BIOS 668

Statistical Methods for High-throughput Genomic Data II

Instructor(s): Mikhail Dozmorov
Duration: 15 Weeks
Dates: Monday/Wednesday class: January 17 to May 2, 2018
Time: 9:00am - 10:20am
Location: One Capitol Square, Rm 5009
Office Hours: Monday/Wednesday 10:30am to 12:00 pm at Biostatistics Office 738

Required course material: posted via Blackboard and on the course' web site https://mdozmorov.github.io/BIOS668.2018/

Software: Unix, R programming environment, GitHub

Course description

The study of genomics and use of next-generation sequencing are at the forefront of biomedical research. Sequencing market is constantly evolving, stimulating the development of new analytical approaches and software tools. Therefore it may not be possible to maintain a stable analysis pipeline throughout a project because the lifetime of software often spans months and even years. To be able to effectively analyze and interpret genomic sequencing data, it is crucial to (1) understand the technologies that produce the data and (2) develop strong computational skills that will be flexible in this dynamic environment.

This course is a continuation of the BIOS 567 https://mdozmorov.github.io/BIOS567.2017/ and will introduce high-throughput genomic assays including DNA sequencing and genome variation analysis, transcriptome profiling with RNA-seq and miRNA-seq, metagenomics, epigenomic analysis including ChIP-seq and methylation assays, single-cell sequencing, and chromatin conformation capture technologies. The course is primarily focusing on human genomics; however, knowledge and skills gained through the course are extendable on genomics of model organisms.

Course Objectives

- Understand the core principles, strengths, and limitations of high-throughput technologies
- Learn statistical models and approaches used in sequencing data analysis
- Gain practical experience in high-throughput Exploratory Data Analysis, visualization, and quality control using Unix and R environment
- Critically evaluate and interpret statistical methods used in flagship tools for sequencing data analysis
- Interpret biological findings provided by different sequencing technologies, and be able to integrate different layers of -omics data

Course topics

- Unix overview
- Genomic technologies and Bioconductor
- Genome sequencing and alignment
- Single nucleotide polymorphism (SNP) analysis
- Copy Number Variant (CNV) analysis
- RNA- and miRNA-seq
- Chromatin Immunoprecipitation (ChIP)-, DNAse-, ATAC-seq

- Metagenomic
- Methylation
- Epigenomics
- Chromatin Conformation Capture analysis
- Single-cell sequencing
- Integrative analysis, TCGA

Homework format

- Assignments will be posted and should be submitted via VCU Blackboard, https://blackboard.vcu.edu
- Assignments should be submitted as reproducible reports in RMarkdown
- Short summaries of Reading assignments should be organized in an RMarkdown report, maintained on GitHub
- Final project should be submitted as a fully reproducible GitHub repository

Class rules

- Attendance is required
- Read all assignments before class
- Bring your laptop and the book to every class

Grading Rubric

For all assignments

In assessing students' work, the instructor will focus on the following factors applicable to all assignments (specific criteria for each assignment will be detailed later):

- Process: Were the drafts iterated and improved based on faculty feedback?
- Quality and Shine: Is it executed with skill and subtlety, and has it been edited well and - polished?
- Organization and Presentation: Is it presented clearly and in a professional manner suitable - for an audience?
- Effort and Application: Has the work been prepared with careful thought and attention to - detail, and does it take appropriate advantage of the relevant tools?
- Punctuality and Completeness: Is it on time and complete, and does it fulfill the assignment?

For Discussions

- Preparedness: Has the student completed the work necessary in preparation for the discussion (viewing assigned video, completing assigned reading or tutorial)?
- Participation: Was the student engaged in the discussion (both paying attention and participating)?

For In-class Exercises

- Effort: Did the student try to complete the exercise to better understand the lesson at hand?
- Participation: If the exercise involves collaboration, did the student contribute?

For the final project

- In addition to the above grading considerations, each student will grade one peer's project, assigned randomly. The goal here is to learn from peer's work while assessing its quality. Grading by students will be averaged with instructor's grading.

Assignment Values

Grading model

Grading for individual assignments is based on the quality of the finished work:

- A being high quality student work with minimal editing required.
- B being good quality student work.
- C being unsatisfactory work.
- D being more than unsatisfactory.
- F being atrocious, unacceptable work.

Grade	Percentage	Performance
A+	97% to 100%	Excellent
A	93% to 96%	Excellent
A-	90% to 92%	Excellent
B+	87% to 89%	Good
B	83% to 86%	Good
B-	79% to 82%	Good
C+	76% to 78%	Unsatisfactory
C	73% to 75%	Unsatisfactory
C-	70% to 72%	Unsatisfactory
D+	67% to 69%	More than unsatisfactory
D	64% to 66%	More than unsatisfactory
D-	61% to 63%	More than unsatisfactory
F	60% and below	Unacceptable

Assignments Grade Overall

Your grade is based on your performance during the semester. It includes readings, participation and in class exercises (that means you must attend all classes) and assignments, broken down as follows.

Assignment	Percentage Value
In-class participation	20%
Reading and homework assignment	50%
Final project	30%
TOTAL	100%

Deadlines policy

Deadlines are mandatory. Homework and reading assignment reviews are due two weeks from the date of assignment, unless specified otherwise. Late assignments will not receive any credit, unless discussed with the instructor.

You must turn in your assignments before the date and time given to you as a deadline. For each day you miss a deadline you will drop half a grade on your assignment. Plan ahead and remember: done is better than perfect. It will always be better to hand in something than nothing. If you are having trouble with your assignment let the instructor know immediately, do not wait until it is too late. In return you will get feedback about your assignment from the instructor within reasonable time.

Plagiarism and Copyright

It is a serious ethical violation to take any material created by another person and represent it as your own original work. Any such plagiarism will result in serious disciplinary action, possibly including dismissal from the VCU. Plagiarism may involve copying text from a book or magazine without attributing the source, or lifting words, code, photographs, videos, or other materials from the Internet and attempting to pass them off as your own. Please ask the instructor if you have any questions about how to distinguish between acceptable research and plagiarism.

In addition to being a serious academic issue, copyright is a legal issue.

Never "lift" or "borrow" or "appropriate" or "repurpose" graphics, audio, or code without both permission and attribution. This guidance applies to scripts, audio, video clips, programs, photos, drawings, and other images, and it includes images found online and in books.

Create your own graphics, seek out images that are in the public domain or shared via a creative commons license that allows derivative works, or use images from the AP Photo Bank or which the school has obtained licensing.

If you're repurposing code, be sure to keep the original licensing intact. If you're not sure how to credit code, ask.

The exception to this rule is fair use: if your story is about the image itself, it is often acceptable to reproduce the image. If you want to better understand fair use, the Citizen Media Law Project is an excellent resource.

When in doubt: ask.

Observe the VCU Honor Pledge in any class- and homework activities

University-wide policies

Importance of Diversity

It's critical that students learn to include a diverse set of voices in their stories – something that is often glossed over when finding stories in spreadsheets and online sources. You are encouraged to look for stories about and voices from communities that are underrepresented.