

BISC 4171 - NeuroGene Explorer: Student Project Packet

Spring 2026 (asynchronous / independent research)

Instructor: Dr. Damien O'Halloran | **Email:** damienoh@gwu.edu | **Office:** SEH 6530 | **Office hours:** by appointment

Start here

- **Download the course repository as a ZIP** (no Git required): repository name **dohalloran/BISC-4171-NeuroGene-Explorer**. The link is posted on Blackboard.
- Unzip the folder and open Terminal (Mac) or Anaconda Prompt (Windows) in that folder.
- Follow the repo **README.md** for step-by-step setup (including a quick-commands section).

Project overview

You will complete a guided bioinformatics research project focused on brain gene expression. You will build a small, reproducible Python pipeline that loads a public gene-count table, performs basic quality checks, ranks genes by differences between groups, and produces publication-style figures.

Research questions (required)

- **Aim 1 (sleep deprivation in WT):** Which genes change in wildtype (WT) prefrontal cortex after 5 hours of sleep deprivation (SD5) compared to homecage controls (HC5)?
- **Aim 2 (genotype comparison required):** Which genes differ between Shank3 mutant and WT under matched conditions (minimum: HC5)? (Optional: repeat within SD5 if time allows.)

Deliverables (graded)

- **Midterm Progress Report** (PDF, 2 pages max) - Aim 1: QC + preliminary ranked gene list (WT SD5 vs WT HC5) + short interpretation of 3-5 genes + one limitation.
- **Final Report** (PDF, 3-4 pages max) - Aims 1 + 2: finalized figures + ranked tables + interpretation + two limitations + one follow-up experiment.
- **Lightning talk** (one-slide PDF) + optional 2-3 minute recording: one key message, one main figure, one proposed follow-up.
- **Code + reproducibility package** (zip folder): include **src/** scripts and **results/** outputs + a short **README** (or text file) listing the exact run commands you used.

See the next page for a suggested timeline and workflow.

Recommended timeline (no weekly submissions)

Weeks	Milestone	What to have done
1-2	Setup	Download repo ZIP; create conda env; create data/ and results/; run download script.
3-5	Aim 1 draft	Load counts; confirm metadata; QC plot; preliminary volcano; top genes list.
6-7	Midterm	Clean Aim 1 figures + short interpretation; submit midterm progress report.
8-10	Aim 2	WT vs Shank3 within HC5; finalize tables/figures.
11-13	Final package	Integrate Aims 1 + 2; finalize reproducibility zip; lightning talk slide; submit final report.

Data and analysis workflow

You will download a processed counts table (no raw FASTQ processing required). The pipeline scripts are provided in the repository under **src/**.

Step-by-step data download (recommended)

- Activate your conda environment (**bisc4171-neurogene**).
- Run the download script to fetch the processed counts into **data/**:

Command: `python src/00_download_gse113754.py --outdir data --series-matrix`

Pipeline scripts (what they do)

- **00_download_gse113754.py** - downloads processed counts into data/
- **01_qc_and_metadata.py** - builds metadata + QC plot; writes results/metadata.csv
- **02_rank_genes_aim1_wt_sleepdep.py** - Aim 1: WT SD5 vs WT HC5 (volcano + ranked table)
- **03_rank_genes_aim2_genotype.py** - Aim 2: Shank3 vs WT within a matched condition (minimum HC5)

AI tools policy (debugging allowed)

You may use AI tools to help debug code (e.g., explain error messages, suggest fixes, refactor for clarity). There is **no AI log requirement**. However, you must be able to explain your final code and results in your own words. If you use AI assistance, include a brief disclosure in your report (1-2 sentences describing what you used AI for).

Midterm Progress Report template (2 pages max)

1. Background (150-200 words)
2. Data + Methods (150-200 words)
3. Results (1-2 figures + top-10 gene table)
4. Interpretation (3-5 genes + one limitation)
5. What is next (3 bullets)

Final Report template (3-4 pages max)

1. Background + hypotheses (200-300 words)
2. Data sources (GEO accession + files) and methods (200-300 words)
3. Aim 1 results (figures + top genes)
4. Aim 2 results (figures + top genes)
5. Interpretation (connect to brain biology; 2 limitations; 1 follow-up experiment)
6. References (2-4)

Note: The statistical approach here is simplified for learning; it is not a full DESeq2-style differential expression pipeline.