

BIOINF 524/525 Winter 2018
Foundations of Bioinformatics and Systems Biology

Course Director: Dr. Peter Freddolino (petefred@umich.edu)

Description: BIOINF524/525 is a three credit three module course designed to provide biological researchers with the computational and statistical tools that they need to efficiently make use of their data, and to integrate publicly available “omics”/high-throughput data sets into their work. The course is intended for students with primarily experimental backgrounds and research interests, and will provide a foundation in computational and quantitative methods suitable for analysis of both low- and high-throughput data sets.

BIOINF524 itself is comprised of modules covering: (1) Introductory programming and exploratory data analysis with R, (2) Introductory statistics for analyzing biomedical data, and (3) Systems biology and analysis of high-throughput datasets. BIOINF525 allows for each one credit module to be registered for separately. Students wishing to take only the later modules should consult with the instructor before registering in order to ensure that they have appropriate background knowledge, as the material is somewhat cumulative.

The course itself consists of weekly lecture sessions and lab sessions, with the latter conducted primarily using the R programming language. Weekly homework assignments will reinforce concepts covered in lecture/lab and provide an opportunity to practice applying the methods that have been learned.

Schedule:

Lectures: Tuesdays 2:30-4:00 PM, Room 2062 Palmer Commons
Labs: Thursdays 2:30-4:00 PM, Room 2062 Palmer Commons
(Note: During Module 1, all sessions will be combined Lecture/Lab)

Module 1: January 9 -- February 2 (8 combined lectures/labs)
Module 2: February 6 -- March 15 (5 lectures, 5 labs)
Module 3: March 20 -- April 12 (4 lectures, 4 labs)

Prerequisites: A familiarity with basic biomedical concepts and basic knowledge of computer usage. No programming skills are needed. You **MUST** bring your own wifi enabled laptop to lectures and labs to fully participate in this class.

Grading: Letter grades will be assigned based on a combination of lecture attendance (20%), lab exercises (40%), and homework performance (40%).

COURSE SCHEDULE

Lecture 0: Introduction and course procedures

Time: Jan. 4 (Thursday), 2:30 -- 4:00 PM

Topics: Introduction to bioinformatics, quantitative biology, and systems biology. Intersection points with other branches of biology, and how to get the most out of this course. Course policies and procedures. Students who are not taking Module 1, but plan to take later modules, should attend if possible, but attendance at this lecture will not be graded. (Delivered by P. Freddolino)

Module 1 (1 credit): Introductory programming and exploratory data analysis with R

Module 1 Instructor: Dr. Stephen Guest (stguest@umich.edu)

Introduction to the R programming language for students with no previous programming experience. Loading datasets from various sources into the R programming environment. Manipulating and restructuring datasets with R. Exploring data and generating informative, interactive data visualizations. Creating and implementing customized data analysis pipelines. Documenting your data analysis and making it reproducible. Accessing the Bioconductor repository of bioinformatics packages.

Lecture/Lab (1-1): R programming basics and the RStudio environment

Time: Jan. 9 (Tuesday), 2:30 -- 4:00 PM

Topics: Why learn R? Working in the RStudio integrated development environment. Getting data into RStudio. Slicing/dicing data and calling functions. Working with different data types.

Lecture/Lab (1-2): Enhancing R functionality with Packages

Time: Jan. 11 (Thursday), 2:30 -- 4:00

Topics: Finding and installing packages. Using the dplyr package to perform common data manipulation tasks. Restructuring data with the reshape2 package.

Lecture/Lab (1-3): Visualizing Data Part 1

Time: Jan. 16 (Tuesday), 2:30 -- 4:00

Topics: Using base R plotting capabilities to visualize data. Using specialized packages for visualizing data.

Lecture/Lab (1-4): Visualizing Data Part 2

Time: Jan. 18 (Thursday), 2:30 -- 4:00

Topics: The grammar of graphics and visualizing data using the ggplot2 package. Making your plots interactive with plotly.

Lecture/Lab (1-5): Writing functions and assembling your own data analysis pipelines part 1

Time: Jan. 23 (Tuesday), 2:30 -- 4:00

Topics: The syntax for writing functions. Sourcing functions and integrating functions into your analysis pipeline. Rules for variable scoping and functions in R.

Lecture/Lab (1-6): Writing functions and assembling your own data analysis pipelines part 2

Time: Jan. 25 (Thursday), 2:30 -- 4:00

Topics: Control structures -- If/else, for loops, while loops, repeat, next and break. Using pipes to write code more efficiently and to improve the readability of your code.

Lecture/Lab (1-7): A cautionary tale of irreproducible research and endangering cancer patient's lives

Time: Jan. 30 (Tuesday), 2:30 -- 4:00

Topics: Poorly documented, irreproducible research and the Duke microarray cancer clinical trial scandal. Flexible and fast analysis documentation with the knitr package. Using R markdown and R notebooks to make your analysis reproducible.

Lecture/Lab (1-8): An introduction to Bioconductor

Time: Feb. 1 (Thursday), 2:30 -- 4:00

Topics: Bioconductor as a repository for R packages that are commonly used in bioinformatics data analysis. Using the biomaRt package to annotate genes. Using biomaRt to retrieve data from multiple biological databases e.g. Ensembl, COSMIC, and dbSNP.

Module 2 (1 credit): Introductory Statistics for Analyzing Biomedical Data

Module 2 Instructor: Dr. Armand Bankhead (bankhead@umich.edu)

Statistical concepts represent an important cornerstone to answering scientific biomedical questions in an objective and reproducible manner. This module introduces students to commonly used statistical methods used to summarize and make inferences from biomedical data. Lectures, labs, and homework will illustrate application of course theory using in vitro experimental data, in vivo experimental data, and patient biomedical data.

Lecture (2-1): Fundamental Statistical Concepts

Time: Feb. 6 (Tuesday), 2:30 -- 4:00

Topics: Basic statistics for summarizing data distributions. Statistical error measurements and normality testing.

Lab (2-1): Summarizing Biological Measurements Using R

Time: Feb. 8 (Thursday), 2:30 -- 4:00

Topics: Application of statistical concepts discussed in lecture to two biomedical datasets including summary statistics calculations and visualization of data distributions.

Lecture (2-2): Testing Statistical Differences Between Measurements

Time: Feb 13 (Tuesday), 2:30 -- 4:00

Topics: Framing scientific questions using hypothesis testing. Common statistical tests,

p-values, multiple testing correction.

Lab (2-2): Answering Biological Questions with Hypothesis Testing Using R

Time: Feb 15 (Thursday), 2:30 – 4:00

Topic: Applying statistical tests to identify significant differences in experimental measurements.

Lecture (2-3): Quantifying Associations Between Measurements Using Contingency Tables

Time: Feb 20 (Tuesday), 2:30 – 4:00

Topics: How and why data sets should be summarized as contingency tables. Statistical tests used to evaluate associations between measurements.

Lab (2-3): Comparing Gene and Phenotype Measurements Using Contingency Tables in R

Time: Feb 22 (Thursday), 2:30 – 4:00

Topic: Applying statistical tests to identify significant associations between phenotypes and experimental measurements.

Lecture (2-4): Using Linear Models and Curve Fitting to Compare Measurements

Time: Feb 27 (Tuesday), 2:30 – 4:00

Topics: Modeling measurement associations using correlation and linear models, modeling relationships between experimental measurements using curve fitting

Lab (2-4): Modeling Gene Expression and Dose Response Relationships Using R

Time: Mar 1 (Thursday), 2:30 – 4:00

Topics: Fitting dose response curves to cell line drug response data. Quantifying gene co-expression relationships using correlation.

Lecture (2-5): Applying Bayesian Modeling Approaches to Biological Data

Time: Mar 6 (Tuesday), 2:30 – 4:00

Topics: Introduction to probability, conditional probability, and Bayes' formula. Simulating data distributions using Gibbs sampling.

Lab (2-5): Asking Probabilistic Biological Questions with a Bayesian Mindset Using R

Time: Mar 8 (Thursday), 2:30 – 4:00

Topic: Using the LearnBayes package to create Bayesian models to analyze experimental data and contrasting with results from analogous frequentist methods.

Module 3 (1 credit): Biological networks and systems-level modeling

Module 3 Instructor: Dr. Peter Freddolino (petefred@umich.edu)

Computational analysis of biological networks, OMICs data (genomics, transcriptomics, metabolomics, proteomics). Application of advanced analysis and modeling approaches to study pathways and networks. Emphasis on using existing high throughput data sets alongside newly generated experimental data to analyze and interpret research findings.

Lecture (3-1): Introduction to systems biology

Time: Mar. 20 (Tuesday), 2:30 -- 4:00 PM

Topics: Definitions of systems biology. Common network motifs and how they contribute to biological behavior. Connection of omics data to systems biology.

Lab (3-1): Network analysis for systems biology

Time: Mar. 22 (Thursday), 2:30 -- 4:00 PM

Topics:

Representation of data as graphs. Pathway and network exploration and visualization with Cytoscape (including the MetScape app and the RCytoscape package) and analysis with igraph.

Lecture (3-2): High throughput sequencing methods in systems biology

Time: Mar. 27 (Tuesday), 2:30 -- 4:00 PM

Topics:

Overview of high throughput sequencing-based methods used to investigate biological networks, along with an introduction to databases and analysis considerations for each.

Lab (3-2): Mapping genetic regulatory networks using high-throughput sequencing

Time: Mar. 29 (Thursday), 2:30 -- 4:00 PM

Topics:

Finding and interpreting RNA-seq and ChIP-seq data sets to study regulatory networks, including identification of differentially expressed genes, location of transcription factor binding sites, and inference of regulatory motifs.

Lecture (3-3): Network modeling for hypothesis testing and generation

Time: Apr. 3 (Tuesday), 2:30 -- 4:00 PM

Topics:

Conceptual understanding of kinetic models and constraint-based models of metabolic networks; approaches for translating biochemical schematics into quantitative frameworks.

Lab (3-3): Modeling and inference of metabolic networks

Time: Apr. 5 (Thursday), 2:30 -- 4:00 PM

Topics:

Students will learn how to represent simple metabolic pathways in SBML, and use the sybil R package to simulate the effects of mutations in a model network.

Lecture (3-4): Machine learning approaches in systems biology

Time: Apr. 10 (Tuesday), 2:30 -- 4:00 PM

Topics:

Introduction to machine learning and overview of key algorithms (linear classifier, SVM, decision trees and random forests). Existing and potential areas for application of machine learning in the

analysis of biological networks.

Lab (3-4): Application of machine learning to biological network analysis

Time: Apr. 12 (Thursday), 2:30 -- 4:00 PM

Topics:

Overview of R modules for machine learning; comparison of a variety of approaches for predicting regulatory modules, protein function, and protein-protein interactions.