MCB4324 Computational Genomics and Epigenomics

Spring 2026 – 3 credits

Class Location

This course and all class materials (e.g. lecture podcasts, discussion papers, quizzes, assignments, exams, and final project) are available online through the Canvas course website https://elearning.ufl.edu/.

Instructor Information

Dr. Meixia Zhao
Room 1006, Microbiology & Cell Science
Phone: 352-273-3715
Email: meixiazhao@ufl.edu
Office Hours: Mondays & Wednesdays 1:00-2:00 pm or by appointment through Zoom Conferences under Canvas.

Teaching Assistant

Mr. Henry Shuket
Email: henry.shukert@ufl.edu
Office Hours: TBD or by appointment through Zoom Conferences under Canvas.

Course Description

Genomics and epigenomics utilize high-throughput sequencing technologies in understanding biology questions. The primary goal of this course is to introduce history, theory, latest advances, and computational approaches in (epi)genomics for conducting large-scale genomic analyses. Course topics include sequence alignment, genome assembly and annotation, variant identification, transcriptomics, small RNAs, DNA methylation, histone modification, open chromatin region, and 3D chromatin interaction.

Course Learning Objectives

After successful completion of this course, students should be able to:

- use basic command skills in UNIX.
- show mastery of the fundamental concepts and methodology of genomics and epigenomics.
- recognize and differentiate the advantages and disadvantages of different computational approaches and methods.
- employ and compare the computational approaches and methods in analyzing different types of high-throughput genomic data.
- interpret data generated by different computational approaches and methods.

Course Prerequisite

BSC2891, or STA2023, or MCB3020, or MCB3023, or PCB3063, or BSC4434C, or MCB4325C, or permission of the instructor. Many of the computational tools we will use are installed on the HiPerGator supercomputers. Every student will be provided with user accounts on the HiPerGator UNIX server. Access to the course UNIX server is required to complete the laboratory exercises and assignments.

Important Dates

- Midterm Exam: Feb 21-Mar 2, 2026.
- Final Exam: Apr 25-May 1, 2026.

Textbook Information

Textbook is not required. Prior to each class, PDFs and other relevant materials will be curated and made available online. In addition, supplementary handouts will be provided for review. Students are also required to read research articles related to the course topics, which will be posted on the course website.

Here are two recommended textbooks that could prove beneficial or engaging:

- Concepts in Bioinformatics and Genomics. 2017. Jamil Momand, Alison McCurdy, Notes by Silvia Heubach, and Nancy Warter-Perez. ISBN: 9780190610548.
- Computational Epigenetics and Diseases. 2019. ISBN: 978-0-12-814513-5.

The following research articles are **OPTIONAL** for you to read:

Weeks 1 & 2:

- Nayfach S. et al. 2021. A genomic catalog of Earth's microbiomes. *Nat Biotechnol* 39, 499-509. https://doi.org/1038/s41587-020-0718-6.
- Gauthier J., Vincent A.T., Charette S.J., Derome N. 2019. A brief history of bioinformatics. *Brief Bioinform* 20, 1981-1996. https://doi.org/1093/bib/bby063.

Week 3:

- Nurk S. et al. 2022. The complete sequence of a human genome. *Science* 376, 44-53. https://doi.org/1126/science.abj6987.
- Zhou X., Ren L., Meng Q., Li Y., Yu Y., Yu. J. 2010. The next-generation sequencing technology and application. *Protein Cell* 1, 520-536. https://doi.org/1007/s13238-010-0065-3.

Week 5:

- Deciphering developmental disorders study. 2017. Prevalence and architecture of de novo mutations in developmental disorders. *Nature* 542, 433-438. https://doi.org/10.1038/nature21062.
- Uffelmann, E. *et al.* Genome-wide association studies. *Nat Rev Methods Primers*1, 59. https://doi.org/10.1038/s43586-021-00056-9.

Week 6:

- Logsdon G.A., Vollger M.R., Eichler E.E. 2020. Long-read human genome sequencing and its applications. *Nat Rev Genet* 21, 597-614. https://doi.org/10.1038/s41576-020-0236-x.
- Nicholas J.D., Wayne A., Kim K., Amanda C., Christopher J.C. 2022. No one tool to rule them all: prokaryotic gene prediction tool annotations are highly dependent on the organism of study, *Bioinformatics* 38, 1198-1207, https://doi.org/10.1093/bioinformatics/btab827.

Week 7-9:

- Love M.I., Huber W., Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550. https://doi.org/10.1186/s13059-014-0550-8.

- Kim D., Paggi J.M., Park C. et al. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37, 907-915. https://doi.org/10.1038/s41587-019-0201-4.
- Liu Y., Zhou J., White K.P. 2014. RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics* 30, 301-304. doi:10.1093/bioinformatics/btt688.

Week 10:

- Lafzi A., Moutinho C., Picelli S., Heyn H. 2018. Tutorial: guidelines for the experimental design of single-cell RNA sequencing studies. *Nat Protoc* 13, 2742-2757. https://doi.org/10.1038/s41596-018-0073-y.
- Vallejos C., Risso D., Scialdone A., Marioni J.C. 2017. Normalizing single-cell RNA sequencing data: challenges and opportunities. *Nat Methods* 14, 565-571. https://doi.org/10.1038/nmeth.4292.

Week 11:

- Chen X., Rechavi O. 2022. Plant and animal small RNA communications between cells and organisms. *Nat Rev Mol Cell Biol* 23, 185-203. https://doi.org/10.1038/s41580-021-00425-y.
- Iwasaki Y.W., Siomi M.C., Siomi H. 2015. PIWI-Interacting RNA: Its Biogenesis and Functions. *Annu Rev Biochem* 84: 405-433. https://doi.org/10.1146/annurev-biochem-060614-034258.

Week 12:

- Field A.E., Robertson N.A., Wang T., Havas A., Ideker T., Adams P.D. 2018. DNA methylation clocks in aging: categories, causes, and consequences. *Mol cell* 71, 882-895. https://doi.org/1016/j.molcel.2018.08.008.
- Tse O.Y.O. et al. 2021. Genome-wide detection of cytosine methylation by single molecule real-time sequencing. *Proc Natl Acad Sci U S A* 118, e2019768118. https://doi.org/1073/pnas.2019768118.
- Yang J.H. et al. 2023. Loss of epigenetic information as a cause of mammalian aging. *Cell* 186: 305-326.e27. https://doi.org/1016/j.cell.2022.12.027.

Week 13:

- Whyte W.A., Orlando D.A., Hnisz D., Abraham B.J., Lin C.Y., Kagey M.H., Rahl P.B., Lee T.I., Young R.A. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153: 307-319. https://doi.org/1016/j.cell.2013.03.035.
- Hnisz D., Abraham B.J., Lee T.I., Lau A., Saint-André V., Sigova A.A., Hoke H.A., Young R.A. 2013. Super-enhancers in the control of cell identity and disease. *Cell* 155: 934-947. https://doi.org/10.1016/j.cell.2013.09.053.
- Skene P.J., Henikoff S. 2017. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* 6: e21856. https://doi.org/10.7554/eLife.21856.
- Meers M.P., Bryson T.D., Henikoff J.G., Henikoff S. 2019. Improved CUT&RUN chromatin profiling tools. *Elife* 8: e46314. https://doi.org/10.7554/eLife.46314.

Week 14:

- Dixon JR, Gorkin DU, Ren B. 2016. Chromatin Domains: The Unit of Chromosome Organization. *Mol Cell* 62, 668-680. https://doi.org/10.1016/j.molcel.2016.05.018.
- Sexton T, Cavalli G. 2015. The role of chromosome domains in shaping the functional genome. *Cell* 160: 1049-1059. https://doi.org/10.1016/j.cell.2015.02.040.

Teaching Support

UF Computing Help Desk & Ticket Number: All technical issues require a UF Helpdesk Ticket Number. The UF Helpdesk is available 24 hours a day, 7 days a week. https://helpdesk.ufl.edu/ | 352-392-4357.

Weekly Course Schedule

Week	Date	Topic	Quizzes and Assignments
Weeks 1-2	Jan 12-25	**UNIX Basics:** 1. Course information 2. Introduction of supercomputer 3. Overview of the UNIX operating system 4. Introduction and hands-on UNIX for shell and command lines 5. Hands-on UNIX for module and running batch jobs 6. Introduction to awk	Quiz 1 Assignment 1, due by 11:59 PM Feb 2
Week 3	Jan 23-Feb 1	**Sequencing technologies and sequences:** 1. Brief history of genomics and epigenomics 2. DNA sequencing (sequencing technologies, FASTA and FASTQ sequences, and tools) 3. Human genome project	Quiz 2 Reading assignment 1, due by 11:59 PM Feb 9
Week 4	Jan 30-Feb 8	**Sequence mapping and alignment:** 1. Sequence alignment (Dynamic programming, local and global alignments) 2. Sequence mapping (SAM/BAM, Samtools, BWA, and Bowtie2) 3. Sequence alignment (MUSCLE, ClustalW, and T-coffee)	Quiz 3 Assignment 2, due by 11:59 PM Feb 16
Week 5	Feb 6-15	**Variant identification and GWAS analysis:** 1. Single nucleotide polymorphism calling (SNPs) 2. Structural variant analysis (InDel) 3. VCF annotation and interpretation (GATK, picard, bcftools and vcftools toolkits) 4. Genotyping by sequencing and genome-wide association studies (plink)	Quiz 4 Reading assignment 2, due by 11:59 PM Feb 23 Assignment 3, due by 11:59 PM Feb 23
Week 6	Feb 13-22	**Genome assembly and annotation:** 1. Genome assembly (*de novo* genome assembly & pan genome) 2. Genome annotation (gene finding) 3. Genome annotation (transposable elements finding) 4. PacBio assembly and gene prediction (canu, circulator, prokka)	Quiz 5 Reading assignment 3, due by 11:59 PM Mar 2 Assignment 4, due by 11:59 PM Mar 2
Week 7	Feb 21-Mar 2	**Midterm exam (Weeks 1-6)**	
Week 7-9	Feb 20-Mar 16	**RNA-sequencing and differentially expressed genes:** 1. RNA-seq experimental design 2. RNA-seq reads quality control, trimming, mapping, and qualification (FASTQC, Trimmomatic, HISAT2, and HTSeq) 3. Identification of differentially expressed genes (DESeq2)	Quiz 6 Reading assignment 4, due by 11:59 PM Mar 23 Assignment 5, due by 11:59 PM Mar 23

Week	Date	Topic	Quizzes and Assignments
		4. R and RStudio 5. Gene ontology (GO) analysis 6. Clustering (Heatmap, K means, and others)	
	Mar 14-22	No classes, spring break.	
Week 10	Mar 20-29	**Single-cell RNA-seq:** 1. Single cell RNA-seq technologies 2. Preprocessing and quality control 3. Read alignment and quantification 4. Dimension reduction and visualization 5. Clustering and annotation 6. Single cell RNA-seq data analysis (Seurat)	Quiz 7
Week 11	Mar 27-Apr 5	**Small RNAs:** 1. Biogenesis and function of small RNAs (microRNAs, small interfering RNAs, and piwi-interacting RNA) 2. Small RNA identification and analyses 3. Target site prediction of small RNAs 4. bedtools	Quiz 8 Reading assignment 5, due by 11:59 PM Apr 13 Assignment 6, due by 11:59 PM Apr 13 Final project idea, due by 11:59 PM March 31
Week 12	Apr 3-12	**DNA methylation in plants and animals:** 1. DNA methylation (Initiation and maintenance of DNA methylation) 2. Techniques to measure DNA methylation 3. Identification of differentially methylated regions (DMRs) 4. Integrative analysis of DEGs, small RNAs and DMRs	Quiz 9 Reading assignment 6, due by 11:59 PM Apr 20 Assignment 7, due by 11:59 PM Apr 20
Week 13	Apr 10-19	**Histone modification:** 1. Chromosome structure and histone tail modifications 2. Histone marks (H3K9me2, H3K27me3, H3K4me3, etc.) 3. ChIP-seq principles and analysis (MACS2, IDR, and IGV) 4. Cut&Run principles	Quiz 10 Reading assignment 7, due by 11:59 PM Apr 27 Final project draft submission, due by 11:59 PM April 14
Week 14	Apr 17-22	**Chromatin interaction:** 1. Chromatin accessibility (DNase-seq, ATAC-seq including single-cell ATAC-seq, and MNase-seq) 2. 3D chromatin interaction (HiC, chromatin loops, topologically associating domains, A/B compartments, chromosome territories, and HiC-Pro analysis)	
	Apr 25-May 1	**Final exam**	Final project submission, due by 11:59 PM April 28

Note: Exact schedule may be changed based on the progress of the class.

Grade Policy
Course grading is consistent with UF grading policies.

Course Grading Structure
The course grading is based on a total of 800 points, distributed as follows:

- **Weekly or Biweekly Quizzes (100 points, 10% of final grade):** Weekly or biweekly quizzes will be administered through the Canvas course website. These quizzes are designed to reinforce your understanding of lectures and course materials.
- **Post-Class Assignments (400 points, 40% of final grade):** Regular assignments will be provided after class to reinforce your practice of computational approaches covered in lectures. Late submissions will incur a 5% penalty per day.
- **Midterm (120 points, 12% of the final grade) and final exam (180 points, 18% of the final grade):** Both exams will focus exclusively on material presented in class, assessing your understanding of fundamental concepts and techniques in computational genomics and epigenomics.

Submission Methods: You need to submit your assignment through our class Canvas website.

Grading Scale

744 – 800	points	93% - 100%	A
720 – 743	points	90% - 92.9%	A-
696 – 719	points	87% - 89.9%	B+
664 – 695	points	83% - 86.9%	B
640 – 663	points	80% - 82.9%	B-
616 – 639	points	77% - 79.9%	C+
584 – 615	points	73% - 76.9%	C
560 – 583	points	70% - 72.9%	C-
536 – 559	points	67% - 69.9%	D+
504 – 535	points	63% - 66.9%	D
480 – 503	points	60% - 62.9%	D-
Less than 480	points	<60%	E

Academic Policies and Resources
Academic policies for this course are consistent with university policies. See https://syllabus.ufl.edu/syllabus-policy/uf-syllabus-policy-links/

Campus Health and Wellness Resources
Visit https://one.uf.edu/whole-gator/topics for resources that are designed to help you thrive physically, mentally, and emotionally at UF.

Please contact UMatterWeCare for additional and immediate support.

Software Use
All faculty, staff and students of the university are required and expected to obey the laws and legal agreements governing software use. Failure to do so can lead to monetary damages and/or criminal penalties for the individual violator. Because such violations are also against university policies and rules, disciplinary action will be taken as appropriate.

Privacy abd Accessibility Policies

- Instructure (Canvas)
  - Instructure Privacy Policy
  - Instructure Accessibility
- Zoom
  - Zoom Privacy Policy
  - Zoom Accessibility