Experimental Planning

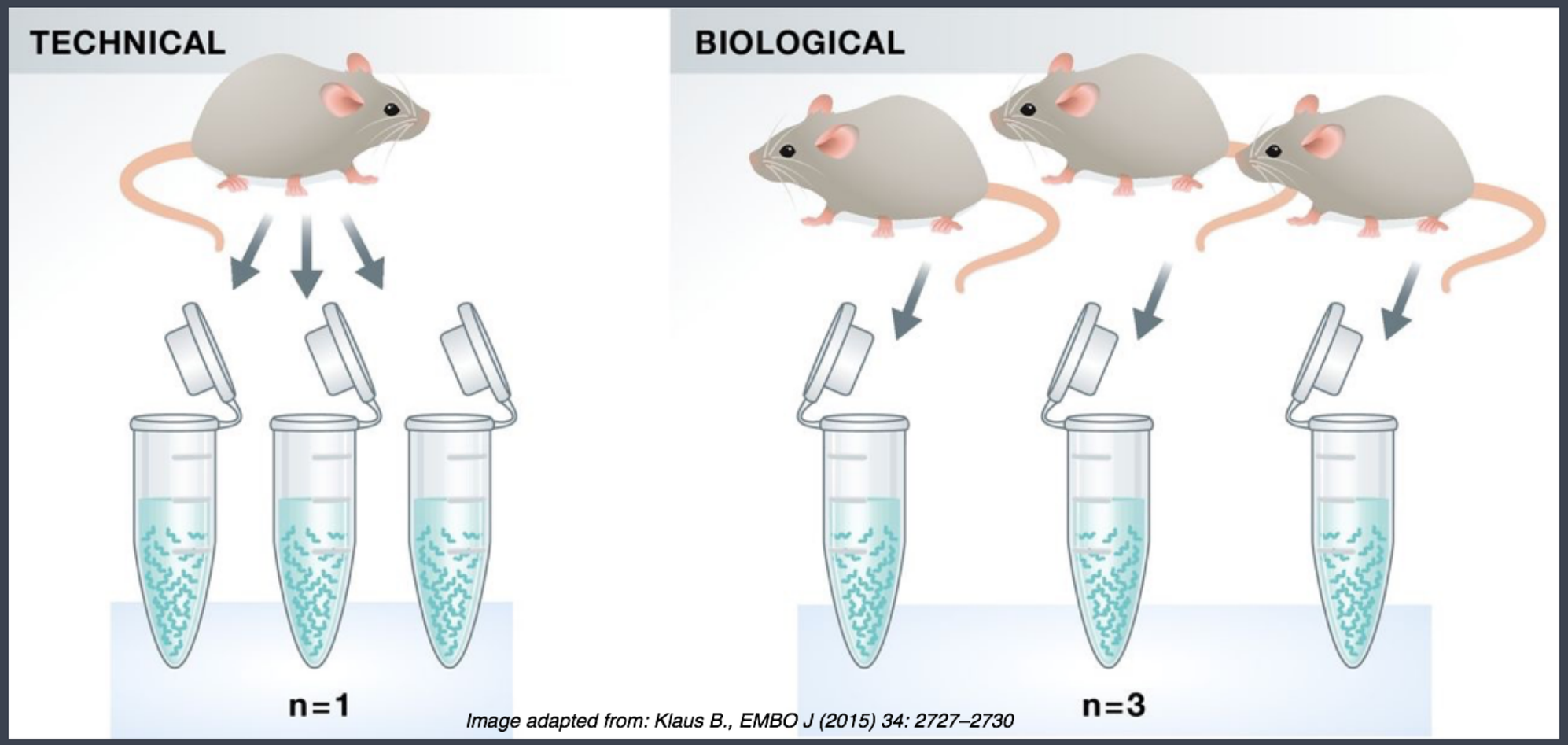


Image adapted from: Klaus B., EMBO J (2015) 34: 2727–2730

Experimental planning considerations

1. Replicates
2. Confounding factors
3. Batch effects

Experimental planning considerations: Replicates

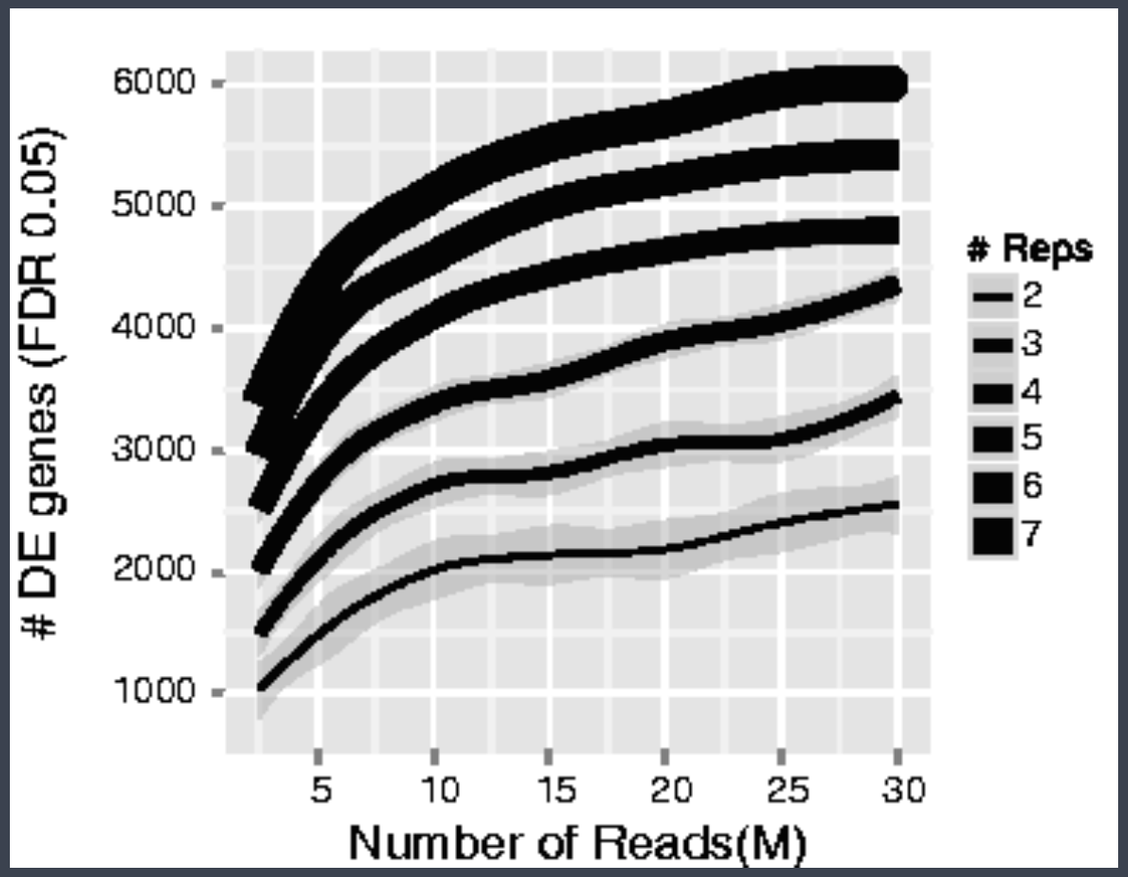
1. Replicates

- Technical replicates: Illumina has low technical variation unlike microarrays, hence technical replicates are unnecessary.

- Biological replicates, are absolutely essential. Have at least 3!
  - For cultured cells try to incorporate as much variation as possible
  - Avoid pooling RNA from multiple biological replicates

Experimental planning considerations: Replicates

More replicates yield more accurate analyses:



Liu, &., et.al., *Bioinformatics* (2014) 30(3): 301-304, https://doi.org/10.1093/bioinformatics/btt688

Experimental planning considerations: Replicates

Replicates versus sequencing depth:

- Replicates are almost always preferred to greater sequencing depth for bulk RNA-Seq.
- General differential gene expression:
  - ENCODE guidelines suggest 30 million SE reads per sample (stranded).
  - 15 million reads per sample is often sufficient, if adequate number of replicates.
  - Spend money on more biological replicates, if possible.

Experimental planning considerations: Replicates

Replicates versus sequencing depth:

- Detection of lowly-expressed genes:
  - Similarly benefits from replicates more than sequencing depth.
  - However, would likely want to sequence deeper with at least 30-60 million reads.

Experimental planning considerations: Replicates

Replicates versus sequencing depth:

- Isoform-level differential expression:
  - More depth (> 60 million reads per sample) and paired-end reads are suggested.
  - However, choose biological replicates over paired/deeper sequencing.
  - Perform careful QC of RNA quality. Be careful to use high quality preparation methods and restrict analysis to high quality RIN # samples.

Experimental planning considerations: Replicates

Replicates versus sequencing depth:

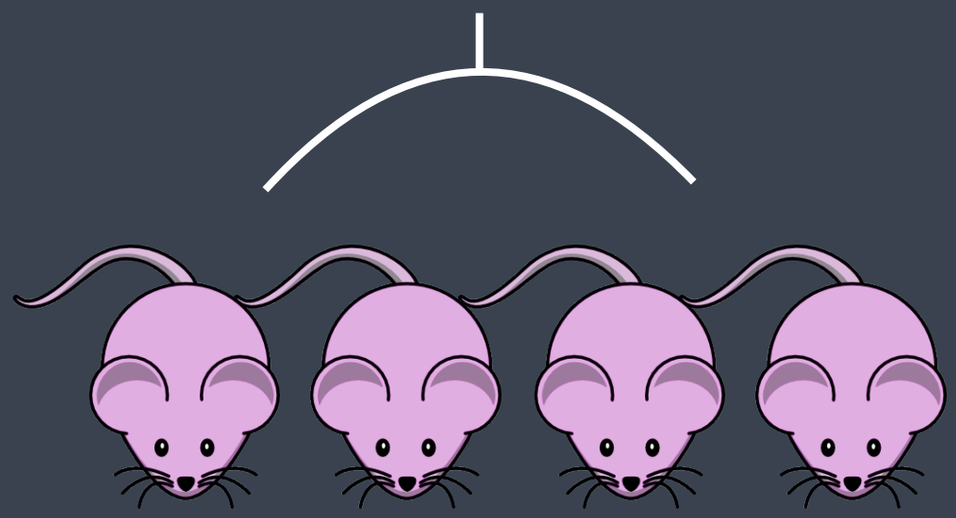
- Other types of RNA analyses (intron retention, small RNA-Seq, etc.) will have different suggestions.

  - But almost always more biological replicates is better!

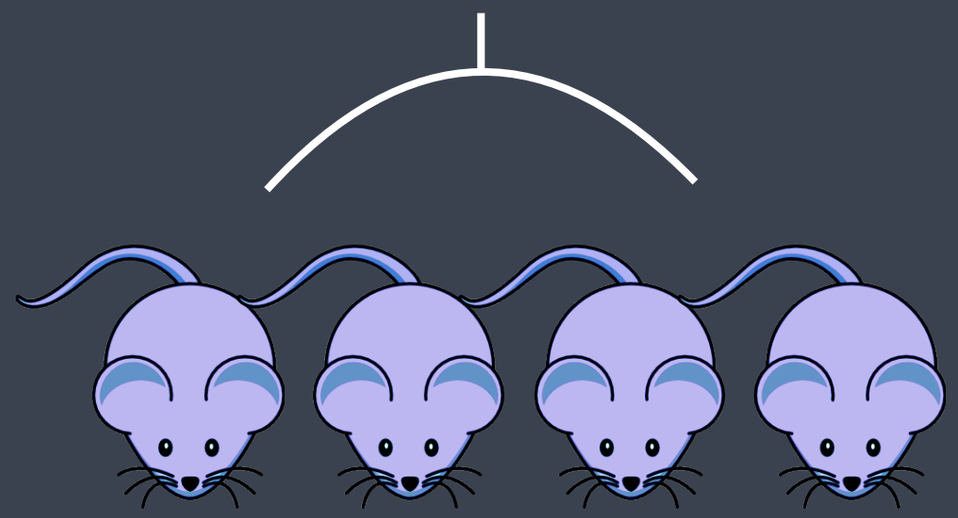
Experimental planning considerations: Confounding

2. Confounding factors

Control group



Treatment group

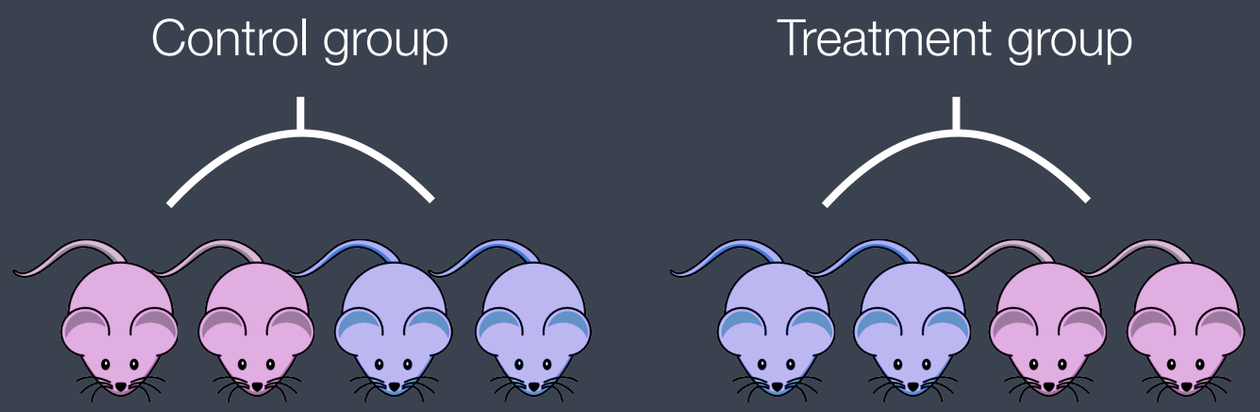


Experimental planning considerations: Confounding

Avoiding confounding:

- Ensure animals in each condition are all the same sex, age, litter, and batch, if possible

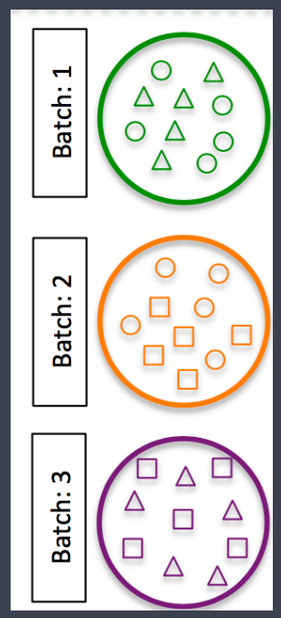
- If not possible, then ensure to split the animals equally between conditions

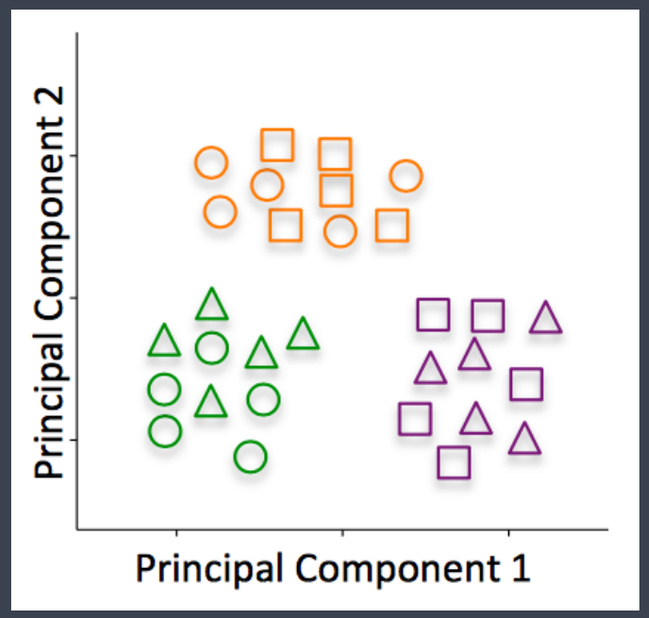


Experimental planning considerations: Batches

3. Batch effects

Batch effects are a significant issue for RNA-Seq analyses. You can see significant differences in expression due to batch effect.





Adapted from Hicks SC, et al. bioRxiv. 2015. doi:10.1101/025528.

Experimental planning considerations: Batches

How to know whether you have batches?

- Were all RNA isolations performed on the same day?
- Were all library preparations performed on the same day?
- Did the same person perform the RNA isolation/library preparation for all samples?
- Did you use the same reagents for all samples?
- Did you perform the RNA isolation/library preparation in the same location?

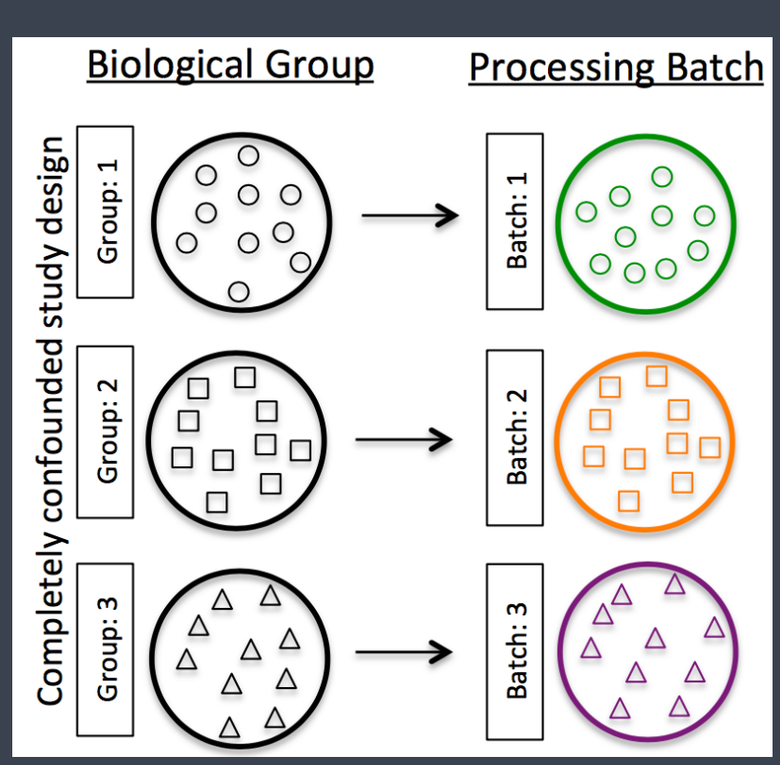
If any of the answers was ‘No’, then you have batches.

Experimental planning considerations: Batches

Best Practice

- Design the experiment in a way to avoid batches, if possible.

- If unable to avoid batches, do NOT confound your experiment by batch

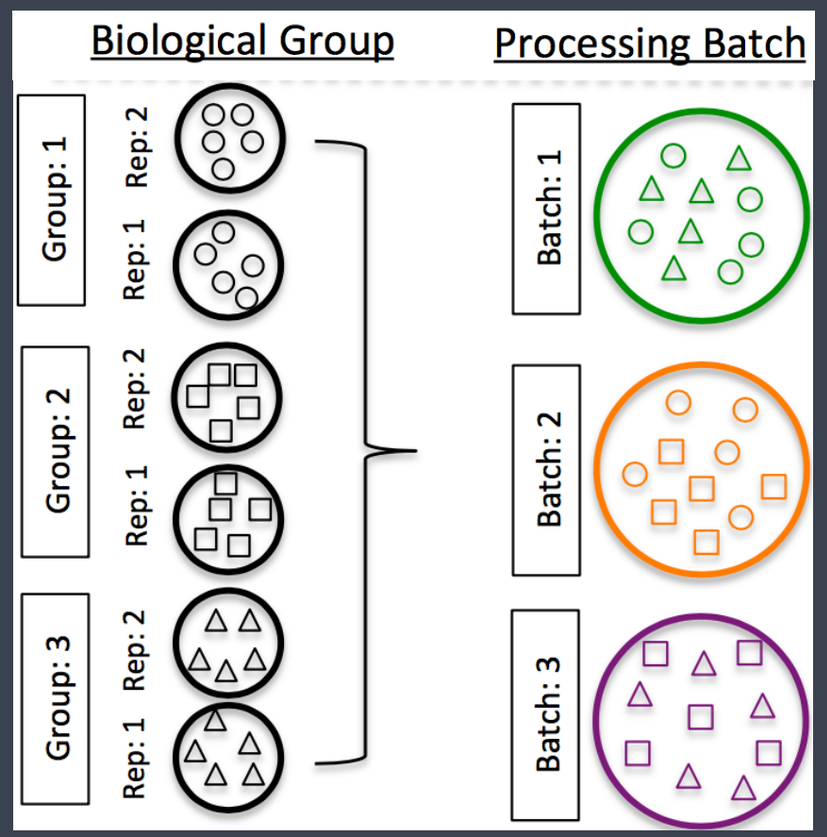


Adapted from Hicks SC, et al. bioRxiv. 2015. doi:10.1101/025528.

Experimental planning considerations: Batches

Best Practice

- If unable to avoid batches, DO split replicates of the different sample groups across batches
  - The more replicates the better (definitely more than 2).



Adapted from Hicks SC, et al. bioRxiv. 2015. doi:10.1101/025528.

Experimental planning considerations: Batches

Best Practice

- If unable to avoid batches, DO include batch information in your experimental metadata
  - During the analysis, we can regress out the variation due to batch so it doesn't affect our results if we have that information.

sample	replicate	condition	batch
sample1	1	control	1
sample2	2	control	1
sample3	3	control	2
sample4	4	control	2
sample5	1	treatment1	1
sample6	2	treatment1	1
sample7	3	treatment1	2
sample8	4	treatment1	2
sample9	1	treatment2	1
sample10	2	treatment2	1
sample11	3	treatment2	2
sample12	4	treatment2	2