

Group Project and Presentation

As we've discussed in class, the recent deluge of genomic data has required many biologists to master computational skills that weren't previously essential to the discipline. To many biologists, version control and documented computational pipelines are somewhat new concepts and there is still some heterogeneity in the extent to which these practices are put in place.

For an example of a well documented study, see the following GitHub repository:

<https://github.com/HuffordLab/NAM-genomes>

This repository accompanied a [paper](#) published in the journal *Science*.

Your Mission: to provide documentation for an undocumented study

Groups of 3 to 5 students will identify a paper that utilizes computational skills in any area of biology but lacks sufficient documentation to easily replicate the analysis pipeline. Collectively, each group will identify analyses from this manuscript to reproduce and document using the computational skills and documentation standards you have learned in class.

This should entail:

1. Downloading, inspecting, and describing the data utilized in the study.
2. Processing the data if necessary to format them for the analysis the group has chosen to reproduce.
3. Rerunning the analysis described in the manuscript using your personal computers or ISU HPC resources.
4. Providing visual summaries (*e.g.*, ggplot figures) of your results.

The graded deliverables that each group will provide will include a **GitHub Repository** and an **In-Class Presentation**.

GitHub Repository:

This documentation should be clear and easy to follow and should adhere to the guidelines for the [ReScience Journal](#). Specifically, please include:

1. A top README.md file that describes the contents of your directory.
2. An author(s)-YEAR.md file that introduces the original paper, explains the technical details of your replication of analyses and summarizes your replication of the original results.
3. A code directory that contains the commented code for the replication.
4. A data directory that contains links to data necessary to run your code.

Should you decide, in consultation with your major professors, to publish your efforts in the ReScience Journal, you will be well on your way to submission if you adhere to these guidelines.

When creating your repository, you can choose one group member to be the owner of the project, in which case, the other group members can be given push rights or fork the project and make changes as pull requests. Alternatively, you can start a new "organization" for your group and create your repository as a group.

Repository Name: Please include "BCB546-Spring2024" in the name of your repository.

In-Class Presentation:

Each group will have ~20 minutes to present their work on either May 1st or May 3rd. Each presentation should include:

1. Background on the biological question being investigated.
2. A description of the workflow carried out by the group.
3. An overview of the group's documentation.
4. Presentation of results including comparison to results from the published paper.

Please feel free to message the instructors on Slack if you are unclear about the requirements of the assignment or regarding the suitability of a particular manuscript for the assignment.

Groups

Your group should be a collaboration. You can form your groups of 3-5 people based on any criteria you choose. We recommend that you find other students with similar research interests or working with similar kinds of data. We have started a channel in Slack for the projects called `#final-project`. On this channel you can let your classmates know what kinds of data you work with and what you're interested in. This may be helpful for forming your group.

To communicate with your group members, we encourage you to use Slack. You can create a new channel for just your group members or you can just chat via direct message.

The deadline for forming groups is **February 28th, 2024**. When you have formed your group,

post your group name and list of members on the Slack `#final-project` channel.

Submission

You must submit the URL to your group's repository on Canvas (to the Final Project assignment that will be created). The grade given to the group presentation will be based on the work as a whole. However, every member of a group must submit the URL on Canvas to ensure that you get credit for your participation in the project.

The deadline for submitting your repository on Canvas is May 3rd, 2024.