

Closing Thoughts

Dr. Matthew L. Settles

Genome Center
University of California, Davis

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

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.

Workshop week reservation

- workshop ACTIVE May 21th, 2018 to May 30 at midnight

My recommendation is to follow all of the instructions again, from the beginning on your own and send emails to

training.bioinformatics@ucdavis.edu

And we will be responsive to answering questions

Future workshops

Bioinformatics: RNA-Seq Workshop 2018

June 18, 2018, 9:30 a.m. - June 22, 2018, 4:30 p.m.

Organizer - Bioinformatics Core

Contact - UC Davis Bioinformatics Core, training.bioinformatics@ucdavis.edu

When	Where	What
Aug 27-31, 2018	UCD	Variant Analysis
Sept 4-5, 2018	UCD	Command Line Workshop
Sept 6-7, 2018	UCD	R Workshop
December, 2018	UCD	Genome Assembly (euk)

<http://registration.genomecenter.ucdavis.edu/>