ZJE 浙江大学爱丁堡大学联合学院
ZJU-UoE INSTITUTE

Deconstruct an Experiment

IBMS3

Integrative Biomedical Sciences 3

Yuchen SHEN

March 31, 2026

Zhejiang University — University of Edinburgh Institute

Contents

1 Before-all preparation 2
2 Windows operations 2
3 Framework and answer elements 3
4 Experiment method/measurement approaches 4
5 Group settings & biological replicates 5
6 Data transformation/normalization 6
7 Statistical Analysis Method 8
8 Ethics 10
9 Potential Nuisance Variables/Factors or Biases 11
9.1 General biases . . . 11
9.2 General nuisance . . . 11
9.3 Sex-related biases . . . 12
9.4 Bias in methods . . . 13
9.5 For mouse model . . . 15
9.6 For cell line model . . . 15
9.7 For human patient model . . . 15
9.8 Randomized controlled trial (Drug test) . . . 15
10 Data analysis questions 17
10.1 Scripts . . . 17
10.2 Data interpretation . . . 17
10.3 Data transformation . . . 17
11 Practice questions (coding) 19
12 Practice questions (experimental design) 20
12.1 Learning from practice questions . . . 21

1 Before-all preparation

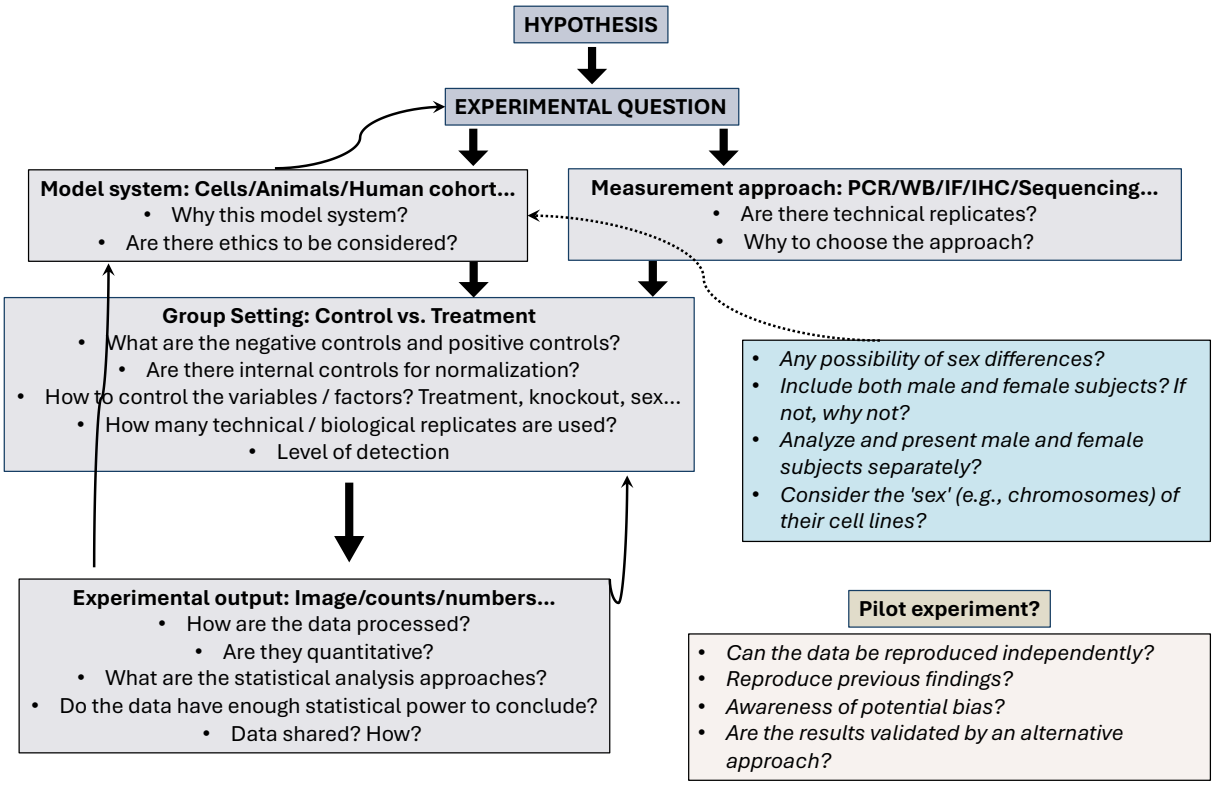
- **ID:** yuchen2.23@intl.zju.edu.cn
- **Password:** xV99@a73
- **Blackboard Learn:** learn.intl.zju.edu.cn

1. Open a word file and a R script file
2. Save them on desktop right away.
3. Do the experimental design questions first, among which the first one is the one encountered already.
4. For each question, draw a schematic first!

2 Windows operations

- Get the to-import file path
  - Unzip the file onto the desktop
  - The file path is on the upper path bar
  - When importing into R, replace the "\" to "/"
- When the blinking input cursor turns into this:_, hit the insert button on the right upper keyboard

3 Framework and answer elements



4 Experiment method/measurement approaches

This section is about answering *what are the steps?*

Answer elements:

1. Formal experiment name if it has one: gain-of-function, loss-of-function, luciferase assay, etc.
2. How to handle the model: Injection, transfection, weeks etc.
3. Collect what tissue? How to process the tissue?
4. What is the metric to measure? How to measure?
5. Briefly state the data format?

- **Randomized Controlled Trial:** Subjects are randomly allocated to groups. Both subjects and experimenters are blinded to the 'treatment'.
- **Cohort Study**
  - Retrospective: Analyses of data already collected for another purpose to test a new hypothesis
  - Prospective: data to be collected to test a specific hypothesis
- **Cross-Over:** Subjects receive both treatments in a random order, with a washout period in between. Each subject serves as their own control, which can increase statistical power and reduce variability.
- **Multi-Batch Study:** Conducting the same experiment in multiple batches to account for batch effects and increase the robustness of findings. Each batch should include all treatment groups to allow for proper comparison and control of batch-related variability.
- **Cross-Sectional:** Data collected at a single point in time to assess the prevalence of an outcome or characteristic in a population. This design is useful for identifying associations but cannot establish causality.
- **Longitudinal/Prospective:** Data collected from the same subjects over an extended period to observe changes and establish temporal relationships between variables.
- **Meta-Analysis:** A statistical analysis that combines the results of multiple scientific studies to derive a more precise estimate of the effect size or to identify patterns across studies. This approach can increase statistical power and resolve uncertainty when individual studies have conflicting results.

5 Group settings & biological replicates

This section is about answering *what is the experiment outline?*

Answer elements:

1. Group settings using bullet points
2. How are subjects grouped: randomly grouped (or stratified) with blinding to avoid bias as possible
3. Subject details: Age (8-12 weeks/2-6 months), sex, tumour size...
4. Describe biological replicates
    - Animals: 8-15 per group would be needed for statistical significance. The exact number should be determined based on power analysis.
    - Cell line: 3-6 replicates per condition would be needed for statistical significance. The exact number should be determined based on power analysis.
    - Human: The sample size would be relative large since the greater variability in human populations. The age, health situation, etc. should be similar to reduce the influence.
5. Describe technical replicates

Special control groups:

- SNP/gene manipulation: wild type allele
- Antibody-related: isotype control

6 Data transformation/normalization

Template: [xxx] should be normalized to [xxx] to account for differences in [A] and [B]

Answer elements:

1. Normalization: Endogenous reference/Internal control
   - Luciferase assay's Renilla luciferase activity, WB's loading control, qPCR's housekeeping gene
2. Normalization: Negative control
   - shRNA's scrambling control, lentivirus-mediated overexpression's empty vector control
3. Physiological data calculation
   - For example: Tumour volume calculation, cell proportion from scRNA-seq data

---

- For ELISA answer, see AI Set B Q3.
- For qRT-PCR (qPCR) answer, see AI Set B Q2 and 2026 formative Q2.
- For luciferase assay answer, see AI Set A Q2.
- For tumour answer, see AI Set A Q3.
- **Reference standard-based normalization**
  - **ELISA:** Quantification was performed using a standard curve generated from serial dilutions of the reference standard.
  - **qPCR (absolute quantification):** Absolute copy number was determined using a standard curve generated from plasmid DNA standards with known copy numbers.
  - **Drug concentration assay:** Drug concentration was determined using calibration standards, with an internal standard included to correct for extraction and analytical variability.
  - **Enzyme activity assay:** Enzyme activity was quantified against a calibration curve prepared with the corresponding product standard.
- **Baseline normalization**
  - **Cell proliferation assay:** Normalized to the initial value of each sample group.
  - **Wound healing assay:** Wound area at each time point was normalized to the initial wound area, and wound closure was calculated as the percentage decrease from baseline.
  - **Tumor growth assay:** Tumor volume was normalized to the volume at treatment initiation and presented as relative tumor volume V_t/V_0.
  - **Time-course assay:** Measurements were normalized to the baseline level of each subject and expressed as percent change from baseline.
- **Genetic engineering & interference**
  - **Non-targeting Control:** Use a non-targeting control (e.g., scrambled gRNA/siRNA) to account for any non-specific effects of the transfection process.
  - **Multi-target verification:** Use multiple independent gRNAs/siRNAs targeting different regions of the gene to confirm that observed effects are due to specific gene disruption rather than off-target effects.
- **Internal control**

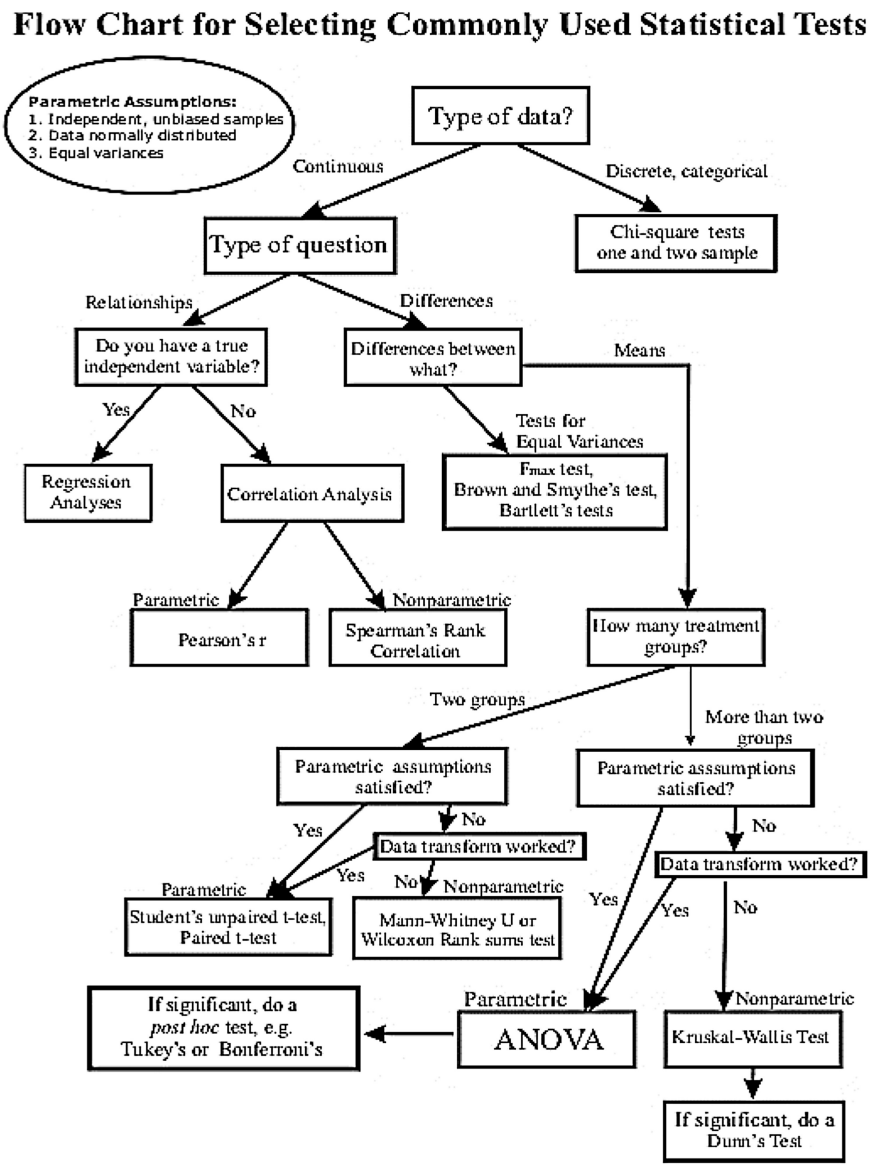
  - **WB:** Loading control (e.g., β-actin)
  - **IF:** DAPI to control for cell number
  - **qPCR (relative quantification):** Housekeeping gene (*GAPDH*) to normalize for variations in RNA input and reverse transcription efficiency.
  - **Flow cytometry:** Use unstained or isotype control to set gates and normalize fluorescence intensity measurements.

- **Immunohistochemistry (IF/IHC/IP/FC)**
  - **Isotype control:** Use an antibody of the same isotype as the primary antibody but without specificity to the target antigen to control for non-specific binding.
  - **Secondary-only control:** Omit the primary antibody and only use the secondary antibody to control for non-specific binding of the secondary antibody.

7 Statistical Analysis Method

Answer elements:

1. What test?
2. Optional: What assumption?
3. Performed between what groups' what data?
4. Aim for what difference?
5. If statistical's assumption is not met, use what test?



Do not forget multiple testing correction (e.g., Benjamini-Hochberg FDR) to control FP in high-throughput data!

- **Survival analysis**

  - **Kaplan-Meier curves:** Estimate survival time; the log-rank test shows differences in survival between two groups.
    * Survival probabilities were estimated using Kaplan-Meier curves, and differences between groups were assessed with log-rank tests.
  - **Log-rank test:** Compare survival distributions between two or more groups.
    * Example: Testing if a new drug (Group A) extends survival compared to standard therapy (Group B) in a randomized clinical trial.
    * Steps:
      · Define the event (e.g., death, relapse) and follow-up time
      · Record event times and censoring status (e.g., patients lost to follow-up)
      · Use the log-rank test to calculate a p-value for differences in survival curves

- **Cox proportional hazards model:** Evaluates the effects of multiple factors on survival.
  - Cox regression analysis was performed to evaluate the influence of multiple variables on survival outcomes while adjusting for confounding factors.

- **Randomized design**
  - **CRD (Completely Randomized Design):** one-way ANOVA, used to compare differences among treatments.
  - **RCBD (Randomized Complete Block Design):** two-way ANOVA (treatment + block), used to separate block effects. It separates variability into treatment effect, block effect (e.g., soil differences, time batches), and random error.

8 Ethics

Cell line: As this experiment involves cultured *in vitro* cell lines rather than live animals or human subjects, ethical considerations are limited, but it must still adhere to relevant institutional and national laboratory safety and bioethics guidelines, including proper handling of potentially biohazardous materials, documented ethical sourcing of all cell lines, verification of origin and consent where applicable, appropriate transfer or repository documentation, authentication of cell line identity, and routine contamination testing.

Animal models: All animal-use protocols would require approval by the Institutional Animal Care and Use Committee (IACUC) and would comply with animal welfare guidelines for the care and use of laboratory animals, including appropriate housing and enrichment, measures to minimize suffering through humane endpoints, anesthesia, and analgesia, and adherence to the 3Rs principles (Replace, Reduce, Refine). The number of mice would also be minimized while still achieving adequate statistical power.

Non-human primate research (e.g., monkeys): It would need IACUC approval, compliance with national and international regulations for primate research, environmental enrichment and social housing where possible, procedures to minimize pain and distress through appropriate anesthesia and analgesia, and use of the minimum number of animals required for scientific validity.

Human cohort: It would need IRB approval, conduct in accordance with the Declaration of Helsinki, written informed consent from all participants or legal guardians when applicable, protection of privacy through anonymization and secure data handling, minimization of study-related risks, and the right of participants to withdraw at any time.

Human primary cell line: It would need IRB-approved collection procedures, informed donor consent covering research use, anonymization of donor information, compliance with applicable tissue-use regulations, and safeguards against misuse of donor materials.

Human medical data: It would need IRB approval, informed consent clearly stating the purpose, risks, and scope of data use, pseudonymization or anonymization of personal identifiers, secure encrypted storage, and restricted access to authorized researchers only.

Human imaging data: It would need IRB approval and participant consent for collection, storage, and research use, anonymization of imaging data through removal of metadata and identifiable facial features where relevant, secure storage in compliant systems, and renewed ethical review for secondary use when required.

Public datasets integration: It would need compliance with the original data provider's ethical and data-use terms, no attempts at re-identification, proper citation and compliance statements, and local IRB exemption or approval for secondary analysis where required.

9 Potential Nuisance Variables/Factors or Biases

9.1 General biases

- **Sampling/Ascertainment/Selection bias:** The sample (data) you get *in the end* does not represent the population, because the process by which the sample enters the pool is not random.
  - Solutions:
    * Define the population from which you want to draw a conclusion
    * Define your inclusion and exclusion criteria
    * Split treatment and control groups after exclusion have been done
    * Run a pilot study first to discover potential problems
- **Presentation/Reporting bias:**
  - y-axis hacking by not starting from zero or increasing gaps
  - Choosing a pie chart over a bar chart when it comes to comparing over a time course
  - Not using Cook's distance to evaluate the effects of outliers on the average level
  - Not showing the data distribution with jitter/density/violin plot
- **Model bias (generalization):**
  - Conclusions drawn from one strain of model organism cannot necessarily be generalized to other strains
  - Conclusions drawn from a lab experiment cannot necessarily be generalized to natural or clinical conditions
  - Conclusions drawn from a cell line cannot necessarily be generalized to primary cells or *in vivo* conditions

9.2 General nuisance

1. **Stratification:** Apply stratification for critical variables (e.g., grouping participants by exercise experience before randomization).
2. **Blinding:**
   - Single-blind: Participants are unaware of their group assignment.
   - Double-blind: Both participants and researchers are unaware of group assignments (e.g., using coded labels instead of explicit group names).
3. **Pilot testing:** Conduct small-scale preliminary studies to identify potential issues (e.g., participant dropout patterns or measurement inconsistencies).
4. **Circadian rhythm effects:** Ensure consistency in data collection at fixed times of day to control for circadian rhythm effects.
5. **Cage effects:** Randomize animals across multiple cages and include cage as a random effect in statistical analyses to control for cage-related variability.
6. If only have marker gene expression, mRNA or protein level data: Functional assays such as cell proliferation, apoptosis, migration, or invasion assays can be performed to validate the functional relevance of the observed molecular changes and strengthen the conclusions.
7. Only proved necessity but not sufficiency: Gain-of-function experiments (e.g., OE) can be performed to demonstrate that the gene/protein is not only necessary but also sufficient to produce the observed phenotype, providing stronger evidence for its role in the biological process.
8. Only proved sufficiency but not necessity: Loss-of-function experiments (e.g., KO or KD) can be performed to demonstrate that the gene/protein is necessary for the observed phenotype, providing stronger evidence for its role in the biological process.
9. Only proved correlation but not causation: Use genetic manipulation (e.g., CRISPR-Cas9) or pharmacological inhibition to directly test the causal relationship between the gene/protein and the phenotype, thereby strengthening the evidence for causality.

10. No rescue experiment: Perform a rescue experiment by reintroducing the gene/protein of interest into the KO/KD model to confirm that the observed phenotype can be reversed, providing strong evidence for the specific role of the gene/protein in the biological process.
11. Will the mouse manipulation protocol cause stress? For example, intraperitoneal injection can cause stress to mice, which can affect the experimental outcomes. To control for this, a vehicle control group that receives the same injection without the drug should be included to account for any effects of the injection procedure itself.
12. If the experiment involves testing a drug, it is possible that there is drug viability inconsistency across batches, which can be controlled by using the same batch of drug for all experiments and including a positive control to verify drug activity.
13. Only one indicator of a broad spectrum is used: For example, only one SASP marker, only one of the cytokine panels.
14. Sex-specific difference in the face of disease is very common, for example, drug-induced liver injury.
15. If use one drug to induce a certain disease model, it is possible that drug-induced disease does not fully represent the disease model.
16. If the drug is microbiome, then the viability might be different.

9.3 Sex-related biases

- **Genetics and genomics**
  - Some Y genes have X homologues but are subtly different. Some Y genes have no X homologues at all.
  - Most genes on the silenced X are inactivated, but some are not... females have double gene dosage.
- **Hormone system**
  - Different gonads
  - Different endocrine systems (Testosterone binds with AR, Estradiol binds with ER). These nuclear receptors up- or downregulate different genes.
- **General differences**
  - Disease
  - Pathophysiology
  - Manifestation
  - Response to treatment
- **Specific differences**
  - Women have higher percentage of adipose tissue (and different distribution)
  - Men have higher muscle mass
  - Men have more RBC, women have more WBC and produce antibodies faster
  - Men's skin is thicker, oilier and has more terminal hair
  - Women have a higher HR and lower BP
- **Pharmacokinetics in women is affected by**
  - Lower body weight
  - Slower gastrointestinal motility
  - Less intestinal enzymatic activity
  - Slower glomerular filtration rate
- **Drugs that women are more sensitive to**
  - Beta blockers
  - Opioids
  - Selective serotonin reuptake inhibitors

  - Antipsychotics
- **Women are 50 to 75% more likely to experience an adverse drug reaction**

9.4 Bias in methods

- **siRNA/shRNA:** Off-target effects. Use multiple independent siRNAs targeting different regions of the gene to confirm that observed effects are due to specific gene knockdown rather than off-target effects.
- **KD/KO/OE:**
  - Efficiency and effects should be validated by WB or qPCR.
  - Perform a reconstitution experiment where you insert again an active version of the gene that was KO/KD to verify that the KD/KO only affected that single gene.
  - To confirm that KD/KO system used did not disrupt other pathways and it is itself responsible for any new phenotype, a control targeting a random sequence or an intergenic sequence in a gene-depleted region should be included.
- **CRISPR-Cas9:**
  - Off-target effects. Use validated guide RNAs and perform off-target analysis. Control cell lines without the target modification but subjected to the CRISPR process will be included. At least two independent sgRNAs targeting the same locus will be used to confirm phenotype reproducibility.
  - Validate the successful CRISPR-Cas9 by sequencing
- **CRISPRi:**
  - Design 3-5 gRNA spanning the region of interest and separately into cells to confirm the effect is due to the specific region rather than off-target effects. Control is gRNA outside this region to confirm that the effect is not due to the CRISPRi system itself.
- **WB:** Antibody specificity, protein loading variations, and subjective band quantification could introduce bias. We will validate antibodies, use loading controls, and implement automated quantification software with blinded analysis.
- **qPCR:** Primer efficiency, reference gene stability, and threshold cycle determination can affect results. We will validate primers, use multiple reference genes, and apply consistent threshold settings across all samples.
- **RNA-seq:** Library preparation biases, batch effects, and computational analysis choices can influence differential expression results. We will prepare libraries simultaneously, include spike-in controls, and apply multiple normalization methods to ensure robustness.
- **Flow cytometry:** Gating strategies, fluorophore compensation, and instrument calibration can introduce variability. We will use fluorescence-minus-one controls, consistent gating protocols, and regular instrument quality control.
- **Microscopy:** Selection bias in field of view, inconsistent exposure settings, and subjective image analysis can skew results. We will implement systematic random sampling, standardized acquisition parameters, and automated quantification algorithms.
- **Animal behavior tests:** Observer presence, time of day, and environmental conditions can affect behavioral outcomes. We will conduct tests at consistent times, use automated tracking systems when possible, and maintain consistent environmental conditions.

- **Mass spectrometry:** Sample preparation variability, ion suppression, and peak identification can bias metabolomic or proteomic analyses. We will include internal standards, run quality control samples, and use multiple technical replicates.
- **ELISA:** Plate position effects, edge effects, and inconsistent washing can lead to systematic errors. We will randomize sample positions, include standard curves on each plate, and use automated washing systems when available.
- **ChIP-seq:** Antibody specificity, chromatin fragmentation variability, and PCR amplification biases can affect results. We will validate antibodies using knockout controls, optimize sonication protocols, and minimize PCR cycles.
- **Single-cell analyses:** Cell isolation methods, doublet contamination, and computational clustering algorithms can bias population identification. We will implement doublet removal strategies, include control populations, and apply multiple clustering approaches.
- **Histology:** Tissue processing artifacts, staining variability, and region selection can skew quantification. We will process all samples simultaneously, include positive and negative controls, and implement systematic sampling of tissue sections.
- **Bioinformatic analyses:** Algorithm selection, parameter settings, and reference database choices can significantly impact results. We will apply multiple analytical approaches, perform sensitivity analyses with different parameters, and validate key findings using alternative methods.
- **Metabolic assays:** Circadian rhythm effects, feeding status, and stress responses can influence metabolic measurements. We will standardize testing times, fasting conditions, and acclimation periods prior to measurements.
- **Transient transfection:** Variable transfection efficiency between samples can affect protein expression levels. We will include transfection controls, optimize conditions, and normalize results to transfection efficiency.
- **Luciferase assays:** Substrate degradation, cell number variations, and instrument sensitivity can bias reporter activity measurements. We will use fresh substrate, normalize to cell number or protein content, and include internal control reporters.
- **Co-IP:** Non-specific binding, antibody cross-reactivity, and wash stringency can affect protein interaction results. We will include IgG controls, validate antibodies, and optimize wash conditions for each interaction.
- **Organoid cultures:** Batch-to-batch matrix variability, size heterogeneity, and differentiation state can influence outcomes. We will use matrix from the same lot, standardize seeding density, and characterize differentiation markers.
- **Calcium imaging:** Dye loading variability, photobleaching, and motion artifacts can bias signal quantification. We will normalize to baseline fluorescence, correct for bleaching, and implement motion correction algorithms.
- **Electrophysiology:** Electrode placement, seal quality, and cell health can affect recording quality. We will establish minimum seal resistance criteria, monitor access resistance, and exclude recordings with unstable baselines.
- **In vivo imaging:** Anesthesia effects, tissue depth, and motion artifacts can influence signal detection. We will standardize anesthesia protocols, implement depth correction, and use respiratory gating when applicable.
- **Pharmacological studies:** Drug solubility, stability, and off-target effects can confound interpretations. We will prepare fresh solutions, verify activity, and use multiple structurally distinct compounds targeting the same pathway.

9.5 For mouse model

Potential variables could include the age and sex of the mice, their genetic background, and environmental factors such as housing conditions, diet, microbiome composition, and stress levels.

To minimize these factors, we would use mice of the same age and sex, and from the same genetic strain. In addition, randomization of the mice into the control and treatment groups can also help to minimize any potential biases. Blinding of the researchers during the data collection and analysis stages can also help to reduce bias.

9.6 For cell line model

- Transfection efficiency between wells: Internal control
- Batch effects in plasmid preparation: Multiple independent plasmid preparation
- Batch effects in reagents and timing of experiments could introduce variability, which will be controlled by using reagents from the same lot when possible and including appropriate controls in each experimental batch.
- The choice of cell type: Validate in a secondary cell line
- Animal cell line cannot fully recapitulate human biology: Validate in primary human cells
- Differences in cell culture conditions, such as variations in media composition, temperature, or CO_2 levels: Ensure that all cell culture conditions are identical for both cell lines.
- Differences in passage number, cell density, and genetic drift over time.: Regular authentication of cell lines and consistent passage protocols will be implemented.
- Clonal variation within cell lines: Use multiple clones for each cell line to ensure that observed effects are not due to clonal artifacts but rather reflect the phenotype of interest.
- If using patient-derived cell line, there might be donor-to-donor variability: Use multiple patient-derived cell lines to ensure that findings are generalizable and not specific to a single donor's genetic background or disease state.
- Cell line cannot faithfully capture chromatin structure and cellular context: Validate findings with *in vivo* models to ensure that the observed effects are not artifacts of the cell culture system.

9.7 For human patient model

Patient demographics including age, sex, ethnicity, socioeconomic status, and comorbidities may influence outcomes and will be recorded and considered during analysis.

Treatment history, medication use, and lifestyle factors (diet, exercise, smoking status) could confound results and will be documented.

Selection bias may occur if recruitment strategies favor certain patient populations. We will implement broad inclusion criteria and diverse recruitment methods.

Recall bias in patient-reported outcomes will be minimized through validated questionnaires and prospective data collection when possible.

Loss to follow-up could skew longitudinal data and will be addressed through multiple contact methods and incentives for study completion.

9.8 Randomized controlled trial (Drug test)

- **Placebo:** Supposed to be a neutral control intervention, but can be associated with positive effects (the placebo effect)
- **Nocebo:** a subject experiences negative effects when given the control intervention
- **Hawthorne effect:** subjects change behavior in response to being observed

- **Demand characteristics:** effects that occur from subjects guessing the aims of the research and acting accordingly
- **'Screw-you' effect:** subjects act in a way that might sabotage the experimenter's aims

10 Data analysis questions

10.1 Scripts

1. Load packages
2. Load data
3. Data cleaning
4. Process data (technical replicates & metrics calculation)
    - Background mean
    - Technical replicates (mean)
    - Subtract background value (normalization)
    - Metrics calculation: OD-concentration mapping is possible
5. Statistical analysis (biological replicates)
6. Visualization

10.2 Data interpretation

- State hypotheses
- First interpret the data *intuitively*, for example, interpret FC
- Report the computed p-value
- State whether rejecting or accepting null hypothesis
- State conclusion
- Critiques (Nuisance/Bias section can be applied here): (1) Sample size, (2) Potential biases for example WB is a semi-quantitative method,
- Follow-up experiments

10.3 Data transformation

10.3.1 qPCR

10.3.2 PCR Efficiency

From the standard curve, fit a linear regression between the mean Ct values and the $\log_{10}$ of the dilution factor:

$$Ct = \text{slope} \times \log_{10}(\text{Dilution Factor}) + \text{intercept}$$

The PCR efficiency is then calculated as:

$$E = 10^{-1/\text{slope}}$$

Ideal efficiency is 2 (i.e., doubling each cycle), corresponding to a slope of approximately -3.32.

10.3.3 Absolute Quantification (Standard Curve Method)

The standard curve is inverted so that concentration can be predicted from Ct. Concentration is expressed as $\log_{10}(1/\text{Dilution Factor})$, and a new linear model is built: $\log_{10}(\text{Concentration}) = \text{slope} \times Ct_{\text{mean}} + \text{intercept}$.

For unknown samples, the predicted concentration is: $\text{Concentration} = 10^{(\text{slope} \times Ct_{\text{sample}} + \text{intercept})}$

This allows the estimation of the absolute amount of mRNA in each sample by interpolation from the standard curve.

10.3.4 Relative Quantification ($\Delta\Delta Ct$ Method)

The calculation proceeds in three steps:

Step 1: Calculate ΔCt

$$\Delta Ct = Ct(\text{GeneX}) - Ct(\text{GAPDH})$$

Step 2: Calculate $\Delta\Delta Ct$

$$\Delta\Delta Ct = \Delta Ct(\text{sample}) - \text{mean } \Delta Ct(\text{Ctrl})$$

Step 3: Calculate Relative mRNA Level

$$\text{Relative mRNA level} = 2^{-\Delta\Delta Ct}$$

The control group's relative expression normalizes to approximately 1, and treatment groups are expressed as fold-change relative to the control.

11 Practice questions (coding)

Question	Data	Additional steps	Test data	Test
AI-A	MTT assay OD value	Calculate FC	Direct DMSO vs. CompK OD value	t-test (var.equal = T)
AI-B	WB intensity quantification	Calculate FC	Vehicle (normalized to beta-actin) vs. FlavP (normalized to beta-actin) WB band intensity	t-test (var.equal = T)
AI-C	ELISA	Standard curve log10 linear model (subtracted background)	Vehicle + LPS (standard curve transformed concentration) vs. Berry + LPS (standard curve transformed concentration) vs. Vehicle (standard curve transformed concentration)	One-way ANOVA + post hoc
2026 F1	ELISA	Standard curve log10 linear model (subtracted background)	Placebo (standard curve transformed concentration) vs. Drug Z (standard curve transformed concentration)	t-test (var.equal = F)
2026 F2	qPCR	-	Vehicle ($\Delta\Delta Ct \rightarrow 2^{-\Delta\Delta Ct}$ value) vs. HormZ ($\Delta\Delta Ct \rightarrow 2^{-\Delta\Delta Ct}$ value)	t-test (var.equal = T)

12 Practice questions (experimental design)

Question	Type	Model	Manipulation	Measurement	Notes
Test drug's effects by injection *in vivo* (2021)	Drug response by indicator	Mouse	Drug injection	Serum metabolite	-
Test drug's effects on a HFD-induced disease model (2025)	Drug response by indicator	Mouse	Drug injection, diet feeding	ELISA	-
Prove causality from PPI correlation data (2022)	Gene-phenotype	Mouse	KO, KD, OE	ND phenotype	PPI master protein
Prove kinase X phosphorylates protein Y using *in vitro* and *in cell* experiments (2022)	PPI	Chemical, cell line	-	WB, radioactivity, IF co-localization, Co-IP	Kinase assay
Two proteins: Redundant or synergy? (2025)	Kinase	Mouse	Kinase inhibitor	Cell proliferation, apoptosis	Tumour-related, kinase
Functional genomics for gene candidate-phenotype validation (2025)	Gene-phenotype	Cell line	si/shRNA, drug treatment	Fluorescence	Functional genomics
Relationship between a mutation and cancer (2025 & 2026)	Mutation-phenotype	Cell line	CRISPR-Cas9	Cell proliferation	-
Test the function of an enhancer (2022 & 2026)	Regulatory element function	Cell line	Plasmid construction; CRISPRi; 3C	Luciferase assay; RT-qPCR; data	-
Test probiotic's effects on a toxicity-induced injury (2025)	Drug response by indicator	Mouse	Probiotic injection, toxicity	ELISA	-
Test whether SNP affects promoter activity (AI-A)	Regulatory element function	Cell line	Plasmid construction	Luciferase assay	-
Receptor's effects in TAM's tumour suppressive effects (AI-A)	Antibody-phenotype	Mouse	Antibody technique	Flow cytometry	Tumour-related
Test lncRNA's necessity in Tgfb pathway (AI-B)	Gene-phenotype	Cell line	siRNA	qRT-PCR, WB	Immune population
Clock gene's role in the face of LPS (AI-B)	Gene-phenotype	Mouse	Conditional KO	ELISA	Inflammatory cytokines
Sox9's effect in therapy resistance (AI-C)	Gene-phenotype	Cell line	shRNA, lentiviral OE	Luminescent ATP-based viability	Cell viability
Microbiome's effect in resisting PD (AI-C)	Germ-phenotype	Mouse	Microbiome transplantation	Multi-indicators	-
Epigenetic memory (2025)	Drug response by indicator	Cell line	Cytokine treatment	BS-seq	Time-course

Dry-lab questions:

- GWAS: 2024 formative question 1

- Loci-phenotype: 2022 formative question 5
- Model-based questions: see below

12.1 Learning from practice questions

- **KO/KD:**
  - Categories
    - Total KO by homologous recombination gene targeting or CRISPR/Cas9
    - Conditional KO by Cre-LoxP system or tamoxifen-inducible Cre-lox (temporal)
  - (1) WB of the protein (gene X's product) would be needed to validate if the system is correctly constructed. WB would require its own loading control (β-actin) to normalize the results.
  - (2) Targeting a random sequence or an intergenic sequence in a gene-depleted region to control that KO/KD does not disrupt other pathways and it is itself responsible for any new phenotype.
  - (3) Must have a reconstitution experiment where you insert again an active version of the gene that was KO/KD to verify that the KD/KO only affected that single gene.
- **OE:**
  - To control that the phenotype A can be measured: Activate a gene known to cause phenotype A
- **Phosphorylation:**
  - Two types of antibody that specifically bind the pY and Y $\rightarrow$ WB.
  - Radioactive phosphate group donor (ATP) $\rightarrow$ Autoradiography.
- **IF co-localization:**
  - Positive: DAPI
  - Negative: Not adding anything
- **Co-IP:**
  - Positive: IP with antibody against a known interactor
  - Negative: IP with IgG
- **Tumour establishment:**
  - Randomize *based on tumour size* to ensure unbiased starting tumour volumes across groups
  - Measure tumor volume & weight
  - Measure body weight as a toxicity indicator
  - Western blot of tumor lysates
- **Cellular proliferation:**
  - Ki-67
  - Caspase-3
  - DAPI (to control for cell number)
- **Staining:**
  - Stain target A and *DAPI*
  - Quantify mean target A intensity per cell, normalized to cell number via *DAPI*
- **siRNA experiment:**

  - Positive: siRNA targeting a gene known to be involved in the phenotype of interest.
  - Negative: Scrambled siRNA to account for any non-specific effects of siRNA transfection.

- **Transfection:**
  - Transfection efficiency
  - Use multiple replicates, and use the normalization approach with the internal control to account for variations in transfection efficiency and cell number

- **Measure T-cell killing capacity:**
  - Co-culture tumor cells with antigen-specific T cells. Perform cytotoxicity assay.
  - Measure T-cell activation markers
  - Measure release of IFN-γ, Granzyme B, TNF-α
  - Measure tumor cell death rate

- **Serum metabolite measurement:**
  - Collect blood serum via bleeding under anesthesia.

- **Experiment types:**
  - *in vivo*: in living model organism
  - *in cell*: use live cells
  - *in vitro*: cell line or chemical systems
  - *in silico*: computer simulation

- **Marker gene expression:** Both mRNA and protein level data are needed.

- **LPS challenge:**
  - Cell line experiment: Cultivate with LPS and collect supernatant
  - Animal experiment: Inject a sublethal dose of LPS and collect serum

- **Gene-phenotype related experiment:**
  - As long as it is RNA, it can be siRNA KD, even though lncRNA
  - Confirm genes and phenotype to measure. Sometimes phenotype can be obscure, such as "therapy resistance".
  - Necessity and sufficiency
    * Necessity: Take it away (KD/KO)
    * Sufficiency: Give it back (OE)

- **Drug cures disease** design experiment is pure disease control, pure healthy control.

- **Cell proliferation/apoptosis/cell number experiment:**
  - Control groups: Initial cell seeding densities.
  - Tumour/cell number experiment applied to this answer template.

$$\text{Cell number} \begin{cases} \text{Proliferation} & \begin{cases} \text{MTT assay} \\ \text{Ki-67 staining} \end{cases} \\ \text{Apoptosis} & \begin{cases} \text{Apoptosis assay (annexin V/PI staining)} \\ \text{Caspase-3 staining} \\ \text{Colony formation assay} \end{cases} \end{cases}$$

- **Model-based questions:**
  - Data collection and preprocess
    - Collect basic information
    - Feature engineering: identification of the training features and target label
    - Preprocess: normalization, data type transformation
    - If we need more data/information, we could design an experiment to get them
  - Model selection
    - Deep-learning based model
    - For that you will need a test set to make sure your model is able to predict the outcome. Keep a subset (~20%) of the data separate to test your model (test set), and another subset (~80%) of the data separate as the training set.
    - Can conduct several models, to choose the better model for this task
    - You should be aware of not overfitting your model: including more and more variables to your models is not necessarily good. Your model may be robust to variables-associated noise.
  - Model evaluation
    - Evaluate the performance of your model using appropriate metrics (e.g., accuracy, precision, recall, F1 score for classification; mean squared error for regression).
    - Use cross-validation to assess the generalizability of your model and to prevent overfitting.
  - Wet lab verification
    - Use wet lab test to verify our model