Computational work

Principles

- Computational research and work is central to the goals of our lab.
- We are committed to reproducible computational research.
- All code in the lab should be written in R and python and each lab member should be comfortable using at least one of these.
- The use of spreadsheets for data analysis and plotting in the lab is forbidden.

Reproducible research

- The results of all computational analysis must be associated with the code that generated it through the use of R markdown or iPython Notebooks. Examples of reproducible research are here and here.
- David will not look at plots that are not associated with the code that generated them.
- Every lab member should maintain their own code on a personal github page.
- All code for publications will be posted on the labs github page
- For routine bioinformatic analyses standard pipelines, which are maintained on github, should be used.

Data Backup

- All University members have unlimited cloud storage space on NYU Google Drive. With the Google Drive App users can selectively mount shared folders on lab computers or smartphones. By using the selective folder sync preference in the Google Drive App you can take full advantage of your unlimited storage without overfilling your hard drive.
- All data and files associated with research in the lab should be maintained in Google Drive.
- Data and files should be shared with David via a shared google drive folder named “GreshamLab_Member”
- External hard drives should not be used for backup

High Performance Computing

- Each lab member must know how to log on, transfer data and submit jobs using slurm to NYU HPC.
- Each lab member is responsible for maintaining an HPC account that has the following structure:
  - /scratch/userid/
  - /home/userid/
  - /archive/userid/
  - each lab member needs to understand the quotas and stability of these directories as well as the best practices for HPC computing in the Biology department as outlined here.
- The lab maintains a shared directory on scratch for long term storage and sharing of files:
  - /scratch/cgsb/gresham
    - each lab member may create a subdirectory within this directory (for example, David's directory is /scratch/cgsb/gresham/David)
    - all lab members have read and write access to all files in /scratch/cgsb/gresham
    - talk to David to request access to /scratch/cgsb/gresham
    - common files and resources are maintained in /scratch/cgsb/gresham/LABSHARE
      - /scratch/cgsb/gresham/LABSHARE/Data
      - /scratch/cgsb/gresham/LABSHARE/Resources
      - /scratch/cgsb/gresham/LABSHARE/Scripts
      - /scratch/cgsb/gresham/LABSHARE/Pipelines
    - jobs should never be launched within /scratch/cgsb/gresham/LABSHARE
    - note that this directory is NOT subject to the flushing rules that operate within /scratch/

Bioinformatic Pipelines

- Standard bioinformatic pipelines are maintained and available in /scratch/cgsb/gresham/LABSHARE/Pipelines and on the GreshamLab github page
- Each time a pipeline is used you should
  - clone the entire repo from github following these instructions: https://help.github.com/articles/cloning-a-repository/
  - modify the code so it is tailored for your analysis
• run the pipeline

Github Usage

• We aim to keep all code in the lab up to date and uniform to enable reproducible research. The lab github account is here.
• The best practices are:
  • maintain all code in repos using git on your local computer or the HPC
  • for every repo maintain a corresponding github repo
  • every time you use code, perform a pull from github to use the most current version of the code
  • commit frequently
  • comment on your commits verbosely by typing >git commit and pressing enter, this will open a text editor that allows a detailed description of the commit (do not use git commit --m ‘Made some bug fixes’)
  • every day repeat: git pull -> git add -> git commit -> git push
  • make changes using a fork and a pull request.
  • use gitk to open a GUI that displays the detailed history of the repo

Support

• Departmental computing support is extensive and is detailed in this document.
• All personal computers that are used for research or teaching are eligible for support by 1) registering with the department using this form and 2) submitting a request for support from the department using the ticketing system via this link.