

Group Metadata Exercise

Exercise 1

- Arrange yourselves in groups of 3 or 4
- Imagine you are submitting a zebrafish RNA-seq experiment to a repository
- What metadata should you submit along with the sequencing data?

Exercise 1

- Arrange yourselves in groups of 3 or 4
- Imagine you are submitting a zebrafish RNA-seq experiment to a repository
- What metadata should you submit along with the sequencing data?
- **What did you have on your lists?**

Course dataset (1/2)

- Throughout the course, will use dataset from:
Mech *et al.* (2022) “Behavioral and Gene Regulatory Responses to Developmental Drug Exposures in Zebrafish.” *Front. Psychiatry*
<https://www.frontiersin.org/articles/10.3389/fpsy.2021.795175>
- Exposed zebrafish to amphetamine, nicotine or oxycodone from 24 hpf to 5 dpf then performed behavioural assays on larvae
- At 5 dpf, 6 samples, each consisting of pools of 6-7 embryos, were collected for each condition (plus unexposed controls)

Course dataset (2/2)

- RNA extracted and sequencing libraries made (Illumina TruSeq Stranded mRNA kit)
- Sequenced on one lane of NovaSeq SP PE50 (16-24M reads / sample)
- Aligned to GRCz11 with STAR (twice)
- Differentially expressed genes determined with DESeq2
- (Files needed to repeat the analysis - download FASTQ, run STAR & DESeq2 - are available from <https://funcgen2022.buschlab.org/downloads/dataset.zip>)

Exercise 2

- Get back into your groups
- Which bits of metadata should be submitted for this particular experiment? Which shouldn't?
- Is there any metadata that you missed?

Exercise 3

- Look for zebrafish RNA-seq experiments on ArrayExpress (<https://www.ebi.ac.uk/biostudies/arrayexpress>)
- Try to find examples of experiments that have good metadata and some with poor metadata
- Make a note of the accession of each experiment

Thank You

Any Questions?