

Course Name: Special Topics in Insect Genomics

Topic: Differential Gene Expression Analysis by RNA Sequencing

Course Number: ENTM699N, 1 credit

Instructor: Megan Fritz, Ph.D., Entomology

Semester: Spring 2022

Course Summary: “Special Topics in Insect Genomics” introduces basic concepts and tools used in the analysis of second and third generation sequencing data for learners with little previous experience. Course topics will vary by semester (see above for spring 2022 topic), and students are welcome to make suggestions for a topic at least 1 semester in advance of the course start date. The goal of this course is to provide learners with (1) a basic understanding of the tools needed to analyze big genomic data, as well as (2) experiential learning as we operate these tools together during highly interactive class sessions. Students will discuss primary literature and review articles that illustrate key concepts in genomic data analyses. They will also follow along with instructor live-coding to learn the basic workflow and commands used to conduct genomic data analysis using open source software from the command shell.

Prerequisites: ENTM699M is a prerequisite for this course.

Course Outcomes: Upon completion of this course learners will be able to:

1. Describe how sequencing technology has changed over the past 50 years, and how these different technologies are used for the study of differential gene expression.
2. Predict how quality control steps (e.g. read filtering and trimming), or lack thereof, could impact the outcome of an experiment.
3. Produce a workflow diagram that takes a sequencing dataset from unfiltered raw reads to a dataset ready for downstream statistical analysis.
4. Present a basic interpretation of the results from a differential expression analysis when provided with 1) the biological context, and 2) normalized read count comparisons.

Course Schedule: Wednesdays from 2-4 pm for the second 7 weeks of Spring 2022. Please arrive 10 min early to set up your computer for class.

Readings: Web links and full references for relevant primary literature found below.

Grading: Student grades will be based primarily upon in-class participation (60%) and on the basis of two assignments given over the course of the semester (40%).

Participation: Students should attend and participate in all course meetings during the semester (8 pts. per period). Student participation grades will be determined based upon the quantity, dependability, and quality of participation. Appropriate forms of engagement include: 1) respectful and positive verbal discourse with the instructor or peers, or 2) insightful questions related to course materials. For more information on what constitutes acceptable in-class participation, see Petress 2006.

Assignment 1 - Paper Presentation: Individually or in groups of 2, students will present a paper for class discussion on one aspect of a differential gene analysis pipeline (see the course schedule below for the list of papers). This assignment will be worth 16 pts (worth 20% of the final grade). On the first day of class, learners will sign up to present one of the papers to be discussed in weeks 2 through 6 of the course. Presenters will be responsible for leading a discussion of the paper background, methods

(both experimental and statistical), results, and overall conclusions for their peers. A powerpoint should be generated and used by the presenters to guide the class discussion. For further information on criteria used for assessment, please see the rubric on ELMS.

Assignment 2 - Hands-on with Differential Gene Expression Analysis: Throughout the course, students will work with a small instructor-provided RNAseq dataset and take it through a bioinformatic pipeline. Students will first have the opportunity to participate in in-class discussions focused on the interpretation of the differential gene expression results from this dataset. The assignment will involve using a new instructor-provided dataset and independently taking their data through the bioinformatic pipeline discussed throughout the course. Students will then prepare a written assignment describing and interpreting the results. The rubric for this assignment will be posted in Week 5 of class and the assignment will be due by ELMS upload on Monday May 9, 2022 at 11:59pm. This assignment will be worth 16 pts. (20% of the final course grade).

Grading Scale:

100-90%	4.0	79-75%	2.5
89-85%	3.5	74-70%	2.0
84-80%	3.0		

Schedule:

Week	Location	Topic	Readings
Mar. 30	Gahan	Introduction to Second and Third Generation Sequencing Platforms, the Basics of Gene Expression Analysis by RNAseq	Heather and Chain 2016 , Nagalakshmi et al. 2008 *
Apr. 6	Gahan	Experimental Design, RNAseq Data File Structures and Assessing Data Quality	Schurch et al. 2016 *
Apr. 13	Gahan	Read Trimming and Filtering - Why Does It Matter?	Williams et al. 2016 *, Conesa et al. 2016
Apr. 20	Gahan	Reference Genome Alignment and Special Considerations for RNAseq Data	Dobin et al. 2013 *
Apr. 27	Gahan	Counting Genome-Aligned Reads and Differential Expression Analysis	Anders et al. 2013 , Rapaport et al. 2013 *
May 4	Gahan	Interpreting your RNAseq Results	McDermaid et al. 2018 *
May 9	NA	Final Assignment Due on ELMS	NA

* indicates that the paper will be presented for discussion in class.

Full References for Readings:

Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, and Robinson MD (2013) Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature Protocols* doi:10.1038/nprot.2013.099

Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczesniak MW, Gaffney DJ, Elo LL, Zhang X, and Mortazavi A (2016) A survey of best practices for RNA-seq data analysis. *Genome Biology* 17:13.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1): 15-21.

Heather JM, and Chain B (2016) The sequence of sequencers: The history of sequencing DNA. *Genomics* 107: 1-8.

McDermaid A, Monier B, Zhao J, Liu B, and Ma Q (2018) Interpretation of differential gene expression results of RNA-seq data: review and integration. *Briefings in Bioinformatics* doi: 10.1093/bib/bby067.

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, and Snyder M (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320: 1344-1349.

Rapaport F, Khanin R, Liang Y, Pirun M, Krek A, Zumbo P, et al. (2013) Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biology* 14(9):R95.

Schurch NJ, Schofield P, Gierlinski M, Cole C, Sherstnev A, Sing V, Wrobel N, Gharbi K, Simpson GG, Owens-Hughes T, Blaxter M, and Barton GJ (2016) How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA* 22: 839-851.

Williams CR, Baccarella A, Parrish JZ, and Kim CC (2016) Trimming of sequence reads alters RNA-Seq gene expression estimates. *BMC Bioinformatics* 17:103.