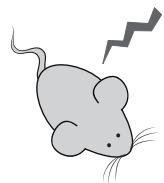# Bioinformatic workflow for Harris *et. al* 2020
Transcriptome analysis of hippocampal subfields identifies gene expression profiles associated with long-term active place avoidance memory

R + GitHub

CGGACCAGCGGGCGGGCTCCAGCGAT
GGAGCTCGACGAC
CGACCGACGCGAC

Texas Advanced Computing Center

UNIXworkflow/00_rawdata.md

UNIXworkflow/01_fastqc.md
+ MultiQC

UNIXworkflow/02_filtrimreads.md

UNIXworkflow/03_fastqc.md
+ MultiQC

data/multiqc/
multiqc_report0204.csv

UNIXworkflow/04_kallisto.md

data/GSE100225_IntegrativeWT2015
 <sample>/
  abundance.h5
  abundance.tsv
  run_info.json

Gene Expression Omnibus
accession: GSE100225

scripts/00_rnaseq_wrangle.Rmd

data/00_Data2013_2016_forAnalysis.csv

scripts/
00_behavior_wrangle.Rmd

data/00_behaviordata.csv

scripts/
1_behavior_analysis.Rmd

figures/fig-1.png
figures/fig-2.png
data/table-1.csv
data/suppltable-1.csv
data/suppltable-2.csv
data/suppltable-3.csv

data/00_countData.csv
data/00_colData.csv

scripts/2_rnaseqQC.Rmd

figures/supplfig1.png

scripts/3_rnaseqSubfield.Rmd

data/suppltable-1.csv
figures/figure-3.png
data/suppltable-4.csv
data/03_[DG,CA3,CA1].vsdtraining.csv

scripts/
4_correlations.Rmd

scripts/
6_candidates.Rmd

scripts/
5_GO.Rmd

figures/fig-4.png
figures/supplfig-3.png

figures/fig-5.png
figures/fig-6.png

data/table-2.csv
data/table-3.csv

Visit https://github.com/raynamharris/IntegrativeProjectWT2015 and click "launch binder" to explore the R workflow in the cloud with an myBinder. R scripts beginning with "00" are all about data wrangling. Focus on scripts beginning with 1-5 to reproduce the manuscript. The UNIX workflow is not executable; rather, it is a narrative containing the UNIX commands that were executed. Scripts and outputs are written in navy blue and burnt orange.