

Course outline

Biochemistry 9545 Macromolecular Informatics. The course will typically meet once per week Wednesdays at 9:30-11:30. There will be an interruption in the course, and additional sessions will be scheduled to account for this interruption.

NOTE: This is intended to be focused largely on theoretical aspects of analysis, and practical information on how to manage and keep track of your work

Instructor:

Greg Gloor, ggloor@uwo.ca, 519-661-3526 (office)

Requirements

Students will be expected to bring a functional laptop to class with the R programming environment already loaded and functional. Students should be familiar with the file system of their computer and be able to locate files associated with R on their own. Students should have a working text editor installed. For Macs this could include Textwrangler, for PCs the could include Notepad++. Students should have a copy of “*Analyzing compositional data with R*” loaded on their computer for reference. We will use many of the principles outlined in that book. In addition, I encourage students to get a copy of “*R in Action*” — you will find this invaluable.

Sessions

1. An introduction to high throughput sequencing and data types generated in high throughput sequencing. An introduction to R, markdown and knitr (functional note taking)
2. A further introduction to R, data types, installing packages, simple plots, exploring multivariate data
3. An introduction to the structure of multivariate datasets. An introduction to compositional data.
4. An introduction to the error structure in high throughput sequencing. PCA and compositional biplots.
5. An introduction to Bayesian thinking in compositional data
6. Correlations in compositional data

Assessments

1. The student will be required to submit a proper R program. Marks will be assessed based on comments, functionality, and conciseness. This must be submitted in the form of a .Rmd document.
2. The student will be required to make and properly interpret a compositional biplot from a dataset provided.
3. The student will be required to examine and interpret the correlation and pairwise abundance of parts of a high-throughput sequencing dataset

The course is largely theoretical . Assessments are to be worked on and handed in individually, and are equally weighted. Ten percent of the marks will be based on attendance, participation and a demonstrated ability to use the tools properly. Students will be expected to conduct independent learning to practice the concepts and tools.

Requirements

Students will need to bring a reasonably recent laptop to class and be willing to install a number of freely available scientific software packages. Students are expected to bring a laptop with R preinstalled (<http://cran.utstat.utoronto.ca>). We will be learning theory and some practice on how to use R and its packages for the analysis of high throughput sequencing data.

I expect that students will learn outside of class, by doing the readings and doing at least some of their own troubleshooting. We will not be installing much, but what we do install will help you to keep track and present your work in the future.

Sources

This should be all you need, aside from some R packages

- R installation: <http://cran.utstat.utoronto.ca>
- Markdown: <http://daringfireball.net/projects/markdown/>
- R studio and markdown: <http://rmarkdown.rstudio.com>

Readings for week 2

- how does multiple testing correction work? Nat. Biotech. 2009 27:1135

- how to make more published research true. PLoS Medicine. 2014 11:e1001747
- its the effect size stupid. <http://www.leeds.ac.uk/educol/documents/00002182.htm>
- the ficke P value generates irreproducible results. Nat. Meth. 2015. 12:179

Due dates

All assignments are due one week after they are given. All three assignments must be handed in and passed to obtain course credit.

Statement on Academic Offences

The statement: “Scholastic offences are taken seriously and students are directed to read the appropriate policy, specifically, the definition of what constitutes a Scholastic Offence, at the following Web site: http://www.uwo.ca/univsec/handbook/appeals/scholastic_discipline_grad.pdf Academic Handbook, Exam, Course Outlines Page 4 Issued: 2011 02

Additionally, “All required papers may be subject to submission for textual similarity review to the commercial plagiarism-detection software under license to the University for the detection of plagiarism. All papers submitted for such checking will be included as source documents in the reference database for the purpose of detecting plagiarism of papers subsequently submitted to the system. Use of the service is subject to the licensing agreement, currently between The University of Western Ontario and Turnitin.com (<http://www.turnitin.com>).”