UCDAVIS Bioinformatics Core

Closing Thoughts

Bioinformatic Core Workshop

7 Stages to Data Science

1. Define the question of interest

2. Get the data
3. Clean the data
4. Explore the data

5. Fit statistical models

6. Communicate the results
7. Make your analysis reproducible

Prerequisites

- Access to a multi-core (24 cpu or greater), ‘high’ memory 64Gb or greater Linux server.
- Familiarity with the ’command line’ and at least one programming language.
- Basic knowledge of how to install software
- Basic knowledge of R (or equivalent) and statistical programming
- Basic knowledge of Statistics and model building

In Bioinformatics

- Know and Understand the experiment
  - "The Question of Interest"
- Build a set of assumptions/expectations
  - Mix of technical and biological
  - Spend your time testing your assumptions/expectations
  - Don't spend your time finding the "best" software
- Don't under-estimate the time Bioinformatics may take
- Be prepared to accept 'failed' experiments

Bottom Line

The Bottom Line:
Spend the time (and money) planning and producing **good quality, accurate and sufficient data** for your experiment.

Get to know to your data, develop and test expectations

Result, you'll **spend much less time** (and less money) extracting biological significance and results during analysis.

Computing cluster (Reminder)

- Course will be conducted on our servers and compute cluster

  tadpole.genomecenter.ucdavis.edu

- Everyone should get an account.
  - https://computing.genomecenter.ucdavis.edu
  - Request an account -> sponsor Bioinformatics Core Workshop
  - If you already have an account on our systems, then please tell us your login
- Cluster usage will be under the slurm reservation ‘workshop’
  - Reservation will last 1 full week after the workshop and allow you to practice or run analyze your own data.

workshop ACTIVE 2018-08-26T00:00:00 2018-09-09T00:00:00 14-00:00:00

Contacts (Reminder)

- **Bioinformatics related questions, include but not limited to**

  bioinformatic methods questions, software use, data questions

  Bioinformatics.core@ucdavis.edu
- Computing Issues, include but not limited to

  User account questions, equipment failure/malfunction, software install, software failures (not related to use)

  helpdesk@genomecenter.ucdavis.edu
- Training courses information

  training.bioinformatics@ucdavis.edu