

## **MCB6937 Computational Genomics and Epigenomics**

**Spring 2024 – 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 (see below).

### **Instructor Information**

Dr. Meixia Zhao

Room 1006, Microbiology & Cell Science

Phone: 352-273-3715

Email: [meixiazhao@ufl.edu](mailto:meixiazhao@ufl.edu)

**Office Hours:** Mondays & Wednesdays 5:00 pm - 6:00 pm or by appointment.

### **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 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 Website**

Login available through Canvas <https://elearning.ufl.edu/>

### **Course Prerequisite**

None. 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 22-25, 2024, Thursday – Sunday.
- Final Exam: Apr 27-30, 2024, Saturday – Tuesday.

### **Textbook Information**

Textbook is not required. Prior to each class, PDF and other relevant documents will be curated and accessible online. In addition, supplementary handouts will be furnished for your review. Students are also required to read the research articles pertinent to the subjects, which will be posted on the course website.

Here are some 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.

Week 1:

- Nayfach S. et al. 2021. A genomic catalog of Earth's microbiomes. *Nat Biotechnol* 39, 499-509. doi: 10.1038/s41587-020-0718-6. (Required)
- Gauthier J., Vincent A.T., Charette S.J., Derome N. 2019. A brief history of bioinformatics. *Brief Bioinform.* 20, 1981-1996. doi: 10.1093/bib/bby063. (Optional)

Week 2:

- Nurk S. et al. 2022. The complete sequence of a human genome. *Science* 376, 44-53. doi: 10.1126/science.abj6987. (Required)
- Gauthier J., Vincent A.T., Charette S.J., Derome N. 2019. A brief history of bioinformatics. *Brief Bioinform.* 20, 1981-1996. doi: 10.1093/bib/bby063. (Optional)

Week 4:

- 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>. (Required)
- Uffelmann, E. et al. 2021. Genome-wide association studies. *Nat Rev Methods Primers* 1, 59. <https://doi.org/10.1038/s43586-021-00056-9>. (Optional)

Week 5:

- 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>. (Required)
- 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>. (Optional)

Week 6-8:

- 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>. (Required)
- 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. (Optional)
- 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. (Optional)

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. doi: 10.1038/s41580-021-00425-y. (Required)

Week 12:

- Field AE, Robertson NA, Wang T, Havas A, Ideker T, Adams PD. 2018. DNA methylation clocks in aging: categories, causes, and consequences. *Mol cell*, 71, 882-895. doi: 10.1016/j.molcel.2018.08.008. (Required)
- Tse OYO, et al. 2021. Genome-wide detection of cytosine methylation by single molecule real-time sequencing. *Proc Natl Acad Sci U S A* 118, e2019768118. doi: 10.1073/pnas.2019768118. (Optional)

Week 14:

- Dixon JR, Gorkin DU, Ren B. 2016. Chromatin Domains: The Unit of Chromosome Organization. *Mol Cell* 62, 668-80. doi: 10.1016/j.molcel.2016.05.018. (Required)
- Sexton T, Cavalli G. 2015. The role of chromosome domains in shaping the functional genome. *Cell* 160: 1049-1059. doi: 10.1016/j.cell.2015.02.040. (Optional)

**Grades and Grade Points:**

For information on current UF policies for assigning grade points, see <https://catalog.ufl.edu/UGRD/academic-regulations/grades-grading-policies/>.

The overall course is based on 1000 points.

- **Weekly or biweekly quizzes (100 points, 10% of final grade):** There will be a weekly or biweekly quiz that needs to be completed on the Canvas course website. These quizzes are designed to enhance your comprehension of lectures and course materials.
- **Literature reading and discussion (50 points, 5% of final grade):** Research articles will be assigned for reading. Each assigned paper requires a concise report (<300 words). We will discuss these papers in the Discussions section within Canvas.
- **Assignments after class (400 points, 40% of final grade):** Regular post-class assignments will be provided. Late submissions will incur a 5% penalty per day. These assignments aim to reinforce your practice of computational approaches covered in lectures.
- **Final project (150 points, 15% of final grade):** There will be a final project at the last quarter of the semester. This project component is to give you practice in applying computational methods to address problems in biology. It can relate to the independent research or any topics of interest. Students are required to design, develop, and conduct the project using the computational approaches learned in this course. Projects can be conducted individually or in groups (up to 4 members). You will submit the final project report in the format of a professional publication, including Abstract, Introduction, Results, Discussion, Methods, and References. If working in a group, each member's contributions should be clearly outlined in an "Author Contributions" section. Draft submission is due by March 31, and the final paper must be submitted by April 28.
- **Midterm (120 points, 12%) and final exam (180 points, 18% of final grade):** Both exams will solely cover material presented during class, assessing your grasp of basic concepts and techniques in computational genomics and epigenomics.

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

930 – 1000	points	93% - 100%	A
900 – 929	points	90% - 92.9%	A-
870 – 899	points	87% - 89.9%	B+
830 – 869	points	83% - 86.9%	B
800 – 829	points	80% - 82.9%	B-

770 – 799	points	77% - 79.9%	C+
730 – 769	points	73% - 76.9%	C
700 – 729	points	70% - 72.9%	C-
670 – 699	points	67% - 69.9%	D+
630 – 669	points	63% - 66.9%	D
600 – 629	points	60% - 62.9%	D-
Less than 600	points	<60%	F

### **Attendance and Make-Up Work**

Requirements for class attendance and make-up exams, assignments and other work are consistent with university policies that can be found at:

<https://catalog.ufl.edu/UGRD/academic-regulations/attendance-policies/>.

### **Tentative Course Schedule**

<b>Week</b>	<b>Date</b>	<b>Topic</b>	<b>Quizzes and Assignments</b>
Weeks 1-2	Jan 8-19	<b>UNIX Basics:</b> 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
Week 3	Jan 22-26	<b>Sequencing technologies and sequences:</b> 1. Brief history of genomics and epigenomics 2. DNA sequencing (sequencing technologies, FASTA and FASTQ sequences, and tools) 3. Human genome project	Assignment 1, due at 11:59 PM Feb 4
Week 4	Jan 29 - Feb 2	<b>Sequence mapping and alignment:</b> 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 2
Week 5	Feb 5-9	<b>Variant identification and GWAS analysis:</b> 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 3 Assignment 2, due at 11:59 PM Feb 18
Week 6	Feb 12-16	<b>Genome assembly and annotation:</b> 1. Genome assembly ( <i>de novo</i> 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 4
Week 7	Feb 22-25	<b>Midterm exam (Weeks 1-6)</b>	

Week 7-9	Feb 19-Mar 8	<b>RNA-sequencing and differentially expressed genes:</b> <ol style="list-style-type: none"> <li>1. RNA-seq experimental design</li> <li>2. RNA-seq reads quality control, trimming, mapping, and qualification (FASTQC, Trimmomatic, HISAT2, and HTSeq)</li> <li>3. Identification of differentially expressed genes (DESeq2)</li> <li>4. R and RStudio</li> <li>5. Gene ontology (GO) analysis</li> <li>6. Clustering (Heatmap, K means, and others)</li> </ol>	Quiz 5 Assignment 3, due at 11:59 PM Mar 10  Assignment 4, due at 11:59 PM Mar 17
Week 10	Mar 11-15	<b>Single-cell RNA-seq:</b> <ol style="list-style-type: none"> <li>1. Single cell RNA-seq technologies</li> <li>2. Preprocessing and quality control</li> <li>3. Read alignment and quantification</li> <li>4. Dimension reduction and visualization</li> <li>5. Clustering and annotation</li> <li>6. Single cell RNA-seq data analysis (Seurat)</li> </ol>	Quiz 6 Assignment 5, due at 11:59 PM Mar 31
Week 11	Mar 18-22	No classes, spring break.	
Week 12	Mar 25-29	<b>Small RNAs:</b> <ol style="list-style-type: none"> <li>1. Biogenesis and function of small RNAs (microRNAs, small interfering RNAs, and piwi-interacting RNA)</li> <li>2. Small RNA identification and analyses</li> <li>3. Target site prediction of small RNAs</li> <li>4. bedtools</li> </ol>	Quiz 7 Assignment 6, due at 11:59 PM Apr 7
Week 13	Apr 1-5	<b>DNA methylation in plants and animals:</b> <ol style="list-style-type: none"> <li>1. DNA methylation (Initiation and maintenance of DNA methylation)</li> <li>2. Techniques to measure DNA methylation</li> <li>3. Identification of differentially methylated regions (DMRs)</li> <li>4. Integrative analysis of DEGs, small RNAs and DMRs</li> </ol>	Quiz 8 Assignment 7, due at 11:59 PM Apr 14
Week 14	Apr 8-12	<b>Histone modification:</b> <ol style="list-style-type: none"> <li>1. Chromosome structure and histone tail modifications</li> <li>2. Histone marks (H3K9me2, H3K27me3, H3K4me3, etc.)</li> <li>3. ChIP-seq principles and analysis (MACS2, IDR, and IGV)</li> </ol>	Quiz 9 Assignment 8, due at 11:59 PM Apr 21
Week 15	Apr 15-19	<b>Chromatin interaction:</b> <ol style="list-style-type: none"> <li>1. Chromatin accessibility (DNase-seq, ATAC-seq including single-cell ATAC-seq, and MNase-seq)</li> <li>2. 3D chromatin interaction (HiC, chromatin loops, topologically associating domains, A/B compartments, chromosome territories, and HiC-Pro analysis)</li> </ol>	Quiz 10
Week 16	Apr 22-26	<b>Final project, review, and discussion</b>	Final project due at 11:59 PM Apr 28
	Apr 27-30	<b>Final exam</b>	

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

### **Online Course Evaluation Process**

Student assessment of instruction is an important part of efforts to improve teaching and learning. At the end of the semester, students are expected to provide feedback on the quality

of instruction in this course using a standard set of university and college criteria. Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Guidance on how to give feedback in a professional and respectful manner is available at:

<https://gatorevals.aa.ufl.edu/students/>. Students will be notified when the evaluation period opens and can complete evaluations through the email they receive from GatorEvals, in their Canvas course menu under GatorEvals, or via <https://ufl.bluera.com/ufl/>. Summaries of course evaluation results are available to students at: <https://gatorevals.aa.ufl.edu/public-results/>.

### **Academic Honesty**

As a student at the University of Florida, you have committed yourself to uphold the Honor Code, which includes the following pledge: "*We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honesty and integrity.*" You are expected to exhibit behavior consistent with this commitment to the UF academic community, and on all work submitted for credit at the University of Florida, the following pledge is either required or implied: "*On my honor, I have neither given nor received unauthorized aid in doing this assignment.*"

It is assumed that you will complete all work independently in each course unless the instructor provides explicit permission for you to collaborate on course tasks (e.g. assignments, papers, quizzes, exams). Furthermore, as part of your obligation to uphold the Honor Code, you should report any condition that facilitates academic misconduct to appropriate personnel. It is your individual responsibility to know and comply with all university policies and procedures regarding academic integrity and the Student Honor Code. Violations of the Honor Code at the University of Florida will not be tolerated. Violations will be reported to the Dean of Students Office for consideration of disciplinary action. For more information regarding the Student Honor Code, please see: <http://www.dso.ufl.edu/sccr/process/student-conduct-honor-code>.

### **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.

### **Services for Students with Disabilities**

The Disability Resource Center coordinates the needed accommodations of students with disabilities. This includes registering disabilities, recommending academic accommodations within the classroom, accessing special adaptive computer equipment, providing interpretation services and mediating faculty-student disability related issues. Students requesting classroom accommodation must first register with the Dean of Students Office. The Dean of Students Office will provide documentation to the student who must then provide this documentation to the instructor when requesting accommodation

0001 Reid Hall, 352-392-8565, [www.dso.ufl.edu/drc/](http://www.dso.ufl.edu/drc/)

### **Campus Helping Resources**

Students experiencing crises or personal problems that interfere with their general well-being are encouraged to utilize the university's counseling resources. The Counseling & Wellness Center provides confidential counseling services at no cost for currently enrolled students. Resources are available on campus for students having personal problems or lacking clear career or academic goals, which interfere with their academic performance.

- *University Counseling & Wellness Center, 3190 Radio Road, 352-392-1575, [www.counseling.ufl.edu](http://www.counseling.ufl.edu)*  
Counseling Services  
Groups and Workshops  
Outreach and Consultation  
Self-Help Library  
Wellness Coaching
- U Matter We Care, [www.umatter.ufl.edu/](http://www.umatter.ufl.edu/)
- *Career Connections Center, First Floor JWRU, 392-1601, <https://career.ufl.edu/>.*
- Student Success Initiative, <http://studentsuccess.ufl.edu>.

### **Student Complaints**

- Residential Course: <https://sccr.dso.ufl.edu/policies/student-honor-code-student-conduct-code/>.
- Online Course: <http://www.distance.ufl.edu/student-complaint-process>