SCRB 152. *Asking Cells Who They Are: Computational Transcriptomics Using RNA-Seq*

Douglas A. Melton (dmelton@harvard.edu), Adrian Veres (adrian_veres@hms.harvard.edu), and Eran Hodis (eran_hodis@hms.harvard.edu)

Course Development Assistant: Rapolas Zilionis (rapolas_zilionis@hms.harvard.edu)

Harvard College/GSAS: 156937
Fall 2016-2017

Location: Sherman Fairchild 268 (FAS)
Meeting Time: Tuesday, Thursday 11:30am - 1pm
Exam Group: FAS04_A

Course Description: This course is a hands-on introduction to computational analysis of RNA sequencing data as a measure of genome-wide expression. We will cover methods spanning the spectrum of RNA-Seq analysis: starting from raw sequencing reads, obtaining gene expression measures, and interpreting biological significance by differential expression analyses, clustering, and visualization. Coursework will consist of programming assignments in Python exploring real datasets. The course will emphasize skills applicable to independent biological research.

Prerequisites: Life Sciences 1a or Life and Physical Sciences A or permission of the instructor; Mathematics at the level of Math 19a or statistics at the level of Stat 102 or above. CS 50 or some previous exposure to programming is preferred but not required.

Overview

First half of course:
learning RNA-seq concepts and analyses.
problem set due (almost) every week.

Second half of course:
applying learned skills to analyze two published RNA-seq papers.
paper-based projects focused on recreating main figures, with some exploration

Problem sets

1. Weekly problem sets or paper-based problem sets requiring approximately 10 hours of work each.
2. Paper reading assignments (2 papers, in the second half of the course.)
3. Re-analyses of central figures from the two papers in <2>.

Midterm / Final exam

No midterm, no final exam. The course requires demanding problem sets. The students' ability to complete the problem sets, together with in-class presentation and discussion, will allow for the best assessment of whether they have learned the material.

Grading

35% Problem sets (divided equally among all 6)
25% Paper #1
25% Paper #2
15% In-class participation

Must complete all assignments (6 problem sets, 2 paper-based projects) to pass the course.

Problem sets will be reviewed and presented in class.

All problem sets and paper-based problem sets must be your own work.

Late problem sets

Each student will be allowed to hand in a single problem set 24 hours late over the course of the semester. Students are encouraged to save this 'late pass' for those days when they find themselves unexpectedly unable to complete the problem set by the deadline. Besides 'late pass', late problem sets will be penalized 25% each day they are late.

Course size and lottery

Course size will be limited, determined by lottery if necessary. Lottery results will be communicated by Monday September 7th. To enter the lottery, students will be required to fill out an online form by Friday September 4th, at 5pm.

Office hours

TBD

Overview of class calendar

	Tuesday	**Thursday**
Aug 30, Sep 1		**First day of class** Lottery Form Due Friday Sep 2, 5pm

Sep 6, 8	PS1 Lecture	PS1 Lecture
Sep 13, 15	**PS1 Review** PS1 Due Sep 14, 5pm	PS2 Lecture
Sep 20, 22	**PS2 Review** PS2 Due Sep 21, 5pm	PS3 Lecture
Sep 27, 29	PS3 Lecture	PS3 Lecture
Oct 4, 6	**PS3 Review** PS3 Due Oct 5, 5pm	PS4 Lecture
Oct 11, 13	PS4 Lecture	PS4 Lecture
Oct 18, 20	**PS4 Review** PS4 Due Oct 19, 5pm	PS5 Lecture
Oct 25, 27	**PS5 Review** PS5 Due Oct 26, 5pm	PS6 Lecture
Nov 1, 3, 5	**PS6 Review** PS6 Due Nov 2, 5pm	**Paper #1 Reading Due**
Nov 8, 10, 12	Special Lecture Topics #1	**Paper #1 Presentations 1**
Nov 15, 17, 19	**Paper #1 Presentations 2**	**Paper #2 Reading Due**
Nov 22, 24	Special Lecture Topics #2	**Thanksgiving (no class)**
Nov 29, Dec 1	Research Day	**Paper # 2 Presentations 1**
One session during reading period	**Paper # 2 Presentations 2 & Feedback**	

Detailed class calendar

Week 0	Thursday September 1, 2016 **Online lottery form due Friday Sept 2 by 5pm.**	Class orientation & goals of the course
		How are cells different from each other?
		RNA: What is it, how do we measure it?
		What is a gene/transcript? How is it turned on/off?

Week 1	Tuesday September 6, 2016	Sequencing by Synthesis *How commonly used short read sequencing works*
		Molecular Biology of Library Preparation *Going from RNA to short reads*
	Thursday September 8, 2016	What is a p-value?
		PDF, CDF
		Multiple hypothesis testing
		Introduction to linear models
Week 2	Tuesday September 13, 2016	**Problem Set #1 review** *Problem Set #1 Due on Monday September 12, 5pm*
		Visualization of aligned RNA-Seq reads
		Understanding the information that aligned short reads give us about expression
	Thursday September 15, 2016	Deriving first-pass measures for gene expression quantification
		What does it mean for a gene to be 'differentially expressed' in one condition versus another?
		RNA composition bias and how to normalize for it
Week 3	Tuesday September 20, 2016	**Problem Set #2 review** *Problem Set #2 Due on Monday September 19, 5pm*
		Expectation and variance, in the context of linear models
	Thursday September 22, 2016	Linear models continued
		Design and contrast parametrizations of linear models
		Formulating hypotheses in linear models
Week 4	Tuesday September 27, 2016	Estimating the variance of read count observations

	Thursday September 29, 2016	Assessing differential expression using linear models
Week 5	Tuesday October 4, 2016	**Problem Set #3 review** *Problem Set #3 Due on Monday October 3, 5pm*
		Covariance and correlation matrices
		Matrix diagonalization
	Thursday October 6, 2016	Projecting data to new dimensions
		Principal component analysis
Week 6	Tuesday October 11, 2016	How can we measure distance between samples or genes?
		Distance measures
	Thursday October 13, 2016	What is clustering ?
		Hierarchical Clustering
		K-means clustering
		Measures of within-group variance
Week 7	Tuesday October 18, 2016	**Problem Set #4 review** *Problem Set #4 Due on Monday October 17, 5pm*
		Smith-Waterman local alignment algorithm
	Thursday October 20, 2016	Seed-and-vote alignment algorithm

Week 8	Tuesday October 25, 2016	**Problem Set #5 review** *Problem Set #5 Due on Monday October 24, 5pm*
		What are gene sets? How are they constructed?
	Thursday October 27, 2016	Methods for ascertaining differential expression of gene sets
		Accounting for inter-gene correlation within sets
		Advantages and limitations of gene set testing
Week 9	Tuesday November 1, 2016	**Problem Set #6 review** *Problem Set #6 Due on Monday October 31, 5pm*
		Introduction to project-based part of course
	Thursday November 3, 2016	**Journal club of Paper #1** (led by students) *A rat RNA-Seq transcriptomic BodyMap across 11 organs and 4 developmental stages*
Week 10	Tuesday November 7, 2016	*Special Topics Lecture #1* Other ways of measuring gene expression (Micro-arrays, qPCR, FISH)
	Thursday November 10, 2016	**Paper #1 Part 1 presentations**
Week 11	Tuesday November 15, 2016	**Paper #1 Part 2 presentations**

	Thursday November 17, 2016	**Journal club of Paper #2** (led by students) *Population and single-cell genomics reveal the Aire dependency, relief from Polycomb silencing, and distribution of self-antigen expression in thymic epithelia*
Week 12	Tuesday November 22, 2016	*Special Topics Lecture #2* Isoform-level quantification using the EM-algorithm and related methods
	Thursday November 24, 2016 (No class)	*Thanksgiving Weekend*
Week 13	Tuesday November 29, 2016	*Research Day* Presentations by Faculty Members relating to research questions using RNA-Seq
	Thursday December 1, 2016	**Paper #2 Part 1 presentations**
Week 14 (reading period)	Session to be scheduled, (tentatively Tuesday December 6, 2016)	**Paper #2 Part 2 presentations**
		Student Feedback on course

Collaboration Policy

You may discuss problems and concepts, but ultimately the work that you submit must be your own. In order to get the most from this class, we strongly encourage you to keep trying to solve problems on your own, even when you are stuck. Likewise, when discussing problems together, students should assume the challenging responsibility of helping others without revealing answers directly, as this may rob their classmates of learning opportunity.