UCDAVIS Bioinformatics Core

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.

UCDAVIS Bioinformatics Core

Workshop week 2 reservation

- workshop ACTIVE Friday September 15th, 2017
- Follow up on Friday September 15th in the GBSF Auditorium

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

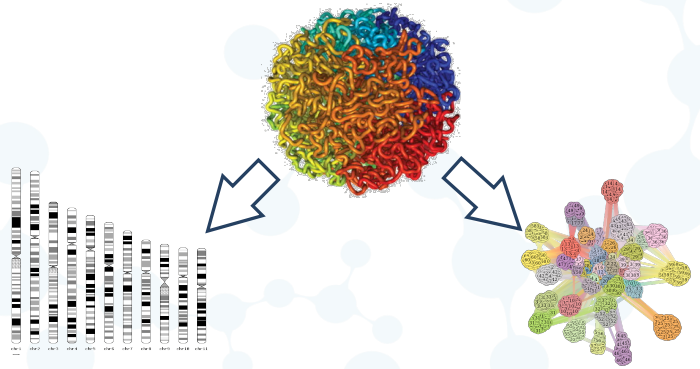
UCDAVIS Bioinformatics Core

Phase Genomics Talk
September 19th
Tuesday, 2:30PM

How Hi-C is transforming genome and metagenome assembly

Chromosome conformation capture methods like Hi-C measure the 3D organization of DNA *in vivo* using a combination of crosslinking, proximity-ligation, and paired-end sequencing.

Because this method captures genomic contiguity on intact chromosomes, the resultant information can be used to generate end-to-end chromosome-scale scaffolds for large genomes. Since Hi-C junctions form within intact cells, any sequences interacting by Hi-C must have originated from the same species/strain in a mixed population, enabling metagenomic deconvolution.



Capturing genomic proximity information *in vivo* removes several major obstacles in genome and metagenome assembly, improving the quality and efficiency of genome discovery efforts.

September 19th 2017
Tuesday, at 2:30 PM
GBSF Auditorium

Ivan Liachko, Ph.D.
CEO, Phase Genomics, Inc.

UCDAVIS Bioinformatics Core

KBASE Workshop
Sept 20th
10am to 3pm
In Coordination with the Library

A collaborative, open environment for systems biology of plants, microbes and their communities