

Human Genetics and Genomics Computational Biology Course

Dennis Hazelett & Simon Coetzee

Spring 2020

Course Abstract

The goal of this course is to take students on an in-depth tour of a single common bioinformatics analysis pipeline. We will deconstruct the RNA-seq analysis pipeline starting with the easiest, highest level visualizations and working our way back through successively more advanced techniques. We will start at the endpoint of the analysis, with an existing, finalized dataset, going step-by-step through the pipeline until we reach the very beginning with alignment and quality control steps. We will emphasize practical utility of the methods, plus exposure to important ideas and concepts for students to explore on their own at a later date.

Course Format

The course takes place over 6 weeks with 1.5 hour lectures on Tuesdays and practical workshops on Thursdays for a total of 12 lectures and workshops. Students should gain an understanding of what is involved with this type of analysis and be able to use their notes to recapitulate it with a future dataset if needed.

Special considerations for COVID-19

Due to social distancing protocols the course will take place online via a password-locked Zoom session. Details to be distributed via email.

Expectations: Attendance, Homework & Grading

The course is worth 100 points. Homework is expected to be turned in on time. Late assignments will be given a maximum of 50% of full credit up to one lecture *after* the original due date. Recommended readings are indicated on the syllabus. During the course, students will choose one or more transcription factors from the CRISPRi dataset and follow the lecture series in completing the steps of an analysis pipeline from data acquisition through hypothesis testing and visualization. Each stage of analysis and results is completed in successive homework assignments (“Labs”) or question sets. Each question set is worth 10 points and will be graded based on completeness. Each student will also be required to present their homework at least once during the course during a workshop. For this informal presentation, the student will be graded on their ability to discuss the figures they generated and any challenges encountered (10 points). A final exam worth 30 points will be given at the 12 lecture.

Attendance is required no exceptions. You may obtain permission ahead of time or with extenuating circumstances after the fact from the graduate school (email to Emma Yates Kassler). Homework assignments are required on time (see schedule below) regardless of attendance or for half credit **one lecture late**. Unexcused absences result in an incomplete grade.

Lecture & assignment schedule:

Tuesday, May 5 *Introduction* **Lecture topics:** Course structure and expectations. Install GENAVI. Availability of course materials. **Homework Assignment I:** (Lab) install genavi locally.

Thursday, May 7 *DGE* **Lecture topics:** ENCODE datasource. TCGA data and TCGAbiolinksGUI. Preparing data for use in GENAVI. Differential gene expression (DGE) analysis using GENAVI. **Homework Assignment II:** (Lab) Perform DGE on a TF of your choice.

Tuesday, May 12 *Review* Workshop / Homework Review

Thursday, May 14 *Exploratory Analysis* **Lecture topics:** Unsupervised clustering. Principle Components Analysis (PCA). Correlation. **Assignment III:** (Lab) Exploratory analysis of your chosen TFs in the CRISPRi dataset.

Tuesday, May 19 *Review* Workshop / Homework Review

Thursday, May 21 *GO Analysis* **Lecture topics:** Gene Ontology (GO) & Pathway Analysis, David, GoRilla, GENAVI. Gene Set Enrichment Analysis (GSEA) and MSigDB. Installing software. **Assignment IV:** (Lab) GO analysis of your TF DGE set.

Tuesday, May 26 *Review* Workshop / Homework Review

Thursday, May 28 *NGS & File Formats (Coetzee)* **Lecture topics:** Log files from alignment and QC analyses. Structure and interpretation of NGS file formats (*e.g.* FASTQ, .sam *etc.*) **Assignment V:** (Question set)

Tuesday, Jun 2 *Review* Workshop / Homework Review

Thursday, Jun 4 *MultiQC (Coetzee)* **Lecture topics:** Performing Quality control with FASTQC and MultiQC. MegaQC for aggregating and databasing MultiQC reports. **Assignment VI:** (Question Set)

Tuesday, Jun 9 *Overview* **Lecture topics:** Question Set review. Course Overview / Review. Course Feedback and Evaluation.

Thursday, Jun 11 *Final Exam*

Oral presentations

During one of the 3 scheduled workshop sessions students are expected *on a volunteer basis* to discuss the results of a homework assignment and any challenges encountered (worth 10 points).

Schedule and Due Dates:

Table 1: **Course schedule and outline.** Numbered assignments are preceded by ‘L’ for labs, ‘QS’ for question sets and ‘E’ for exams. Optional times for oral presentation are during workshops.

day	date	lecture	hmwk	due
Tue	05/05	Intro	L1	
Thu	05/07	DGE	L2	L1
Tue	05/12	Wkshp		L2
Thu	05/14	Expl Anls	L3	
Tue	05/19	Review		L3
Thu	05/21	GO Anls	L4	
Tue	05/26	Wkshp		L4
Thu	05/28	Formats	QS5	
Tue	06/02	Wkshp		QS5
Thu	06/04	MultiQC	QS6	
Tue	06/09	Review		QS6
Thu	06/11	Exam		E1

Course Materials

All course materials, including this syllabus, lab workflows, question sets, and lecture slides are available online through the graduate school website or github.