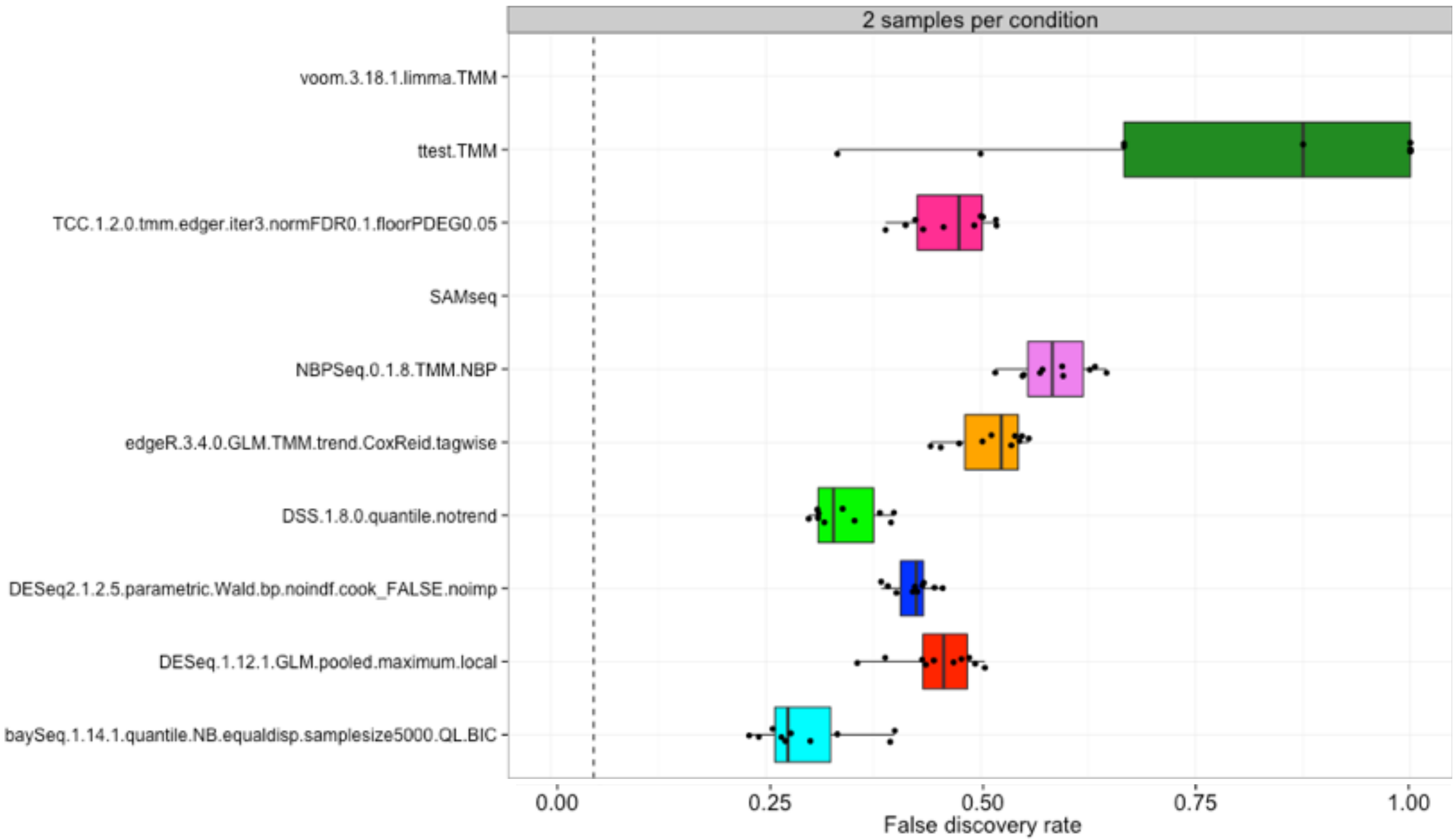
- As always, it depends…

  - on what we want to do (differential gene expression, variant detection, GWAS, …)

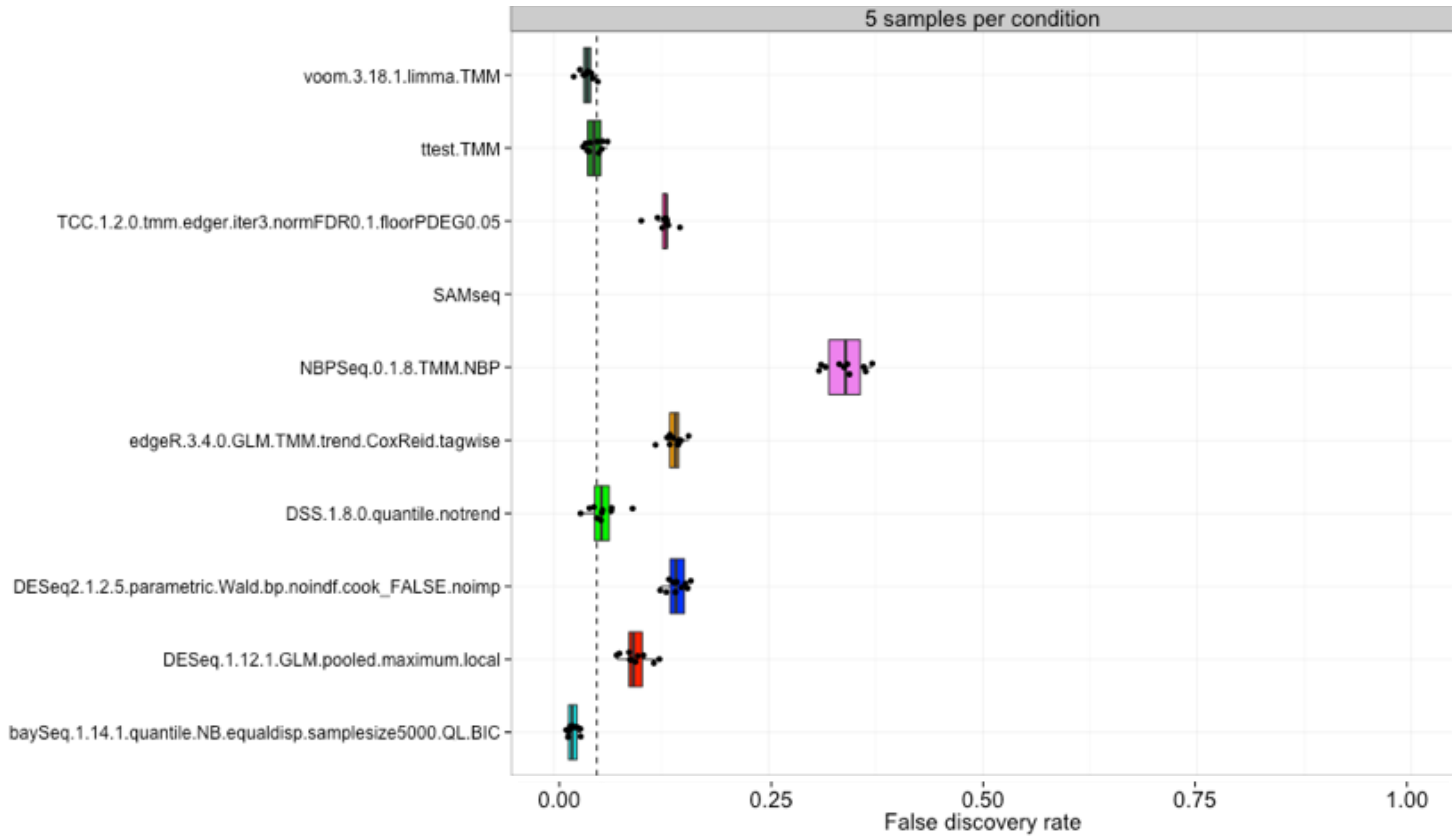
  - on the variability between samples (cell lines, inbred animals, patients, …)

  - on the magnitude of the expected effect

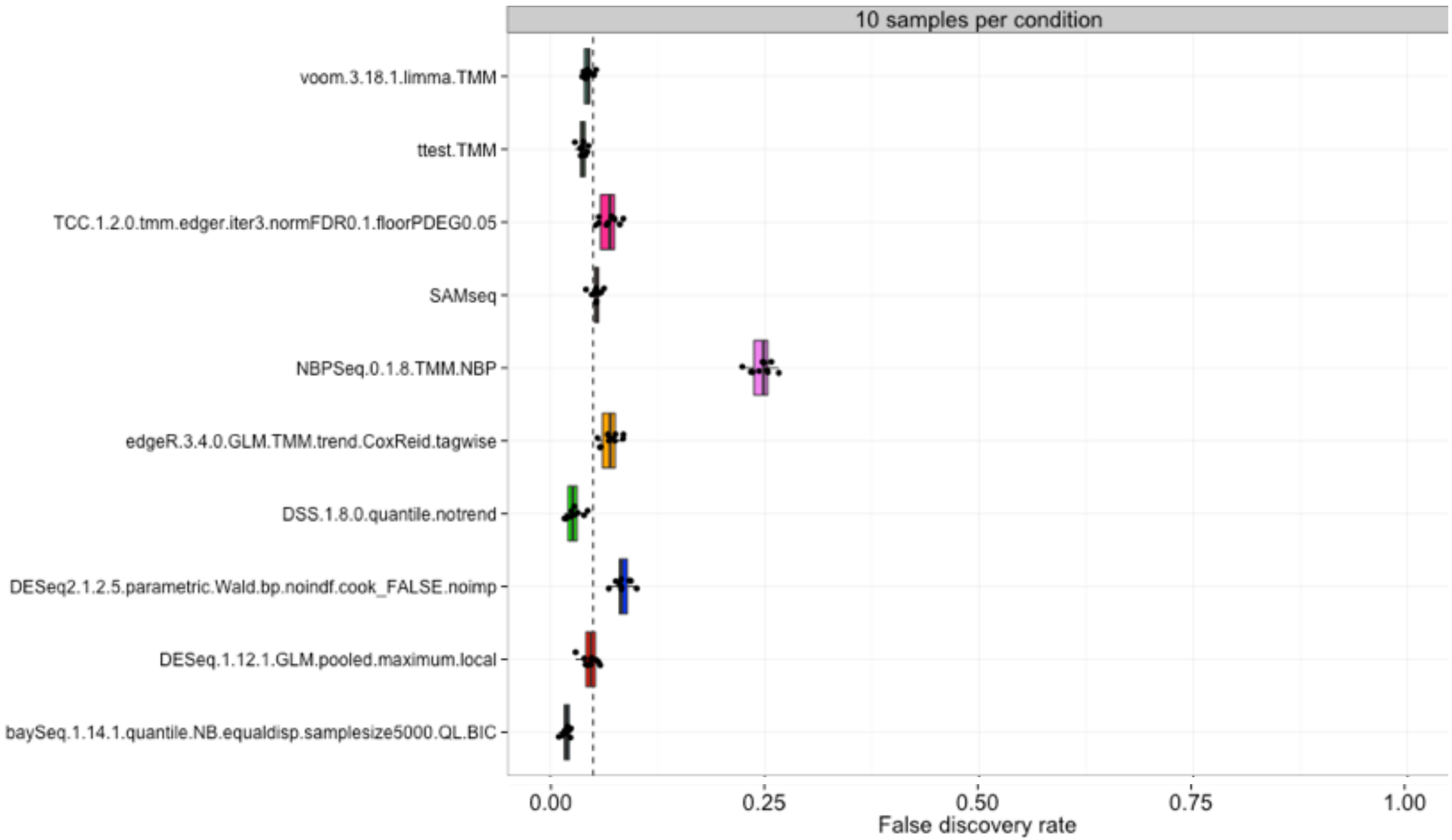
FDR, 2 replicates/condition



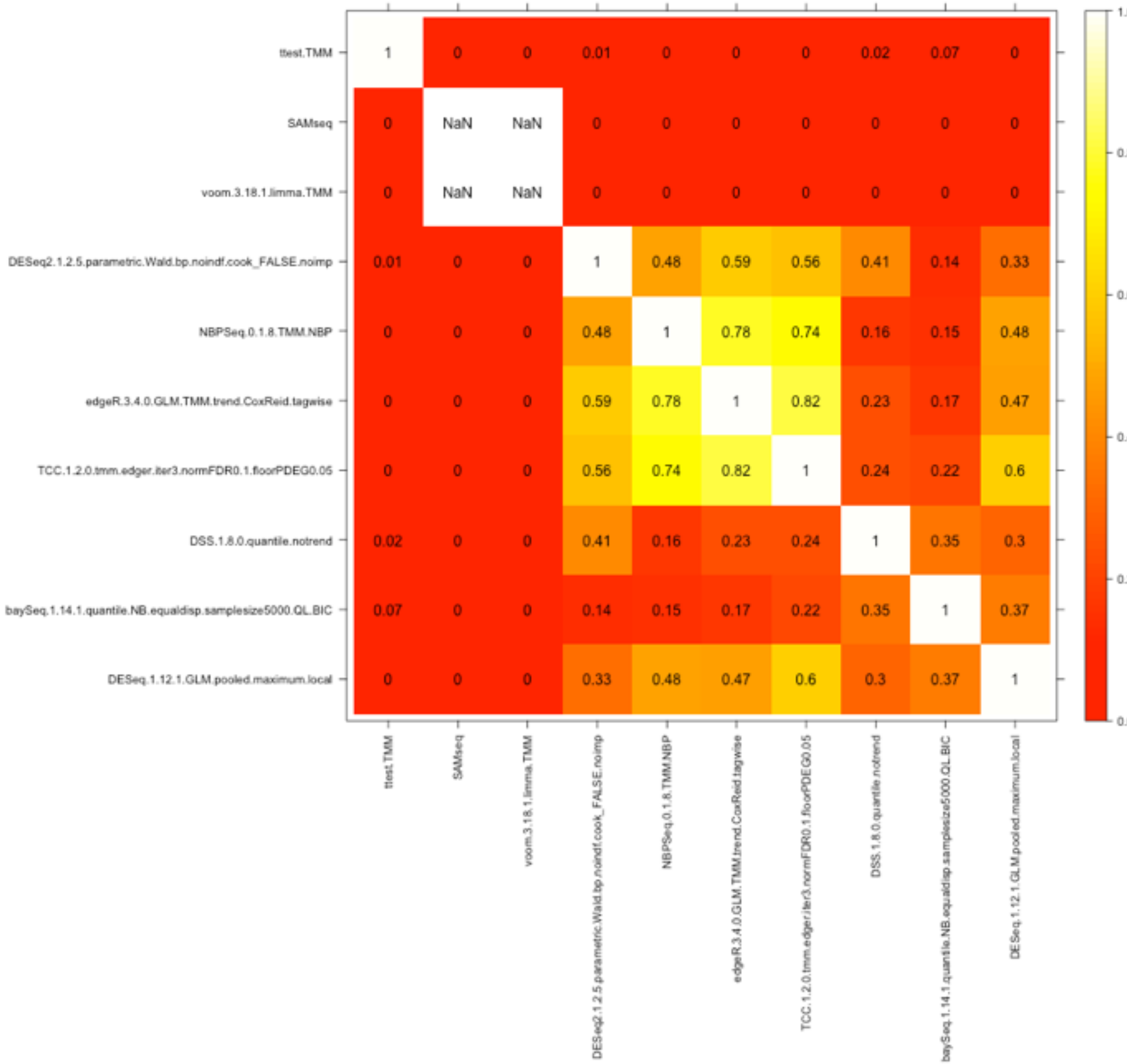
FDR, 5 replicates/condition



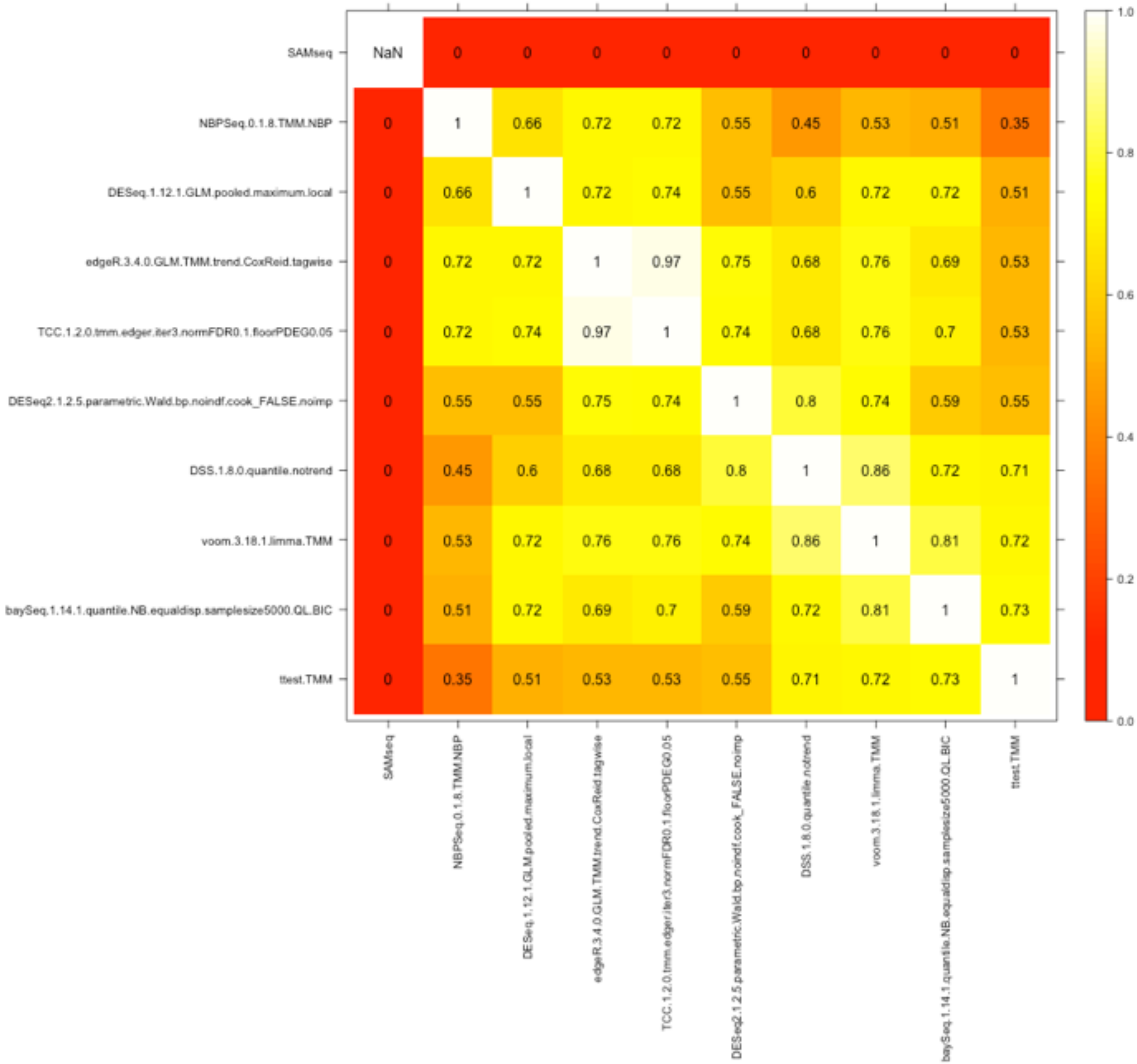
FDR, 10 replicates/condition



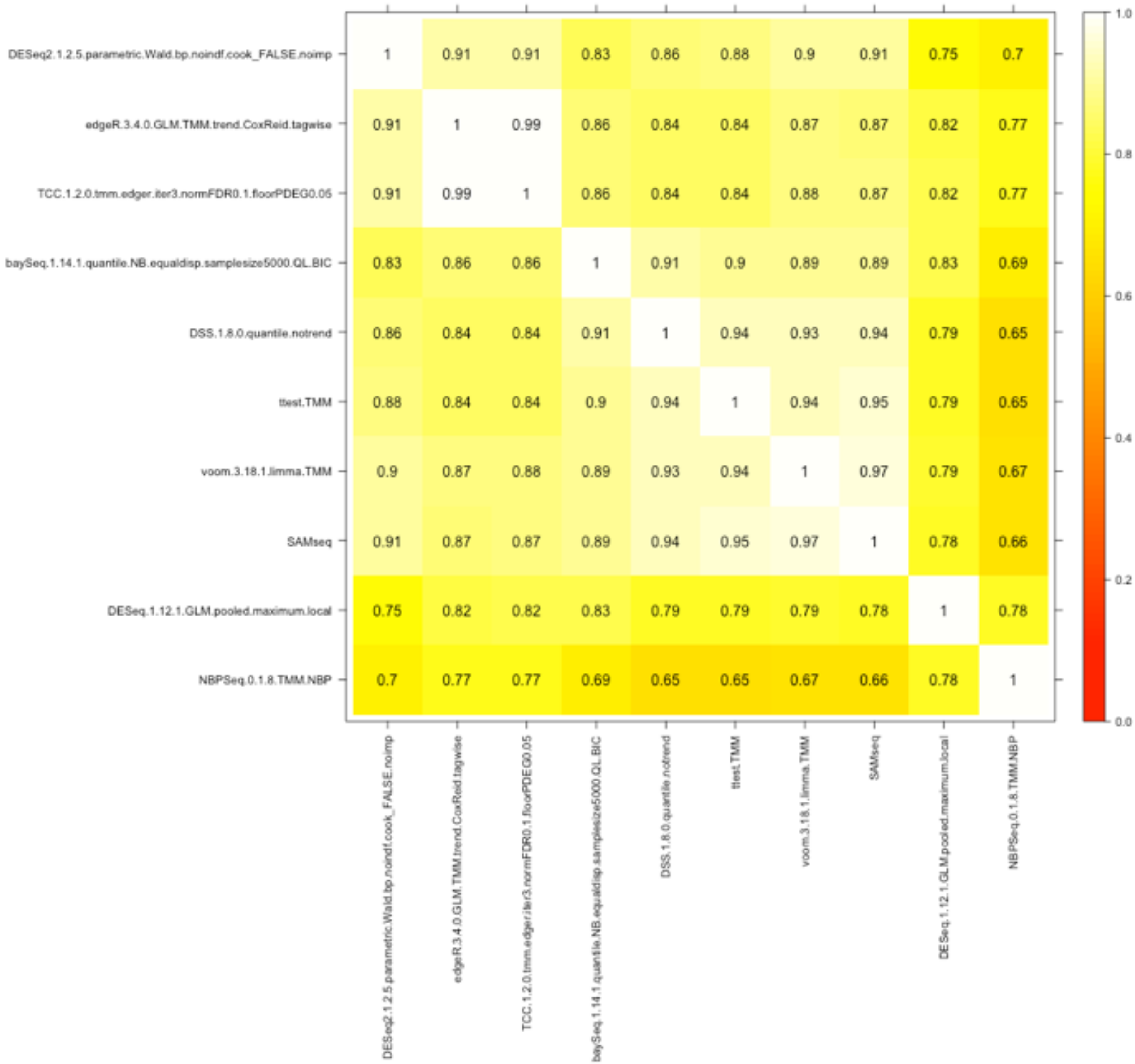
Similarity between sets of DEGs, 2 replicates/condition



Similarity between sets of DEGs, 5 replicates/condition



Similarity between sets of DEGs, 10 replicates/condition



References

- Akay et al.: On the design and analysis of gene expression studies in human populations. Nature Genetics 39(7): 807-808 (2007)
- Nygaard et al.: Methods that remove batch effects while retaining group differences may lead to exaggerated confidence in downstream analyses. Biostatistics 17(1):29-39 (2016)
- Danielsson et al.: Assessing the consistency of public human tissue RNA-seq data sets. Briefings in Bioinformatics 16(6):941-949 (2015)
- Larsen et al.: Microarray-based RNA profiling of breast cancer: batch effect removal improves cross-platform consistency. BioMed Research International vol. 2014, article ID 651751 (2014)
- Leek et al.: Tackling the widespread and critical impact of batch effects in high-throughput data. Nature Reviews Genetics 11(10:733-739 (2010)
- Leek & Storey: Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genetics 3(9):e161 (2007)
- Leek: svaseq: removing batch effects and other unwanted noise from sequencing data. Nucleic Acids Research 42(42):e161 (2014)
- Risso et al.: Normalization of RNA-seq data using factor analysis of control genes or samples. Nature Biotechnology 32(9):896-902 (2014)
- Gagnon-Bartsch & Speed: Using control genes to correct for unwanted variation in microarray data. Biostatistics 13(3):539-552 (2012)
- Schurch et al.: How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA (2016)
- Lazic: Experimental Design for Laboratory Biologists: Maximising Information and Improving Reproducibility. Cambridge University Press (2016).