January – May 2026

SCHEDULE OF ACTIVITIES

WEEK 1	M	1-12	First day of classes: Introduction to the Command Line Students will gain an understanding of the basic skills required to use the command line interface, including how to log into to the Vermont Advanced Computing Core (VACC), a high-performance computing cluster. Emphasis will be placed on why command-line proficiency is essential for large-scale genomics analyses.
	W	1-14	Navigating and Working with Files Students will be introduced to the Linux file system, including directory structures, relative vs absolute paths, and essential commands for navigating, creating, moving, copying, and deleting files and directories on an HPC system.
WEEK 2	M	1-19	Martin Luther King Holiday
	W	1-21	Wild cards & Introduction to Viewing Files An overview of wildcard characters (e.g., *, ?) will be provided to efficiently select multiple files. Students will also learn how to inspect file contents using common Unix utilities, laying the groundwork for working with large text-based genomics files.
WEEK 3	M	1-26	Searching & Redirection Students will learn how to search and manipulate text files using command-line tools, including grep, pipes, and input/output redirection. Basic text editors will be introduced to support script development and workflow documentation.
	W	1-28	Intro to Shell Scripting This session introduces shell scripting as a way to automate repetitive tasks. Students will learn how to capture commands into scripts, define and use variables, and run scripts on the command line and cluster.

WEEK 4	M	2-02	Considerations when selecting an RNA-Seq dataset from GEO Experimental design principles and biological considerations for selecting appropriate RNA-Seq datasets from public repositories will be discussed, including controls, replicates, sequencing depth, and metadata quality.
	W	2-04	Accessing Public Data Students will learn how to access datasets from the NCBI Gene Expression Omnibus (GEO), submit jobs using SLURM (a Linux-based workload manager), and manage data on an HPC system. Common next-generation sequencing (NGS) file formats (FASTQ, SAM, BAM) will be introduced.
WEEK 5	M	2-09	Preprocessing and Quality Control of sequence reads Introduction to FASTQC, a widely used open-source tool for assessing the quality of high-throughput sequencing data. Students will learn how to interpret key QC metrics and identify common sequencing issues.
	W	2-11	Class Questions: Homework Assignment #5 Dedicated question-and-answer session focused on aiding students for the upcoming HW#5 submission.
WEEK 6	M	2-16	Presidents’ Day Holiday
	W	2-18	Trimming and Filtering Students will learn how to trim and filter FASTQ reads using Trimmomatic, including removal of low-quality bases and sequencing adapters to improve downstream alignment performance.
WEEK 7	M	2-23	Genome Indexing and Alignment: FASTQ to BAM Introduction to genome indexing and alignment algorithms. Provide an overview of tools (`STAR, HiSAT2, BWA-mem`) commonly used. Hands-on exercises in genome indexing and alignment will be performed during class. Post-alignment processing to remove low-quality reads, adapter sequences, and contaminants will also be covered.
	W	2-25	QC of RNA-Seq Data

			Students will use RSeQC to assess RNA-Seq alignment quality, including read distribution, gene body coverage, and strand specificity.
WEEK 8	M	3-02	Class Questions: Homework Assignment #6 Dedicated question-and-answer session focused on aiding students for the upcoming HW#6 submission.
	W	3-04	RNA-Seq: BAM to counts files This week we will process BAM files using `HTSeq`. `HTSeq` is a python package that can be used to count the number of reads that overlap with features such as genes or exons.
WEEK 9	M	3-09	Spring Recess
	W	3-11	Spring Recess
WEEK 10	M	3-16	The R basics: A Review Students have previously been exposed to R programming in the prerequisite course, MMG3310. We will review R syntax (variables, functions, arguments), and data wrangling (data structures, vectors, data frames).
	W	3-18	Data Visualization with `ggplot2` Overview of the grammar and how `ggplot2` is implemented, the building blocks of `ggplot2` plot. How to create plots including scatterplots, bar charts, and box plots will be covered. Students will come to understand aesthetics, scales, and coordinate systems of `ggplot2`.
WEEK 11	M	3-23	R and the `tidyverse` Students will learn data manipulation and summarization using `dplyr`, and integrate `tidyverse` workflows with `ggplot2` for exploratory data analysis.
	W	3-25	Class Questions: Homework Assignment #7 Dedicated question-and-answer session focused on aiding students for the upcoming HW#7 submission.
WE	M	3-30	Replicating Results from Paper #1 in R: Advanced RNA-Seq

			Students will replicate key figures and analyses from a published RNA-Seq study, reinforcing best practices in reproducible research.
	W	4-01	Paper #1 : RNA-Seq analysis in R continued Continuation of RNA-Seq analysis, interpretation of results, and discussion of biological conclusions.
WEEK 13	M	4-06	Replicating Results from Paper #2 in R: ChIPSeq Introduction to ChIP-Seq analysis concepts and replication of selected results from a published study.
	W	4-08	Paper #2 : ChIPseq analysis in R continued
WEEK 14	M	4-13	Replicating Results from Paper #3 in R: Metagenomics Overview of metagenomics workflows and replication of analyses from a published metagenomics study.
	W	4-15	Paper #3 : Metagenomics analysis in R continued
WEEK 15	M	4-20	Catch-Up Week Time reserved to complete analyses and prepare final presentations. Students will meet individually with Dr. Rodriguez for a required 30-minute one-on-one check-in. A sign-up sheet will be made available one week in advance.
	W	4-22	Catch-Up Week
WEEK 16	M	4-27	Class Presentations
	W	4-29	Class Presentations - Last day of classes