

# 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 new concepts and there is still great 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/timbeissinger/Maize-Teo-Scripts>

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

## Your Mission: to provide documentation for an undocumented study

Groups of 3 or 4 students will identify a recent 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 a subset of 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

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

## **GitHub Repository:**

This can be organized however each group sees fit, but documentation of analysis should be clear and easy to follow. Please include:

1. A README file in Markdown format that describes the workflow throughout the project and lists initials by tasks undertaken by each group member.
2. Files or a folder with any scripts written to process data, run an analysis, and produce figures.
3. pdfs or jpegs generated to help visualize results.
4. Slides for the group's in-class presentation.

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 "BCB546X-Spring2017" in the name of your repository.

## **In-Class Presentation:**

Each group will have ~20 minutes to present their work on either April 25th or 27th. 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 3-4 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 will start 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 **March 3, 2017**. When you have formed your group, post your group name and list of members on the Slack `#final-project` channel.