BST281: Genomic Data Manipulation, Spring 2018

Wednesday 01: Bioinformatics and computational experiments

Quantitative biology: umbrella term for biological research that requires data nuance.

Computational biology and bioinformatics often used interchangeably.

Sometimes the former refers more to biological applications and the latter to algorithms / biophysics.

Also includes biostatistics and general quantitative thinking.

Computational experiments should be handled much like laboratory experiments.

Express hypotheses, design first, and be constructively skeptical.

Be constructively skeptical: if a result looks too good to be true, it probably is.

Modularize computational processes.

Look at the data going in and coming out of every step.

Include positive and negative control inputs with known and null outputs, respectively.

Document everything: electronic lab notebook, data, and code.

Can be assisted by literate programming tools, simple readme files, and comments.

Keep your workspace clean: use consistent file layouts and naming conventions.

# Python

Python (or any program) consists of instructions and data.

Instructions include operators, keywords, and functions.

Keywords are special built-in commands; operators are specific instructions abbreviated using punctuation.

Functions are complex, reusable instructions: recipes that can be executed (called) or stored (written).

Called using **function\_name( arguments )** or **argument1.function\_name( arguments )**

Defined using the **def** keyword - like writing a recipe, doesn't make the food yet...

Data are stored in variables, which are named buckets (not the data themselves).

**strString = "sequence of characters"**

**iInteger = 123**

**dFloat = 1.23**

**fBoolean = True**

# Textbooks

Python data: Haddock and Dunn, Chapter 7 p105-120 (stop before Input And Output),  
Chapter 8 p125-135 (stop before Controlling String Formatting), Chapter 9 p141-172,  
Chapter 10 p188-192 (start at Creating Custom Python Functions, stop before Packaging Data)

# Literature

[A quick guide to organizing computational biology projects. Noble, PLoS Comp. Bio. 2009.](https://www.ncbi.nlm.nih.gov/pubmed/19649301)

[Genomic signatures to guide the use of chemotherapeutics. Potti et al, Nature Medicine 2006.](https://www.ncbi.nlm.nih.gov/pubmed/17057710)

[Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology. Baggerly and Coombes, Annals of Applied Statistics 2010.](https://arxiv.org/abs/1010.1092)