Design of high-throughput experiments and their analysis

Katharina Imkeller - Goethe University Franfkurt

CSAMA 2026

Good experimental design begins with the end in mind.

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.

Key Considerations:

- 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)?

Learning objectives

1. Confounding factors and batches

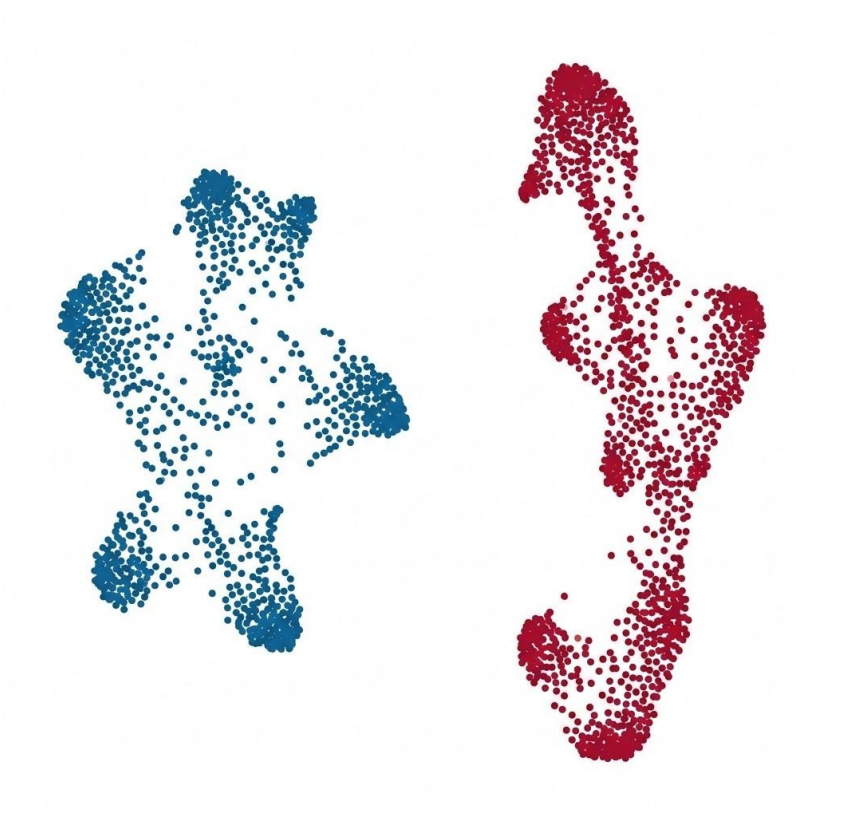
2. Replicate types and their application

3. Considerations for design of multi-omic experiments

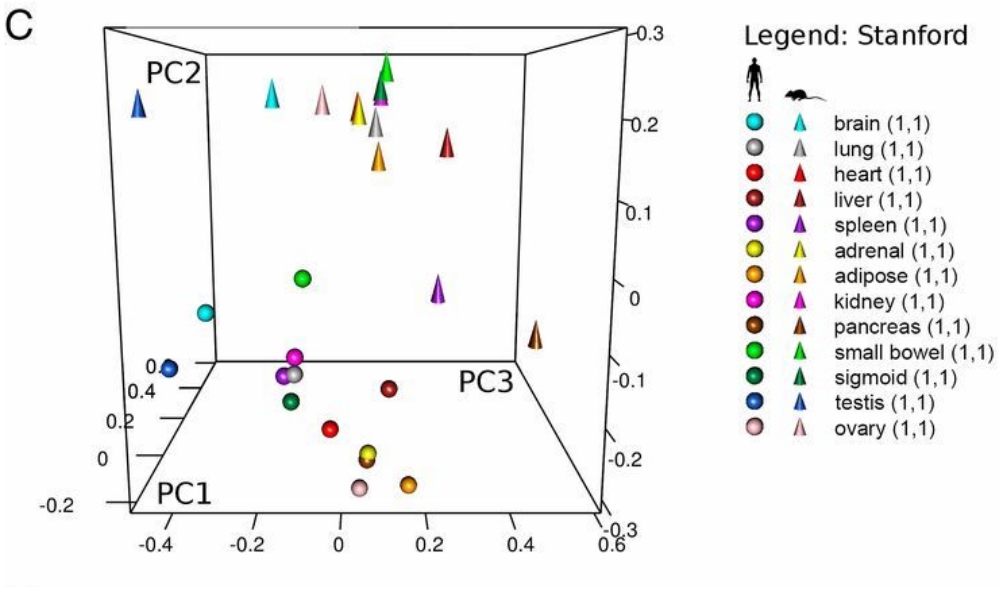
4. Data Management
Reproducible research practices

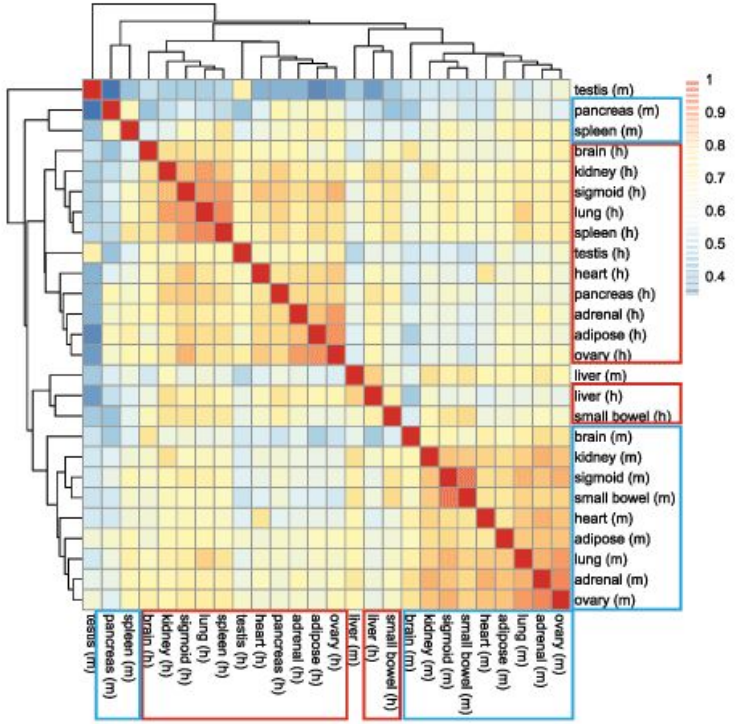
Learning objectives

1. Confounding factors and batches



Example of a flawed experimental design: Comparing mouse and human transcriptomes

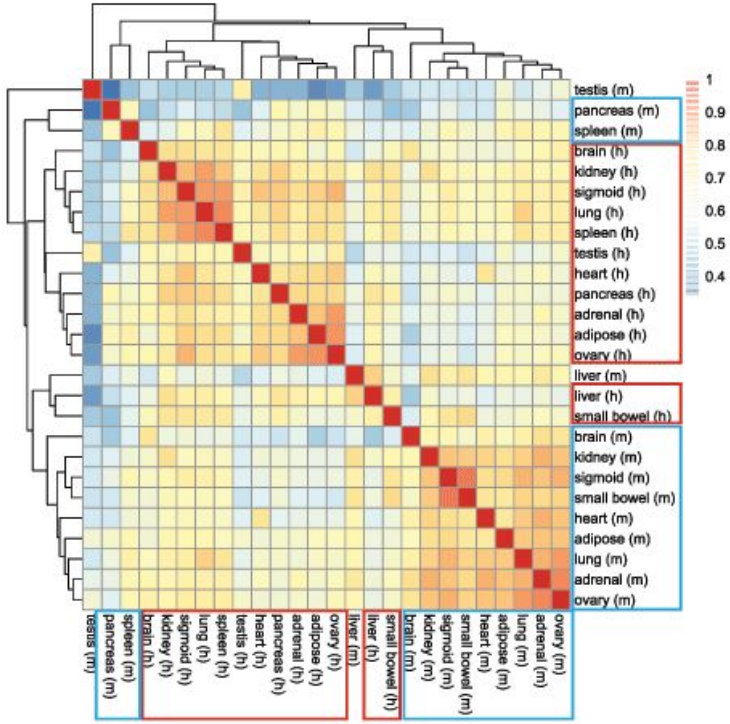




Lin et al (2014), doi:10.1073/pnas.1413624111
Gilad & Mizrahi-Man (2015), doi:10.12688/f1000research.6536.1

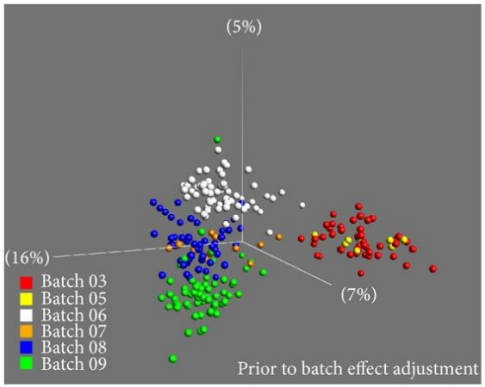
Example of a flawed experimental design: Species and sequencing experiment are confounded

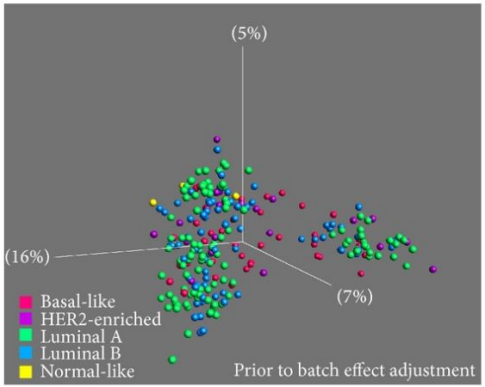
D87PMJN1 (run 253, flow cell D2GUAACXX, lane 7)	D87PMJN1 (run 253, flow cell D2GUAACXX , lane 8)	D4LHBFN1 (run 276, flow cell C2HKJACXX , lane 4)	MONK (run 312, flow cell C2GR3ACXX , lane 6)	HWI-ST373 (run 375, flow cell C3172ACXX , lane 7)
heart	adipose	adipose	heart	brain
kidney	adrenal	adrenal	kidney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



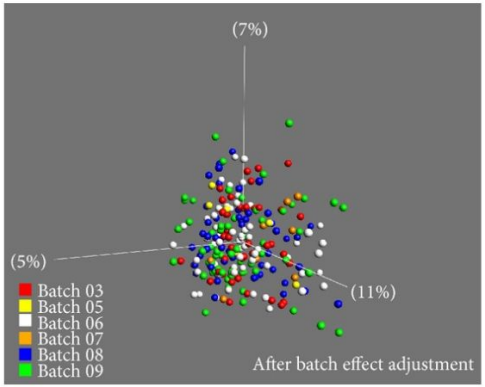
Lin et al (2014), doi:10.1073/pnas.1413624111
Gilad & Mizrahi-Man (2015), doi:10.12688/f1000research.6536.1

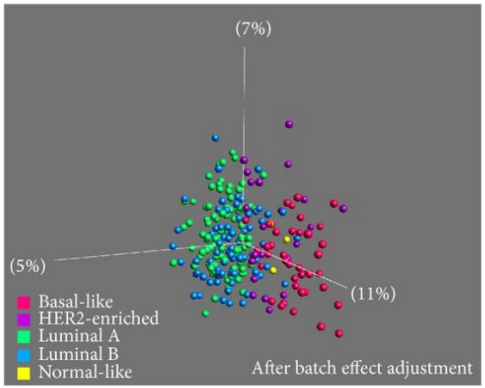
What do we mean by “batch correction”?





Batch correction





Larsen et al (2014), doi:10.1155/2014/651751

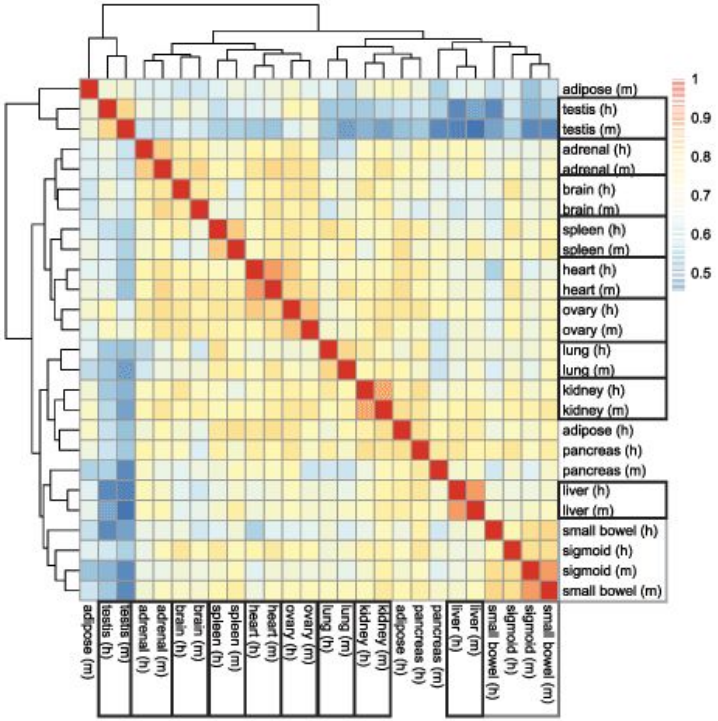
How do we adjust for batch effects in statistical modelling

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

$$y = \beta 0 + \beta 1 \cdot batch + \beta 2 \cdot condition + \varepsilon$$

- Note that if **batch** and **condition** coincide, we can at best hope to estimate β1 + β2.

After batch correction, the difference between species is lost



Batch Effect Adjustment Logic

Adjustment effectively *"equalizes"* or *"aligns"* the different batches.

By construction, this removes systematic shifts between different sequencing runs.

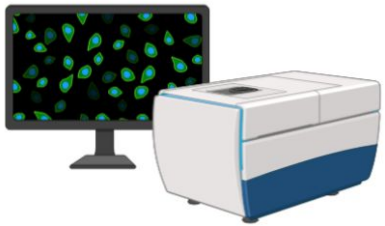
Critical Issue:

Unavoidably, this also removes the **species effect** because species were sequenced in different batches.

There is no way to determine the source of the signal based on this data alone.

Gilad & Mizrahi-Man (2015), doi:10.12688/f1000research.6536.1

Experimental design: treatment of cell lines



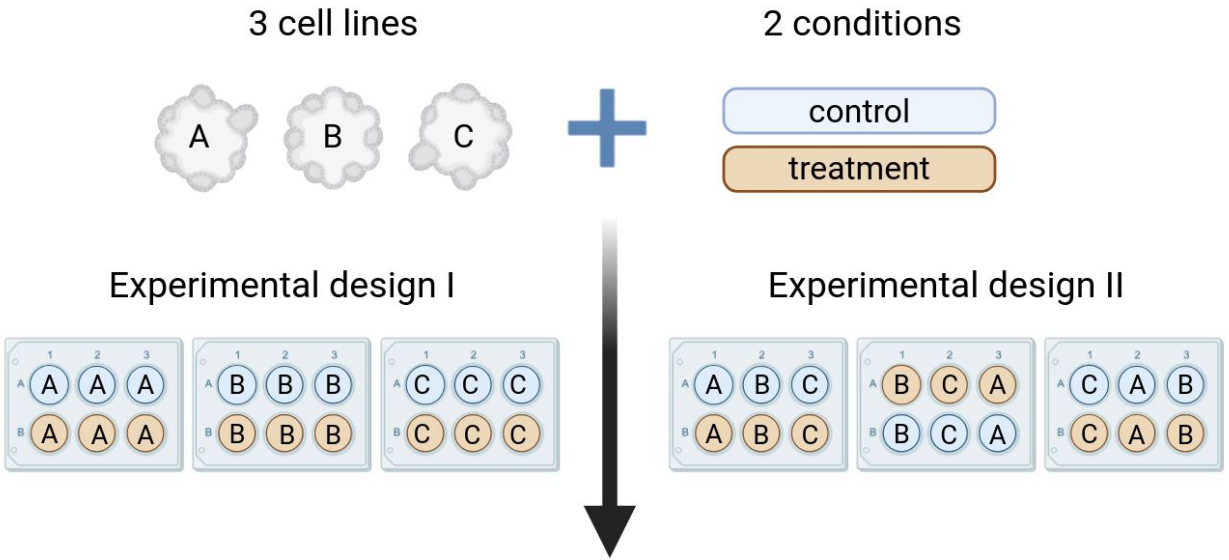
High content imaging

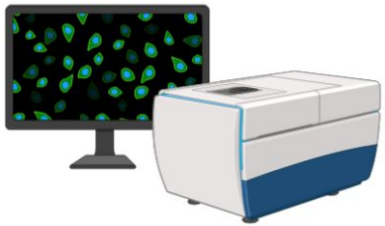
Readout: Features describing cell morphology

Hypothesis: The treatment alters the morphology of the cells over time.

Can only measure one plate at a time.

Experimental design: treatment of cell lines





High content imaging

Readout: Features describing cell morphology

Hypothesis: The treatment alters the morphology of the cells over time.

Can only measure one plate at a time.

Beware of confounding

Careful design is always important - beware of confounding e.g. in situations where we:

- 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 conditions)
- Work in collaborative projects with distributed generation
- Change sample/library preparation protocols
- Work on instruments with drift or degradation over time
- Work with samples that degrade over time (e.g., RIN)

Learning objectives

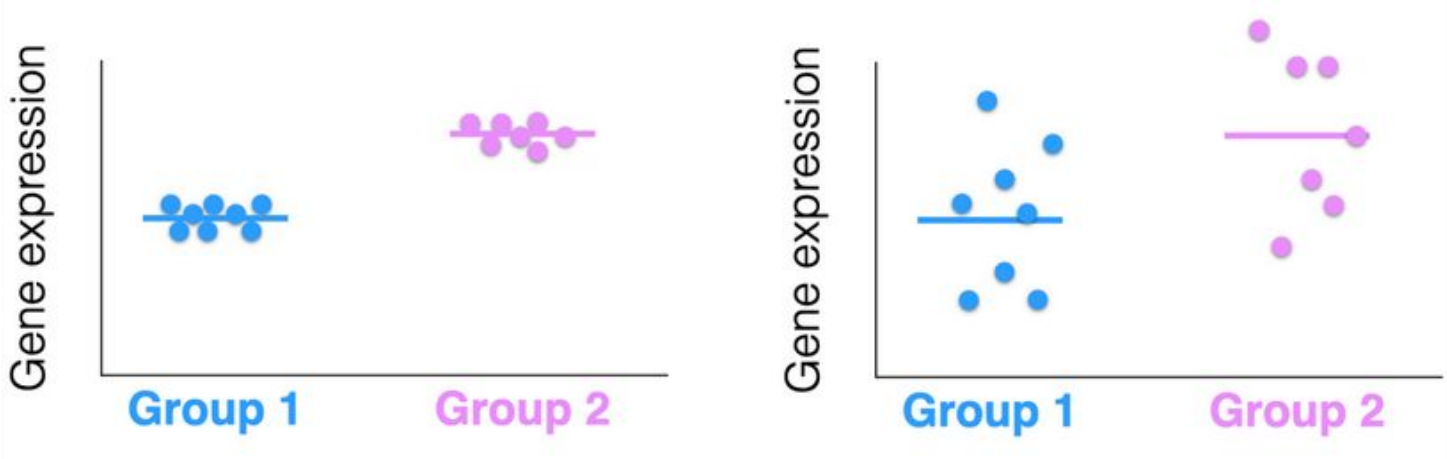
1. Confounding factors and batches

2. Replicate types and their application

Types of Replicates

Why do we need replicates?

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



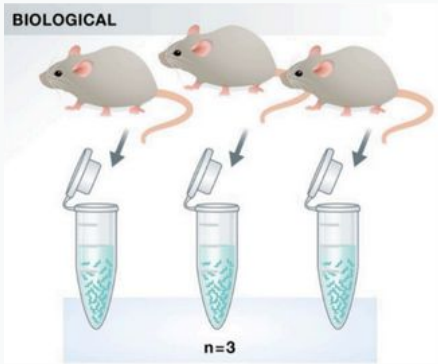
Technical vs. Biological Replicates

Technical Replicates

Definition: Multiple measurements of the same biological sample.

Purpose: Controls for **technical noise** (e.g., pipetting, equipment variability).

Biological Replicates



Definition: Measurements of independent biological samples.

Purpose: Captures **biological variation** within a population; required for inference.

Klaus 2015. doi:10.15252/embj.201592958

Another classification approach for replicates

Definition

Biological Unit (BU)

Entities we want to make inferences about (e.g., an individual or an animal).

Experimental Unit (EU)

The smallest entity that can be independently assigned to a treatment / condition (e.g., a single cage or a specific well).

Observational Unit (OU)

Entities at which measurements are made. (e.g., blood sample or single cell sequencing).

BU replicates required to make a generalizable statement

EU replication = true replication (independent error term)

Another classification approach for replicates

Definition

Biological Unit (BU)

Entities we want to make inferences about (e.g., an individual or an animal).

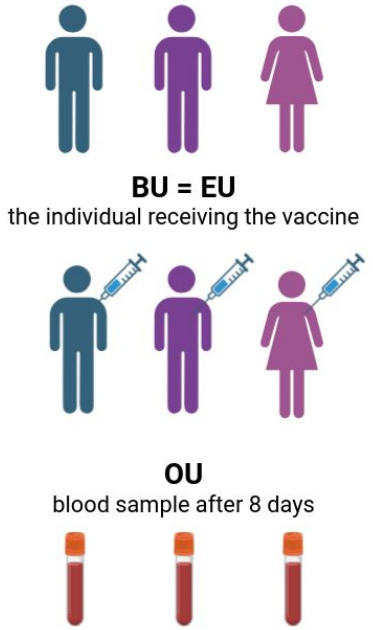
Experimental Unit (EU)

The smallest entity that can be independently assigned to a treatment / condition (e.g., a single cage or a specific well).

Observational Unit (OU)

Entities at which measurements are made. (e.g., a blood sample or a single cell reading).

Example I:
B cell reponse after vaccination



Another classification approach for replicates

Definition

Biological Unit (BU)

Entities we want to make inferences about (e.g., an individual or an animal).

Experimental Unit (EU)

The smallest entity that can be independently assigned to a treatment / condition (e.g., a single cage or a specific well).

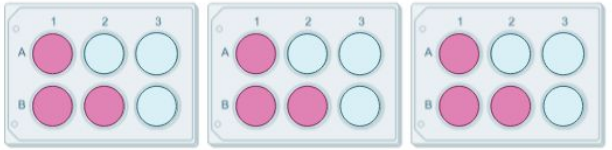
Observational Unit (OU)

Entities at which measurements are made. (e.g., a blood sample or a single cell reading).

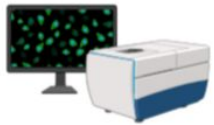
Example II:
Treatment effect on organoid model

BU
tumor (organoid culture)

EU = parts of BU
sample of organoid culture



OU
image of one well after 3 days



Another classification approach for replicates

Definition

Biological Unit (BU)

Entities we want to make inferences about (e.g., an individual or an animal).

Experimental Unit (EU)

The smallest entity that can be independently assigned to a treatment / condition (e.g., a single cage or a specific well).

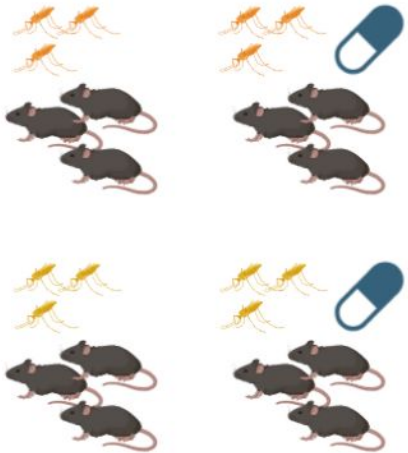
Observational Unit (OU)

Entities at which measurements are made. (e.g., a blood sample or a single cell reading).

Example III:
Drug against parasite infection

BU: mouse

EU = groups of BU
mice treated with same parasite batch



OU
Survival of mice

Another classification approach for replicates

Definition

Biological Unit (BU)

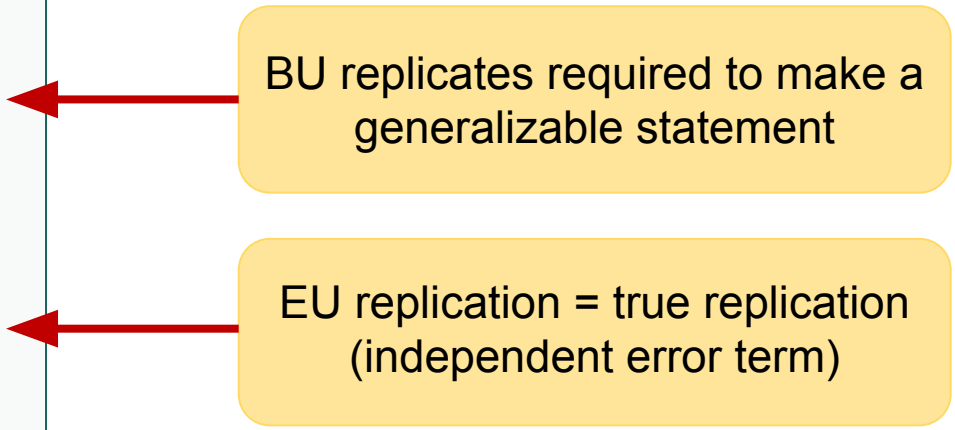
Entities we want to make inferences about (e.g., an individual or an animal).

Experimental Unit (EU)

The smallest entity that can be independently assigned to a treatment / condition (e.g., a single cage or a specific well).

Observational Unit (OU)

Entities at which measurements are made. (e.g., blood sample or single cell sequencing).

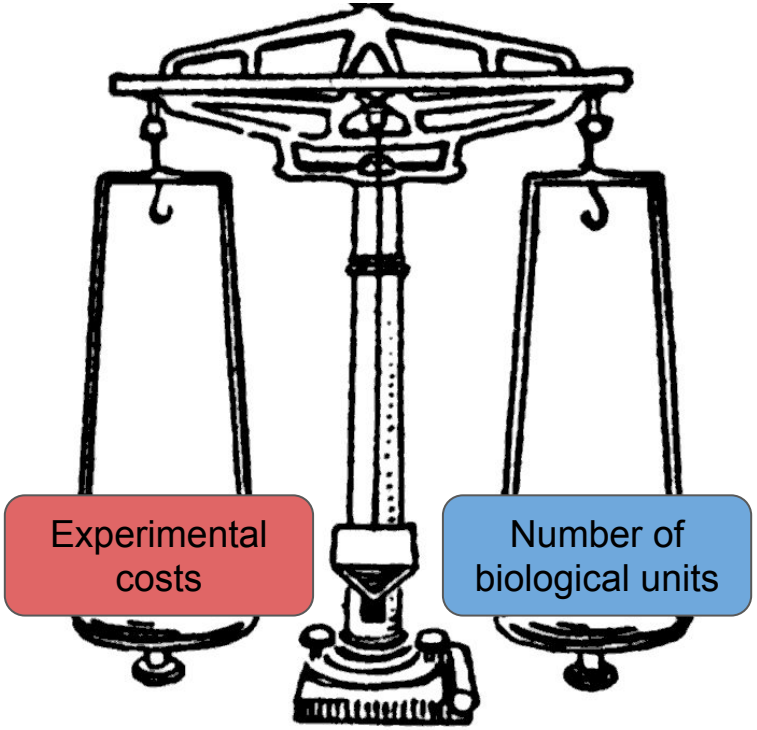


Learning objectives

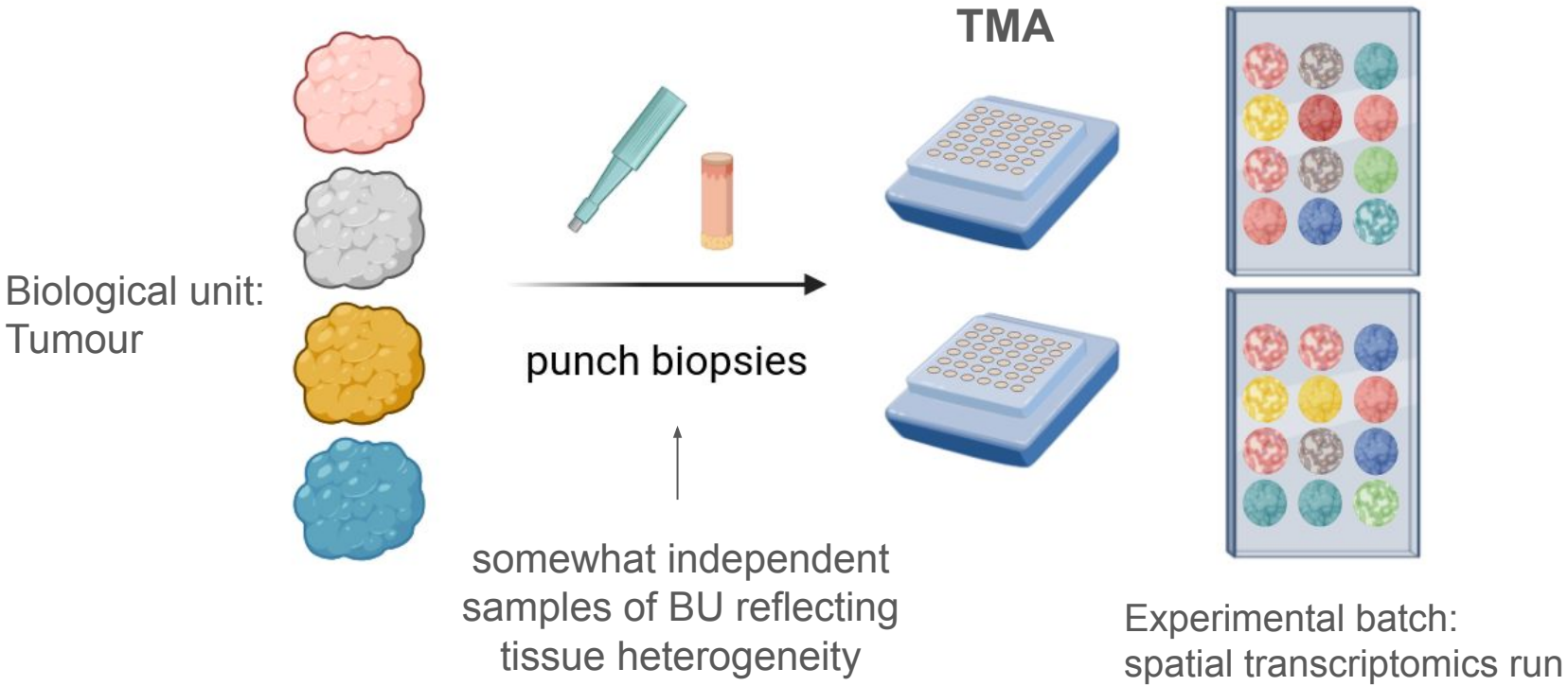
1. Confounding factors and batches

2. Replicate types and their application

3. Considerations for design of multi-omic experiments



Tissue microarrays as a tools for good experimental design

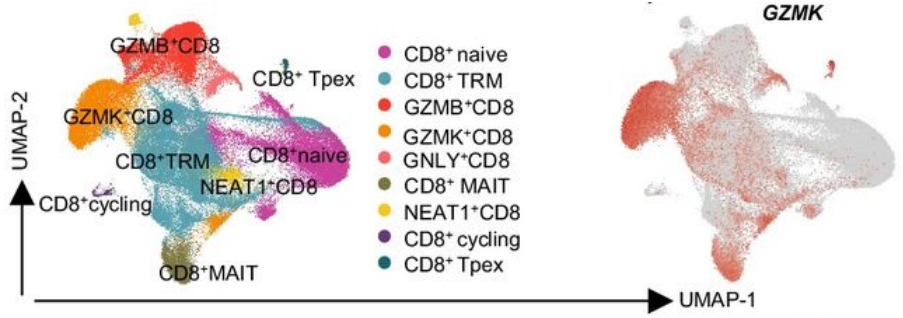


Example: Vannan et al. 2025. doi: 10.1038/s41588-025-02080-x

Combine expensive multi-omic techniques with alternative methods

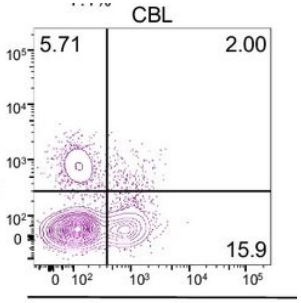
Hypothesis generation on transcriptome level

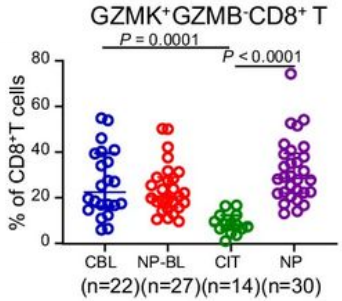
Single-cell sequencing
(n = 5, 8, 4, 8 per group)

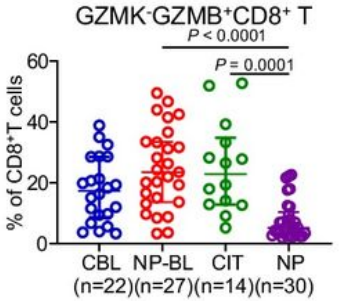


Validation on protein level

FACS
(n = 22, 27, 14, 30 per group)







Guo et al. 2024. doi: 10.1038/s41467-024-54685-1

Learning objectives

1. Confounding factors and batches

2. Replicate types and their application

3. Considerations for design of multi-omic experiments

4. Data Management
Reproducible research practices

Data Management

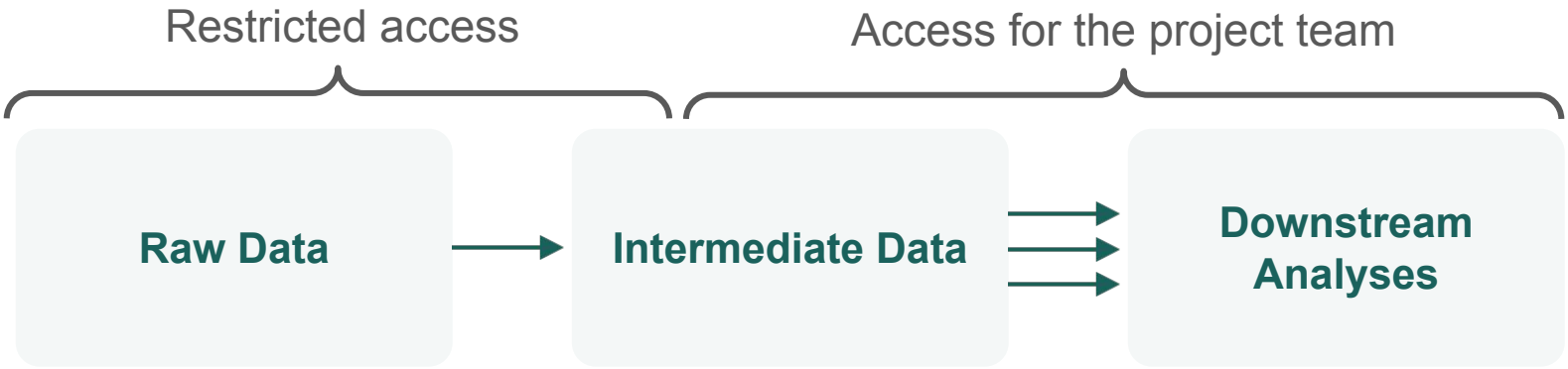
Start a comprehensive metadata annotation early on

- Use unique identifiers compatible with your sample information system (e.g. patient registries available in most hospitals/biobanks)
- Document experiments in electronic labbook and link to data (e.g. elabFTW)

data_id	Imaging date	treatment	experiment ID	Organoid ID	Tumour surgery	Patient ID	…
image1	14.06.2025	control	005_2025	ORG1	14.05.2025	837672	…
image2	14.06.2025	drug	005_2025	ORG1	14.05.2025	837672	…
image3	26.11.2025	control	005_2025	ORG1	14.05.2025	837672	
…	…	…	…	…	…	…	…

Observational unit Experimental unit Biological unit

Data Management Strategy



Central Storage Infrastructure

Use of common intermediate data object.

Sharing of code and results.

Reproducible research and version control

Use markdown reports to document your analyses (include data and software versions).

Use git for version control of analysis scripts.

Maintain a complete history of changes to your analysis scripts, to ensure reproducibility and transparency.