

# Experimental design

Charlotte Soneson

Computational Biology Platform, Friedrich Miescher Institute for Biomedical Research

June 28, 2024



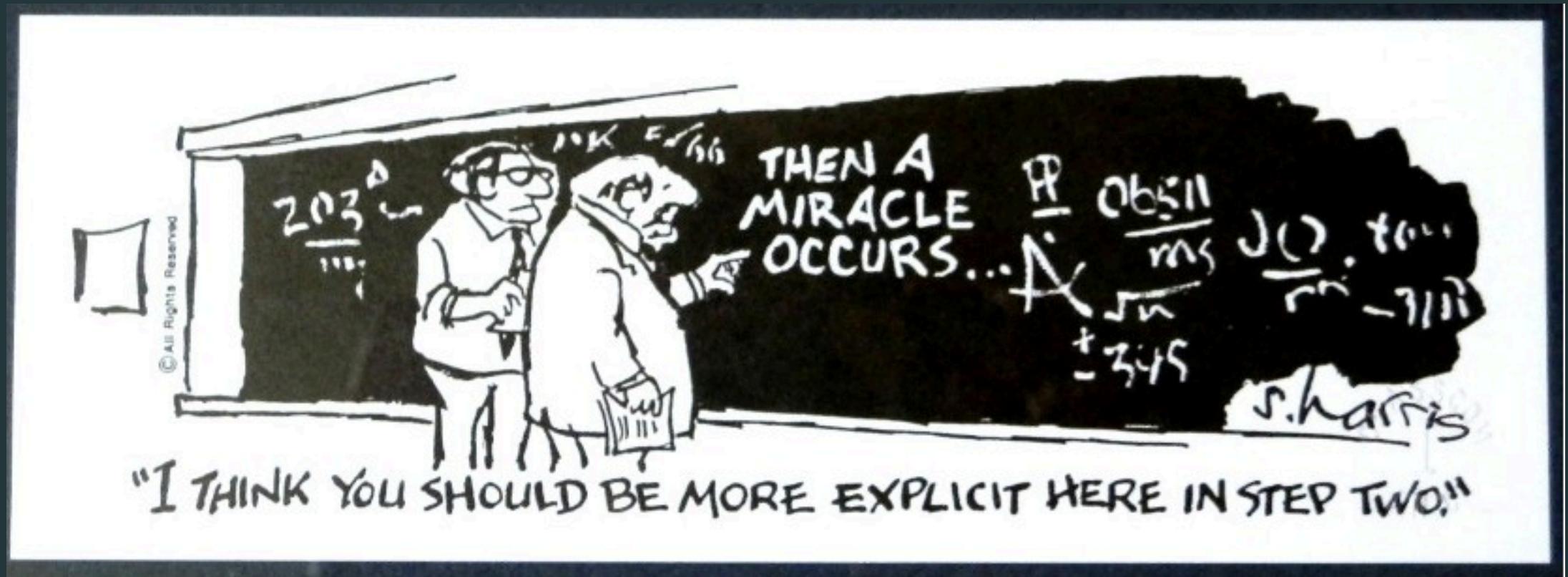
# What is experimental design?

"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."

"Good experimental design begins with the end in mind."

- What is my **hypothesis**?
- What am I going to **measure** (and how is this relevant to the hypothesis)?
- What is a suitable **control**?
- What **other factors** affect this outcome measure (and am I controlling these appropriately)?

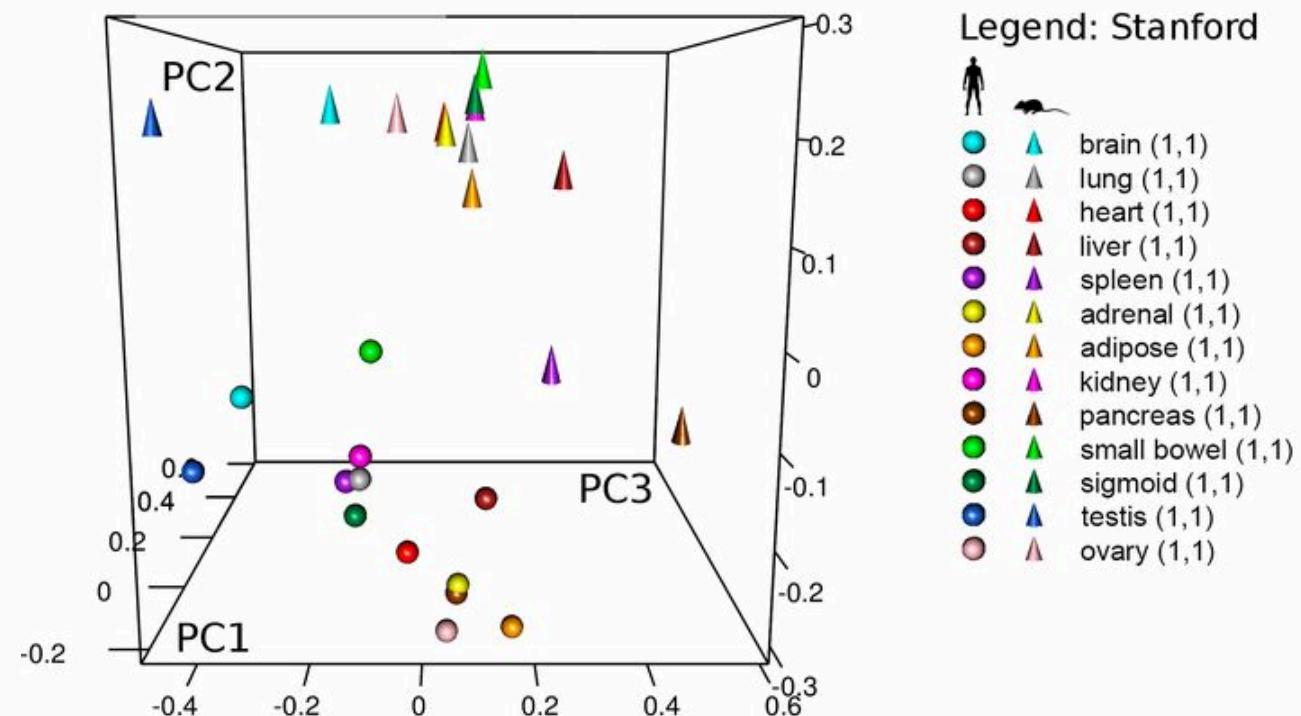
# A couple of motivating examples



# Case study: Lin et al (PNAS, 2014)

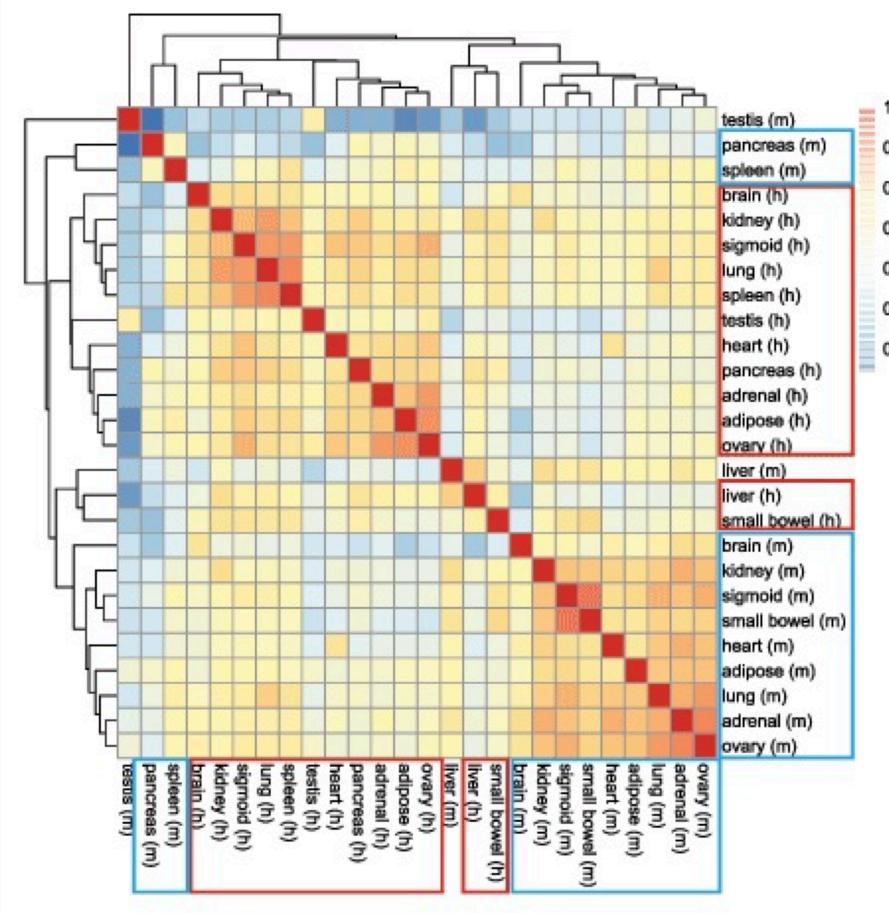
"In this study of a broad number of tissues between humans and mice, high-throughput sequencing assays on the transcriptome and epigenome reveal that, in general, differences dominate similarities between the two species."

[...] well beyond what was described previously, likely reflecting the fundamental physiological differences between these two organisms."



Subset of 13 mouse tissues and 13 human tissues with data generated at the same site (Stanford)

# Gilad & Mizrahi-Man (F1000Research 2015)



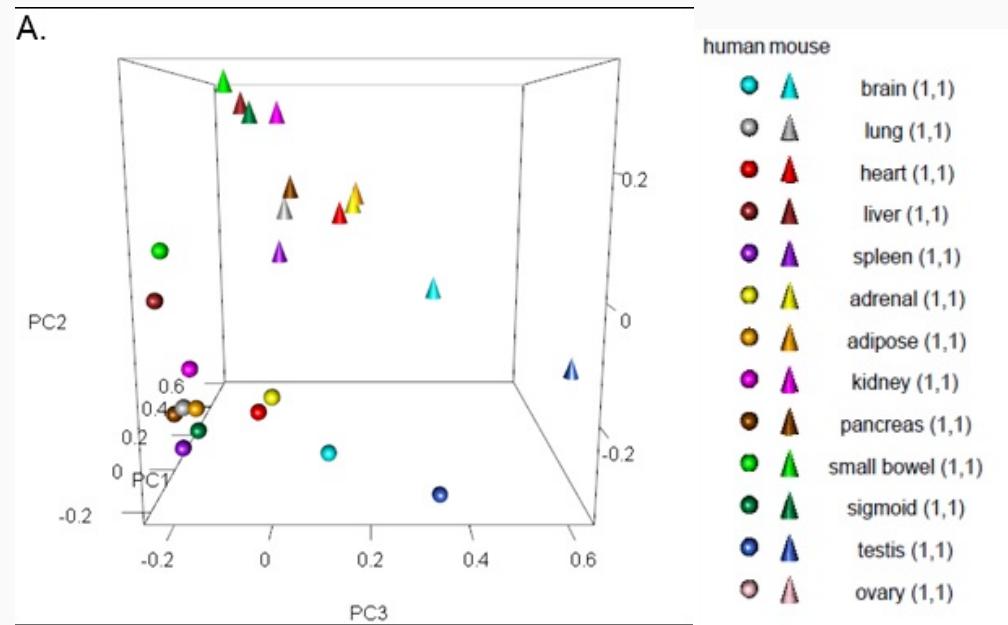
D87PMJN1 (run 253, flow cell D2GUAACXX, lane 7)	D87PMJN1 (run 253, flow cell D2GUAACXX , lane 8)	D4LHBFN1 (run 276, flow cell C2HKJACXX , lane 4)	MONK (run 312, flow cell C2GR3ACXX , lane 6)	HWI-ST373 (run 375, flow cell C3172ACXX , lane 7)
heart	adipose	adipose	heart	brain
kidney	adrenal	adrenal	kidney	pancreas
liver	sigmoid colon	sigmoid colon	liver	brain
small bowel	lung	lung	small bowel	spleen
spleen	ovary	ovary	testis	
testis		pancreas		

Species is **confounded** with sequencing machine.

- Differences on the sample level
  - human - heterogeneous set of deceased individuals (18-66 years), tissue flash frozen, RNA for three tissues acquired commercially.
  - mouse - single strain, similar age (10 weeks), unclear whether tissue was flash frozen.
- Tissues not sex-matched

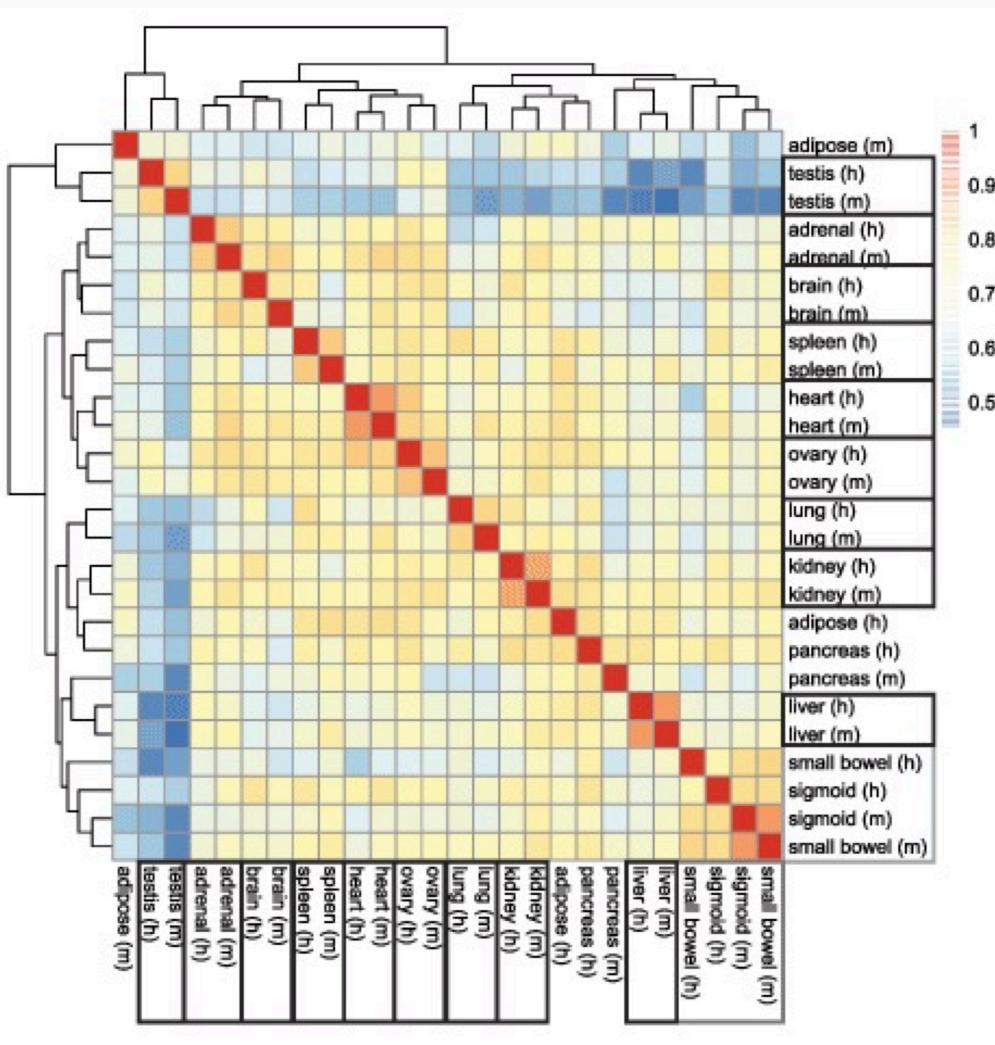
# Follow-up experiments triggered

- Resequencing of a subset of the tissue samples with a non-confounded sequencing design - similar conclusions.
  - Doesn't address the underlying sample differences



- Analysis of Fantom5 CAGE data (matched human and mouse) - same conclusions.
- Analysis of GTEx tissue samples
  - "Neither sex nor age contributed to the clustering in the first two principal components. Thus, these effects are likely to be small and have not been included in the studies of others as well."

# Can we fix it in the analysis?



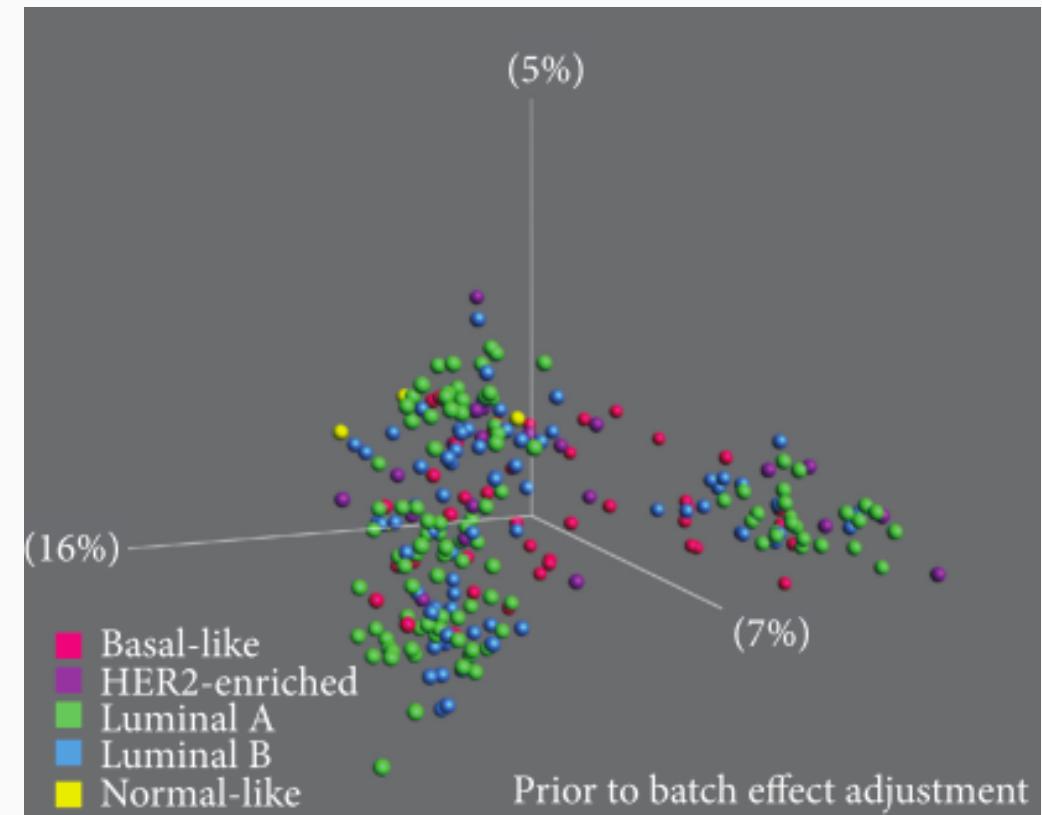
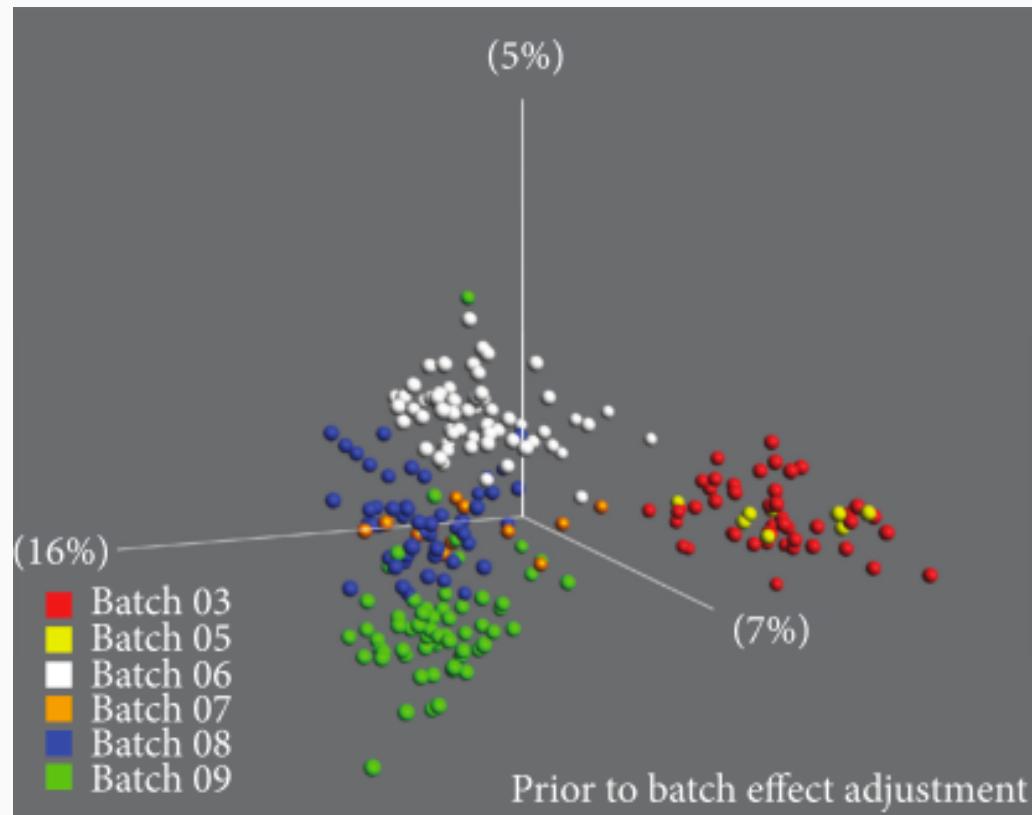
- Batch effect adjustment effectively "equalizes" or "aligns" the different batches.
- By construction, this removes systematic shift between different sequencing runs.
- **Unavoidably**, this removes also of the species effect (the species are sequenced in different batches).
- There's *no way* to determine the source of the signal based on this data alone.

# So - what can we conclude?

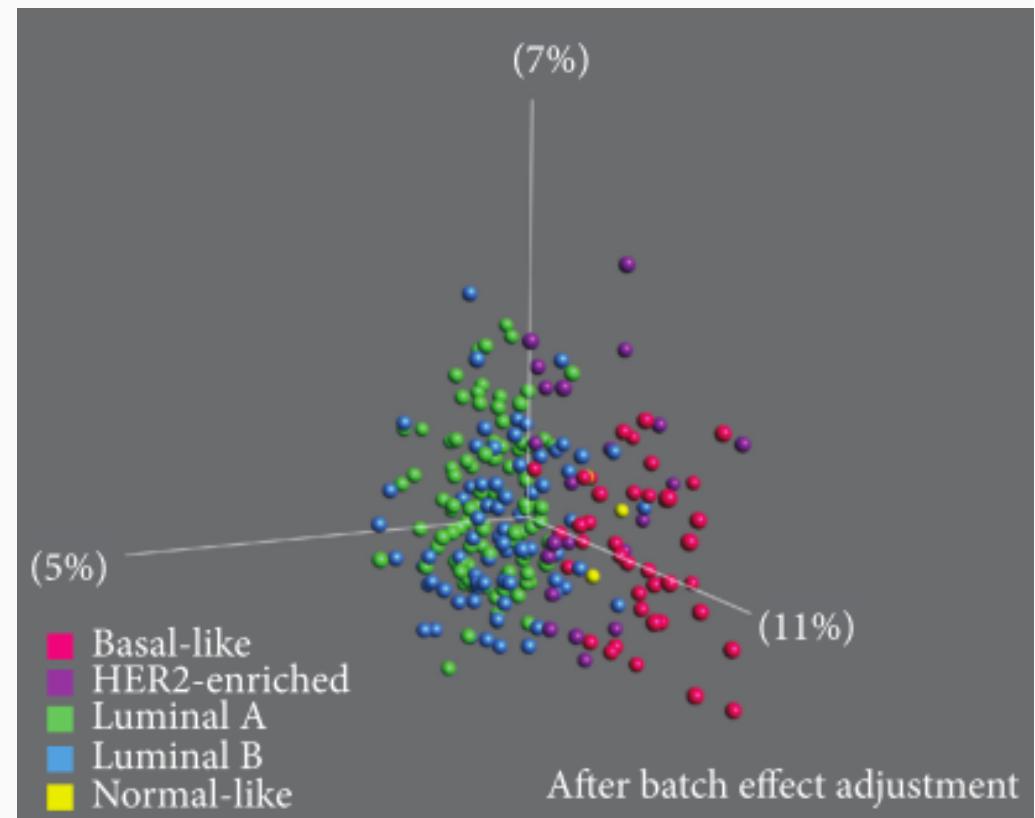
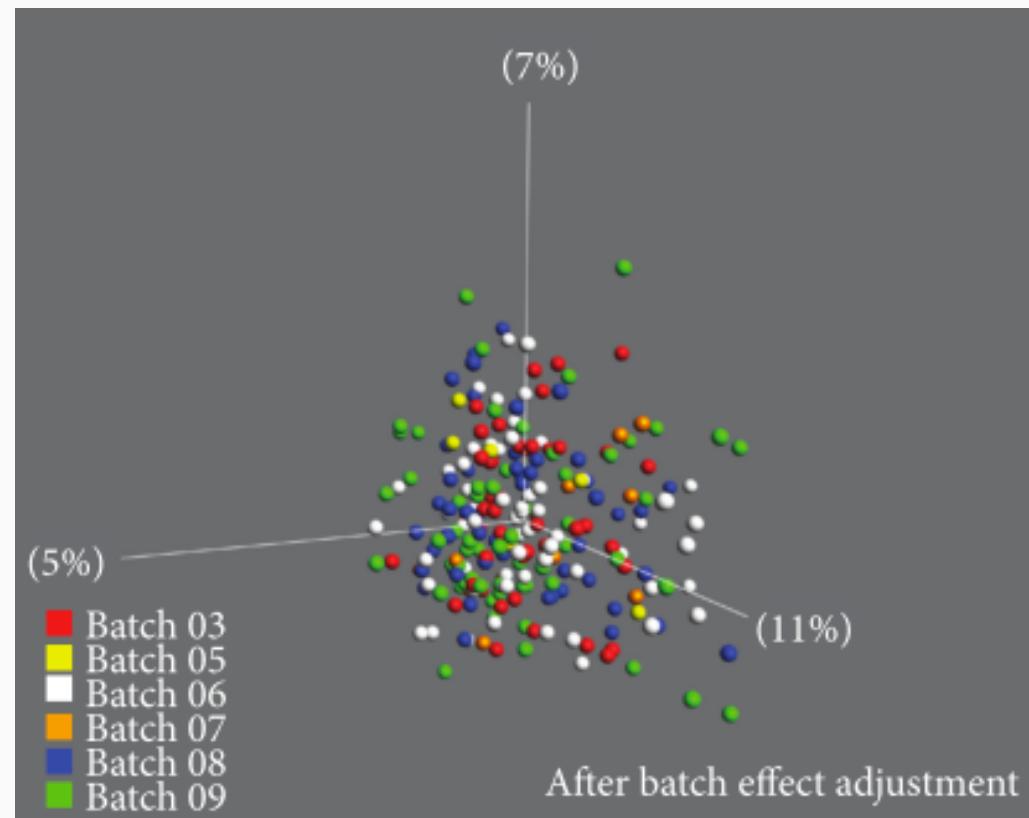


# Case study II (Larsen et al, BioMed Res Int 2014)

- Gene expression data from samples from five breast cancer subtypes, processed across six batches.
- Batch signal dominates in uncorrected data.

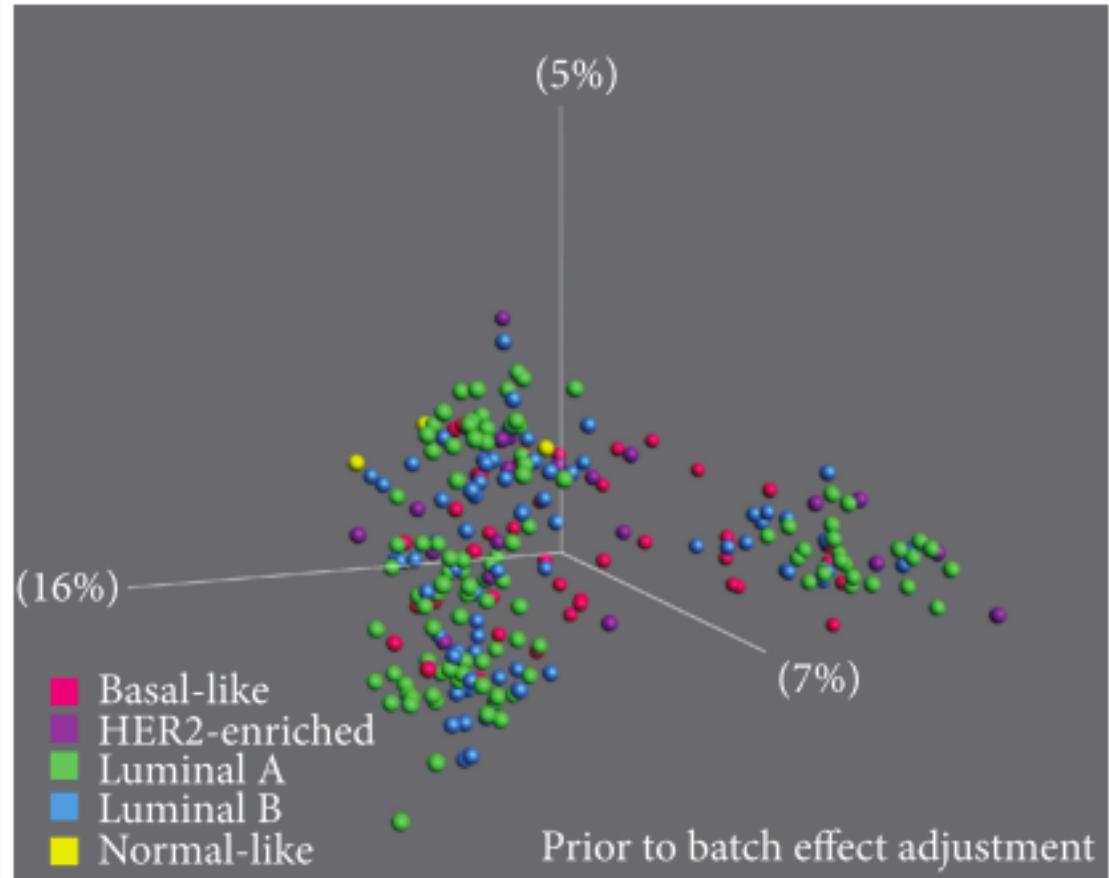


# Case study II (Larsen et al, BioMed Res Int 2014)

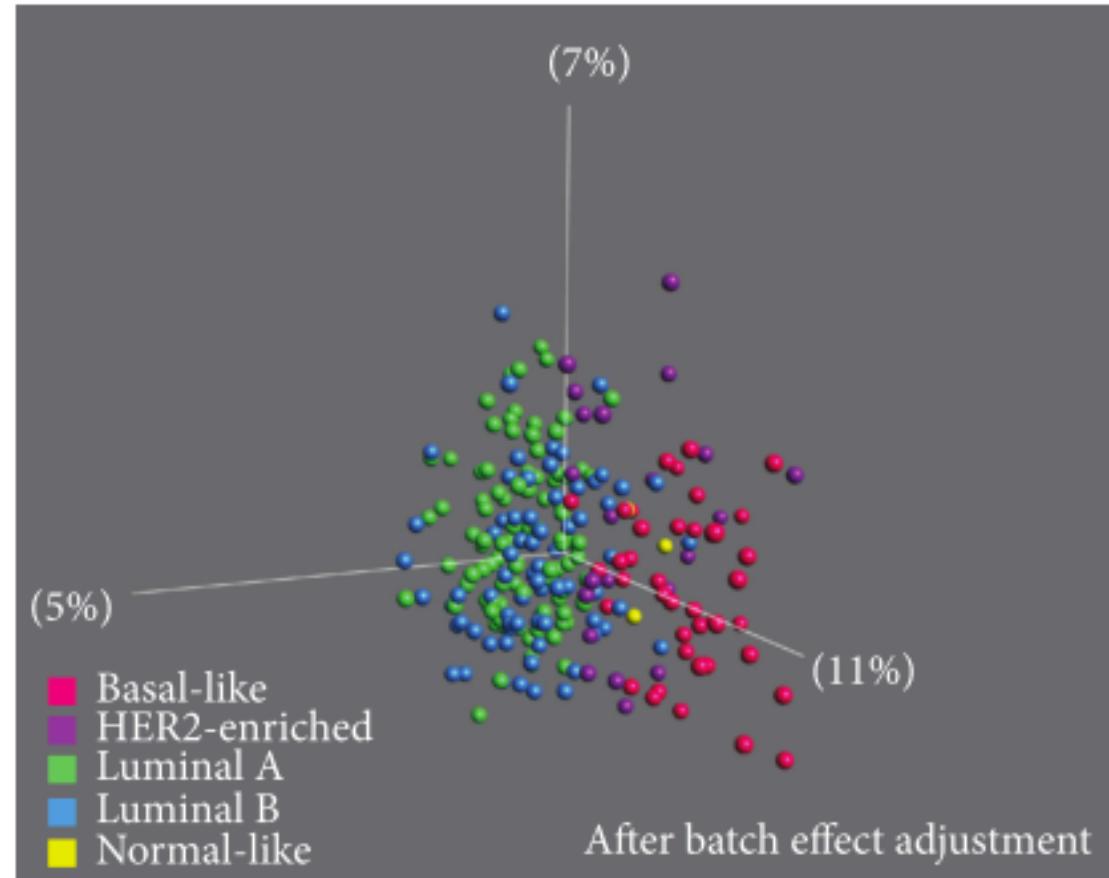


- After batch correction, the subtypes represent the strongest signal in the data.
- This was **only** possible since the subtypes were spread randomly across the analysis batches!
- This illustrates that it's not (mainly) the presence of batch effects *per se*, but the confounding with the signal of interest, that causes problems for the analysis.

# How do we 'adjust for' batch effects?



(c)



(d)

# How do we 'adjust for' batch effects in statistical modeling?

- In statistical modeling (e.g., differential expression analysis), batch effects can be included as covariates (additional predictors) in the model.

$$y = \beta_0 + \beta_1 \cdot \text{batch} + \beta_2 \cdot \text{condition} + \varepsilon$$

- Note that if batch and condition coincide, we can at best hope to estimate  $\beta_1 + \beta_2$ .
- For exploratory analysis, we often attempt to "eliminate" or "adjust for" such unwanted variation in advance, by subtracting the estimated effect from each variable (e.g. the expression of a gene).
  - Not recommended before statistical modelling, since the variance will be underestimated.

# What can we learn from this?

- Acknowledge that lots of factors can affect measurements in (biological) experiments

Outcome = <u>Treatment effects</u> + <u>Biological effects</u> + <u>Technical effects</u> + <u>Error</u>			
Environment Compound Inhibitor siRNA Dose Time	Sex Age Weight Litter Genotype Species Cell line	Technician Batch Plate Cage Array Day Order Source	Experimental Treatment Sampling Measurement

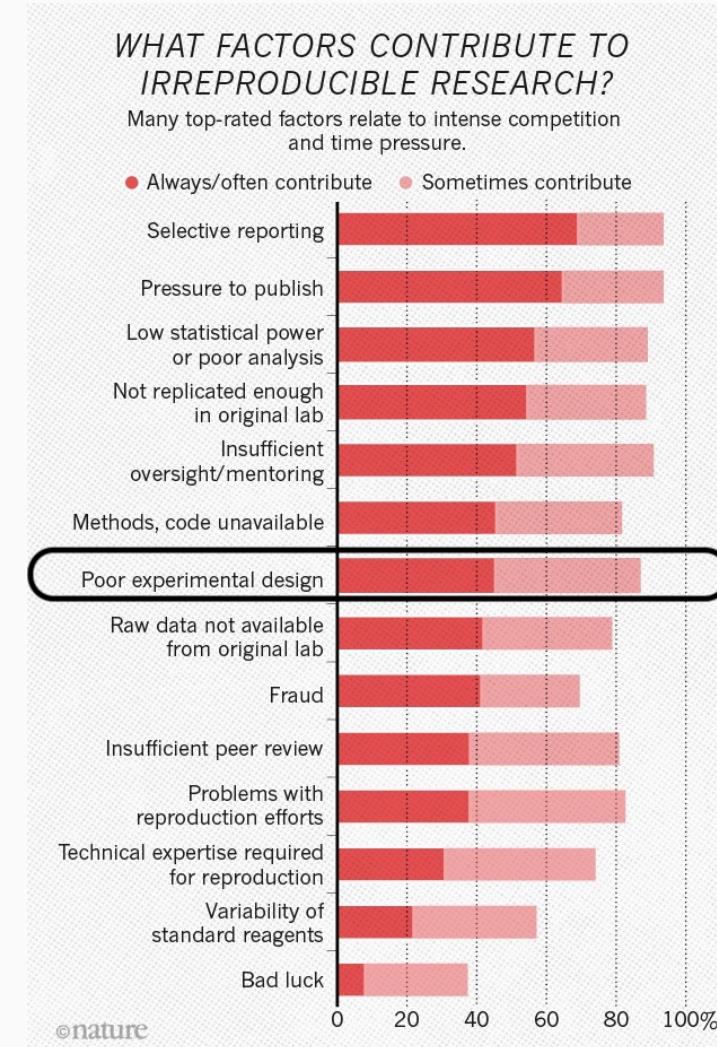
- We (typically) can't change the fact that these factors *do* affect the measurements, so we have to adapt accordingly.
- Spending some extra time thinking through the experiment *before* starting to collect the data can save a lot of time and headache later on.
- Always record as much information as possible about the samples, and share it with whoever analyzes the data. If you are the data analyst - ask questions!
- Avoid confounding between the factor of interest and any source of unwanted variation.

# Always beware of confounding

- Careful design is always important - beware of confounding e.g. in situations where we
  - have to split a large data set into multiple analysis batches
  - use a control group from another (public) data set
  - use different reagent batches
  - collect an expanding data set (adding one condition at a time)
  - work in collaborative projects, where each center generates a subset of the data
  - change some aspect of the sample/library preparation protocol
  - work on an instrument with a drift/degradation over time
  - work with samples that degrade over time
  - ...

# So why should we care about experimental design?

- Improved reproducibility of results
- Improved validity and generalizability of conclusions
- Efficient and ethical use of resources
  - Make sure that the design is adequate for answering the question of interest
  - Avoid having to repeat poorly designed experiments
  - Avoid spending resources on non-interpretable experiments



# Experimental design principles



# Different types of experiments

- Learning experiments

- Discover as much as possible about a condition/situation/phenomenon, generate new hypotheses.
- Example: "Does stress affect rodent behaviour?"
- Heterogenous subjects and environment.
- Several treatments/time points.

- Confirming experiments

- Verify/validate an observation (typically from a learning experiment or literature).
- Example: "Does fox urine odour affect the amount of food Wistar rats consume during the first 24 hours after exposure?"
- Standardized subjects and environment.
- Small number of treatments/time points.

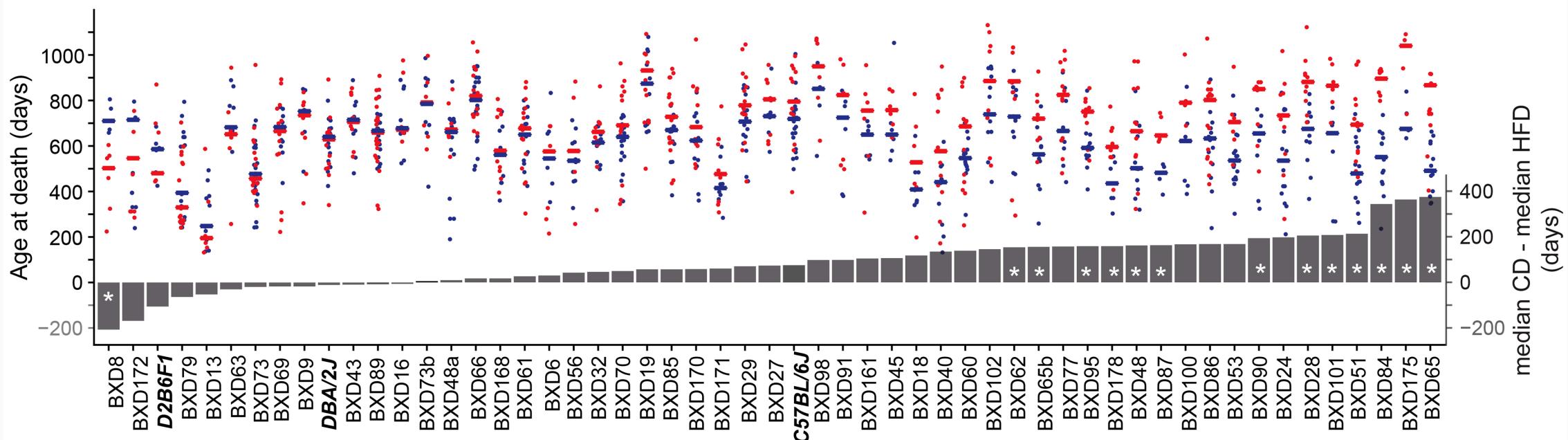
- Both types of experiments are valuable.
- Most high-throughput experiments today are learning experiments.
- Beware of learning experiments that are *presented* as confirming experiments.
  - Example: many protocols are tested, only the results from the "best" one are shown.
- Data from a learning experiment should not be reused in a follow-up confirming experiment.

# How to deal with the impact of unwanted variation?

Outcome = <u>Treatment effects</u> + <u>Biological effects</u> + <u>Technical effects</u> + <u>Error</u>			
Environment Compound Inhibitor siRNA Dose Time	Sex Age Weight Litter Genotype Species Cell line	Technician Batch Plate Cage Array Day Order Source	Experimental Treatment Sampling Measurement

# How to deal with the impact of unwanted variation?

- Minimize it by holding factors constant (e.g., only considering female mice). This reduces unwanted variability, but we can't generalize our conclusions to groups that are not included!

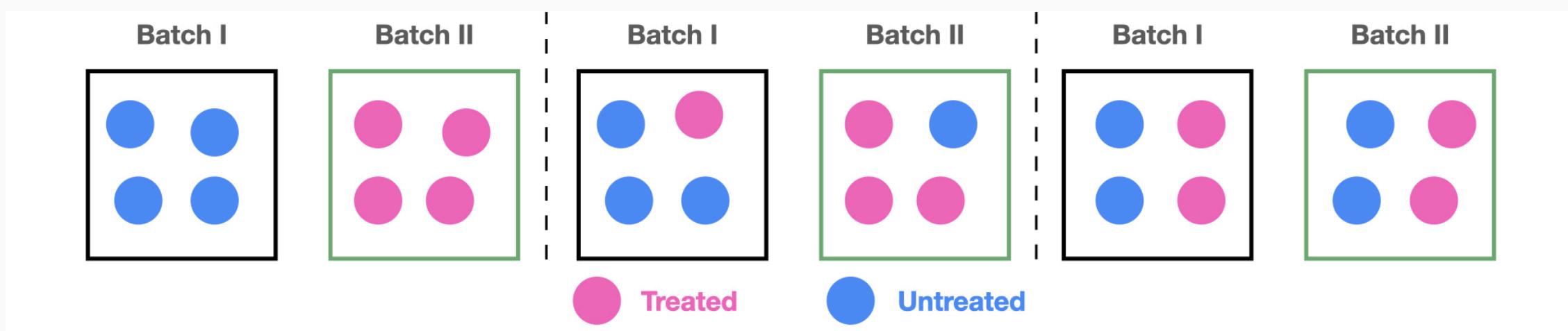


# How to deal with the impact of unwanted variation?

- **Minimize** it by holding factors constant (e.g., only considering female mice). This reduces unwanted variability, **but we can't generalize our conclusions to groups that are not included!**
- Use **blocking** to balance the factor of interest across confounding factors. E.g., assign treatment group randomly within each sex group.
  - Extreme example: blocking *within each biological sample* (e.g., take measurement from each mouse before and after a treatment). This leads to a **paired design**.
- Use **randomization** to hopefully balance out the impact of any unknown confounder. Useful for unknown effects as well as effects that can only be measured at the end of the experiment.

# Example

- Let's say that you are collecting data from four untreated and four treated mice.
- You can only collect and prepare data from four mice per day/session.
- How do you split your mice? Why?
  - All replicates of one treatment in the same session to maximize the similarity among replicates
  - Completely random assignment
  - Replicates with each treatment evenly distributed between the two sessions

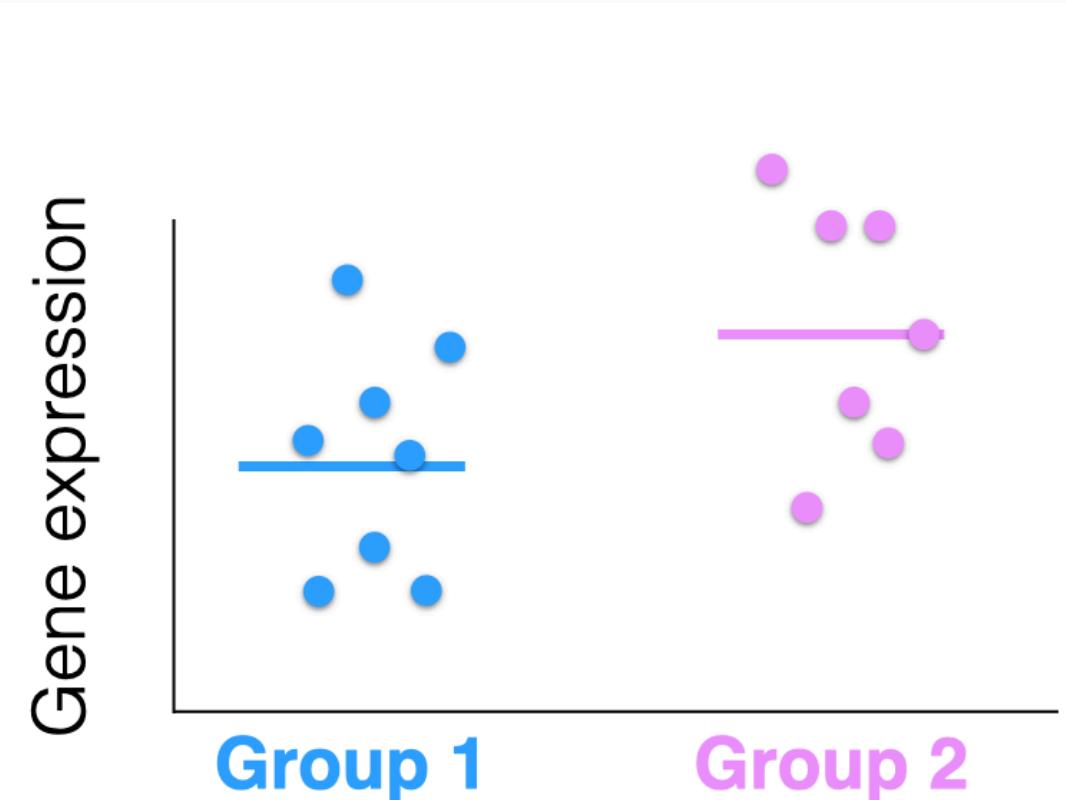
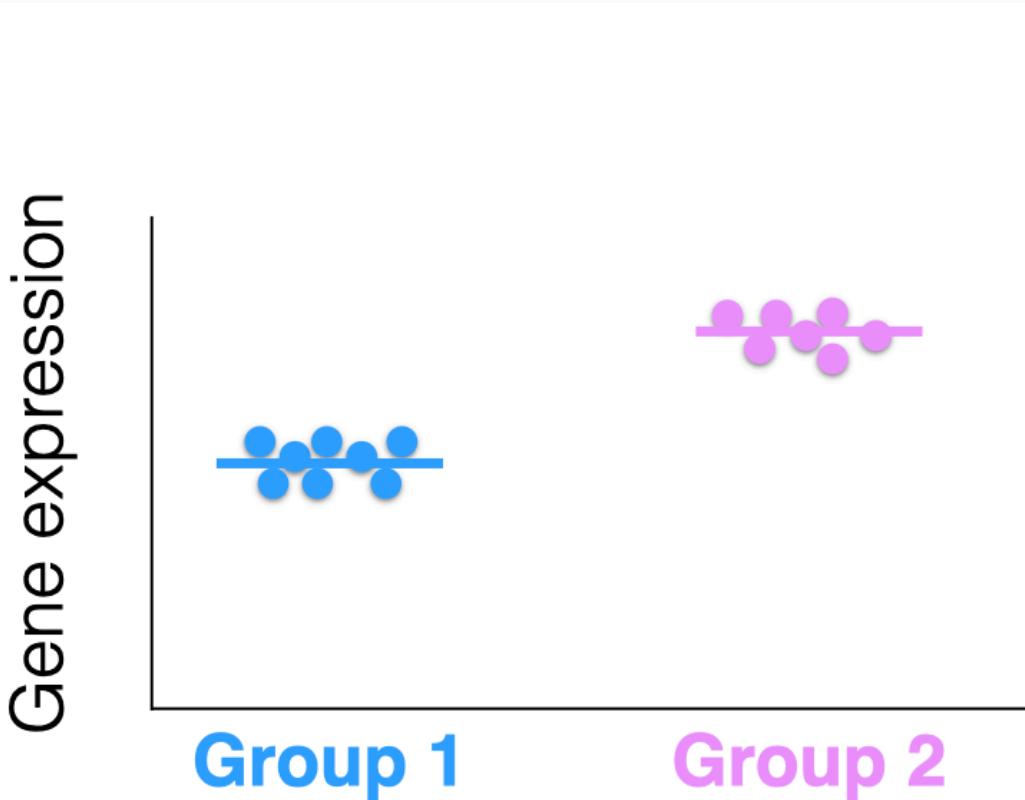


# Replication



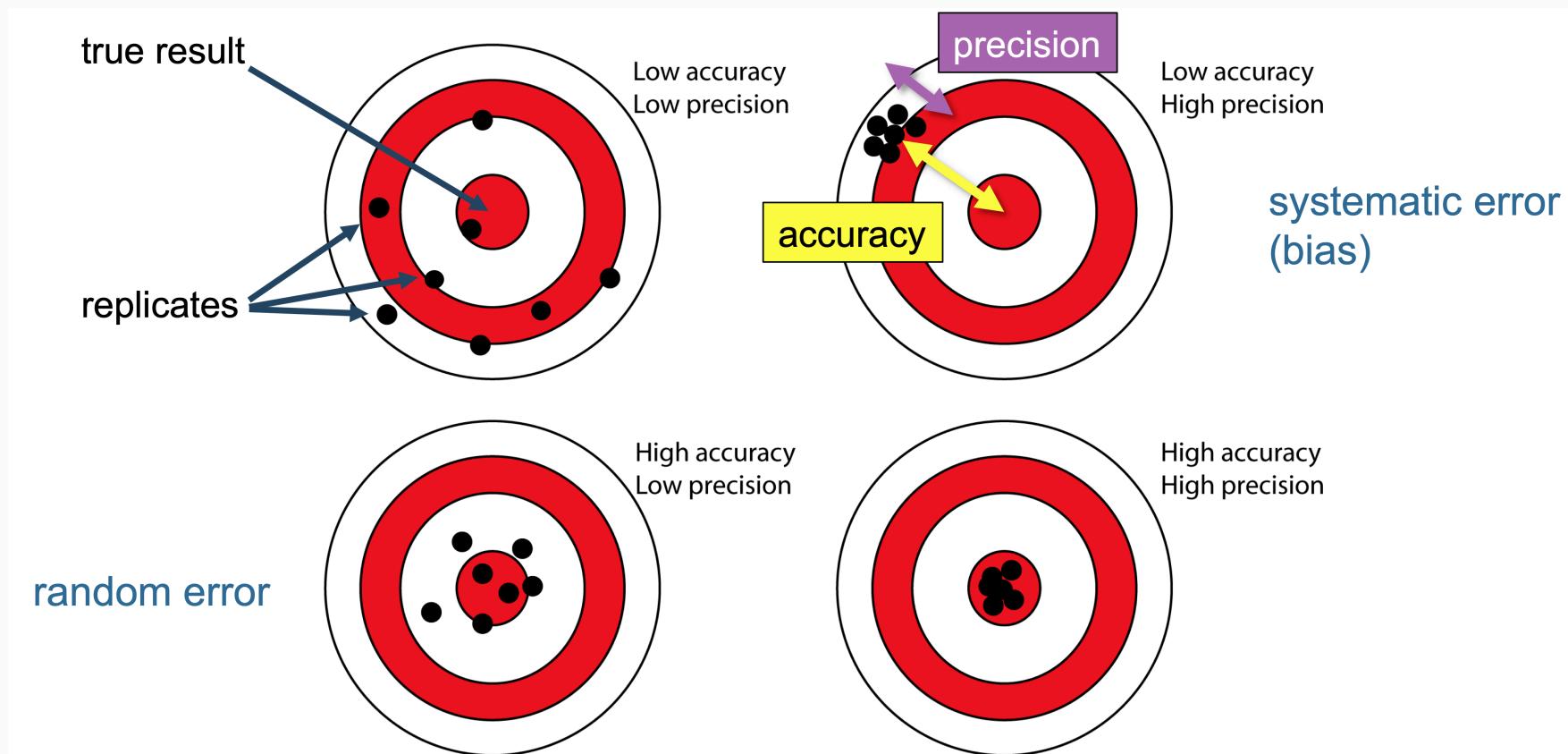
# Why do we need replicates?

- Required to estimate within-group variability, which is then compared to the observed between-group difference in statistical tests.



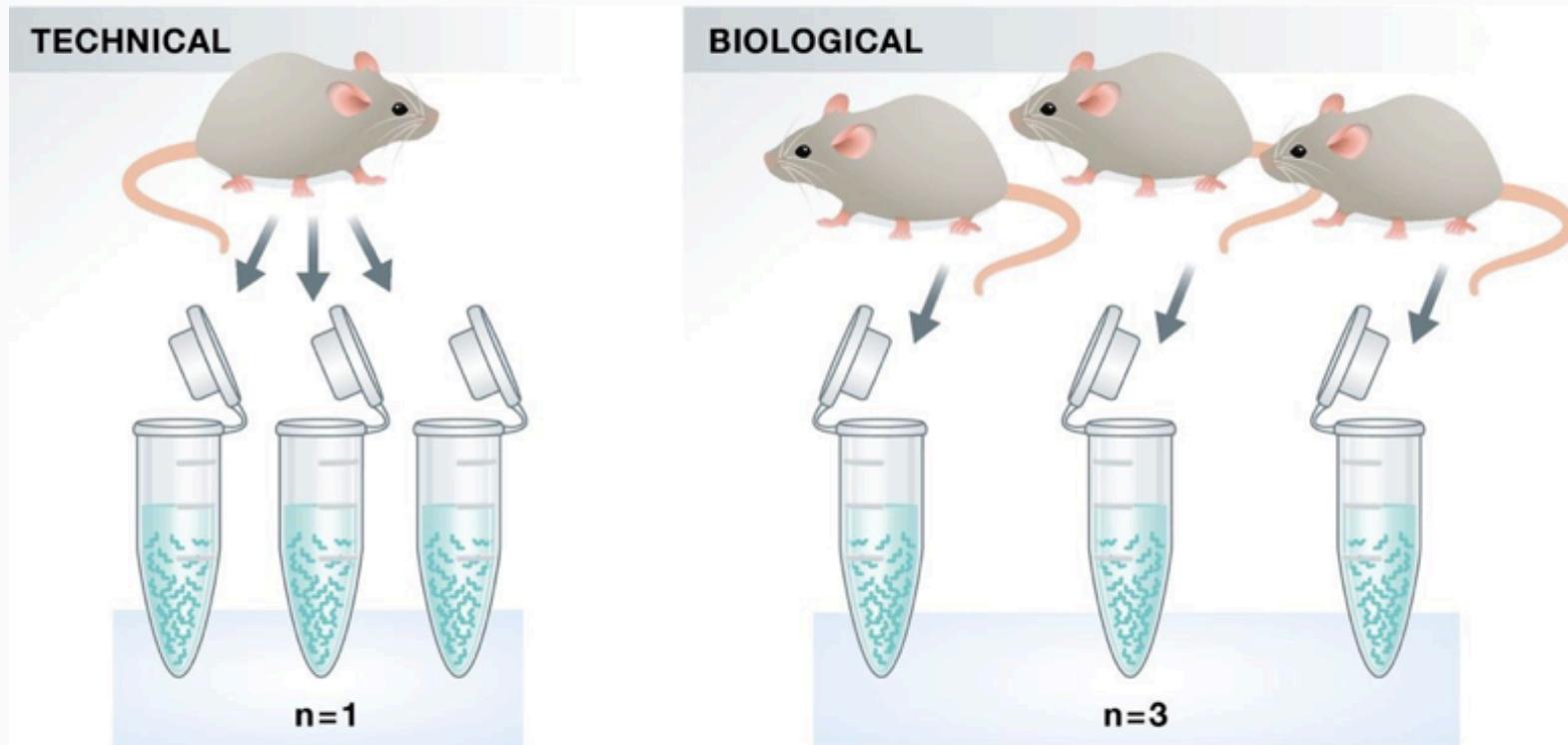
# What do we gain by increasing the number of replicates?

- More precision in our summary estimates (e.g., the mean).
- Replicates should be *representative* of the population of interest!
- Precision is not the same as accuracy!



# Not all replicates are equal!

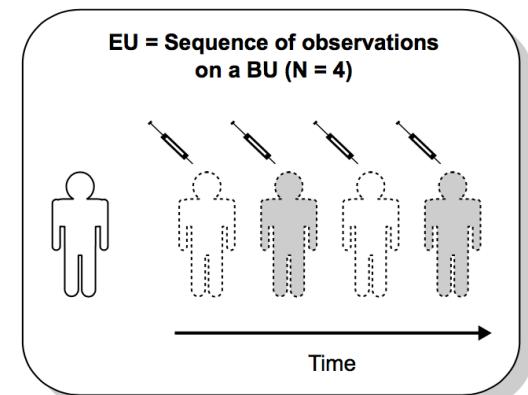
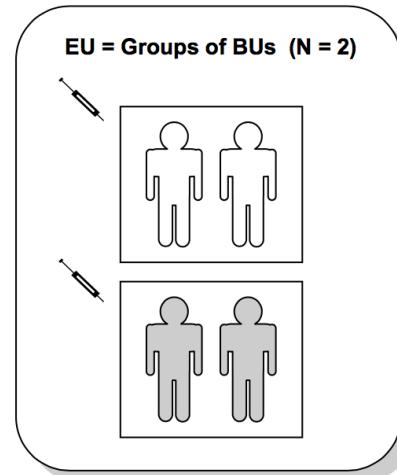
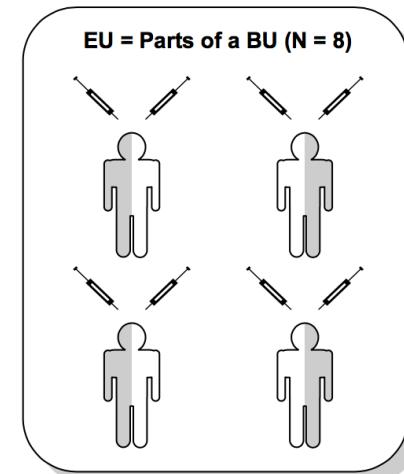
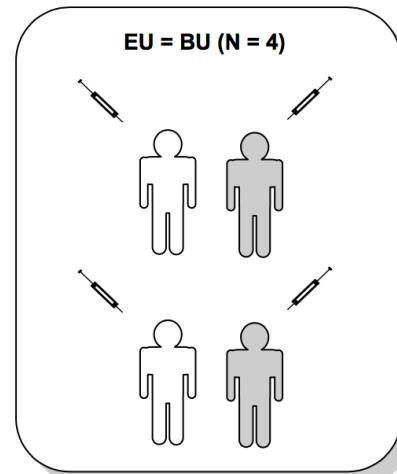
'Technical' vs 'biological' replicates



Common terminology, but sometimes confusing/not specific enough.

# Another classification approach

- **Biological units (BU)** - entities we want to make inferences about (e.g., animal, person).
- **Experimental units (EU)** - the 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.



# Another classification approach

- **Biological units (BU)** - entities we want to make inferences about (e.g., animal, person).
- **Experimental units (EU)** - the 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.

- Replication of **biological units** is required to make a general statement of the effect of a treatment - you can't draw a conclusion about the effect of a compound on people in general by observing the effect in one person.
- Only replication of **experimental units** is true replication (independent error terms).

# Experimental design checklist

- I have a well-defined scientific question or hypothesis to investigate (recall learning vs confirming experiments), and I have thought about how the type of data I will collect helps me to address this question.
- I have thought about how the data will be analyzed (e.g., which comparisons will be done, and what can be learnt from each of them). In particular, I have considered what a suitable 'control' group is for the effect I am interested in.
- The groups that I am planning to compare differ only in the aspect I am interested in investigating. Other confounders (technical/protocol differences, batch effects) are accounted for to the fullest extent possible.
- I have considered which biological or technical factors that might influence my measurements, and decided how to handle them (ignoring, eliminating, blocking, randomizing). Sources of (unwanted) variation are not confounded with the effect of interest.
- I have a suitable type and amount of replication, which allows me to generalize the results to the desired population.

Always record and report as much information as possible.