Genomics and Applied Bioinformatics

BIOS 4150 A/AL & BIOL 6150 A/AL, 3 credits

Mon & Wed, 2:00 pm - 2:50 pm, Ford Environmental Sci & Tech (ES&T) L1205
Fri, 12:30 pm - 3:15 pm, Ford Environmental Sci & Tech (ES&T) L1205

Instructor Information

Instructor	Email	Drop-in Hours & Location
King Jordan	king.jordan@biology.gatech.edu	EBB 2115B, 4:00 pm, Mon
Teaching Assistant(s)	**Email**	**Drop-in Hours & Location**
TBD		ES&T L1205, 2:00 pm, Fri
TBD		ES&T L1205, 2:00 pm, Fri

General Information

Description

This course provides a practical introduction to foundational analysis techniques in genomics and bioinformatics, using the human genome as a working model. Course topics include: human genome structure and variation, genetic epidemiology, common and rare disease genetics, pharmacogenomics, population genomics, and genetic ancestry. The course will emphasize best practices in next-generation sequence (NGS) analysis and reproducible bioinformatics, along with high-performance and cloud computing. Lectures will cover the conceptual background, and lab sessions will cover the implementation and use of bioinformatics methods.

Pre- &/or Co-Requisites

The course assumes prior working knowledge of biology, genetics, scientific computing, and programming in R and Python. The course is considered as complementary to the concurrently offered course BIOL 7200 Programming for Bioinformatics, and it will build on bioinformatics skills taught in BIOL 7200.

Course Goals and Learning Outcomes

Upon successful completion of this course, you should be able to:

- Conduct collaborative bioinformatics projects using high-performance and cloud-computing resources
- Conduct and document reproducible bioinformatics analyses using Jupyter Notebook
- Perform foundational techniques for NGS analysis: data access, quality control, read mapping, and variant calling
- Statistically model disease outcomes
- Clinically interpret human genome sequence variation
    - Common and rare diseases
    - Drug response
- Characterize genome-wide patterns of genetic diversity and ancestry

Course Requirements & Grading

Students will be graded based on ten projects performed using Jupyter Notebook. Jupyter Notebooks will be run and evaluated on the class high-performance compute (HPC) cluster. Projects will be performed in small student groups, and students will have one week to complete each project. The project schedule is shown as part of the class schedule (below).

Grading Scale

Your final grade will be assigned as a letter grade according to the following scale. Grades will not be curved.

A	90-100%
B	80-89%
C	70-79%
D	60-69%
F	0-59%

Course Materials

Course Text

Course readings will be taken from the primary scientific literature and provided to students via Canvas.

Additional Materials/Resources

Students will be given access to a dedicated compute cluster as part of the Georgia Tech PACE-ICE cloud HPC environment. Students will be provided with markdown tutorials and Jupyter Notebook frameworks for genomics and multi-omics methods.

Course Website and Other Classroom Management Tools

The course will use the Georgia Tech Canvas classroom management tool.

Course Expectations & Guidelines

Academic Integrity

Georgia Tech aims to cultivate a community based on trust, academic integrity, and honor. Students are expected to act according to the highest ethical standards. For information on Georgia Tech's Academic Honor Code, please visit http://www.catalog.gatech.edu/policies/honor-code/ or http://www.catalog.gatech.edu/rules/18/.

Any student suspected of cheating or plagiarizing on a quiz, exam, or assignment will be reported to the Office of Student Integrity, who will investigate the incident and identify the appropriate penalty for violations.

Accommodations for Students with Disabilities

If you are a student with learning needs that require special accommodation, contact the Office of Disability Services at (404)894-2563 or http://disabilityservices.gatech.edu/, as soon as possible, to make an appointment to discuss your special needs and to obtain an accommodations letter. Please also e-mail me as soon as possible in order to set up a time to discuss your learning needs.

Attendance and/or Participation

Attendance and participation at Monday and Wednesday lecture sessions and Friday bioinformatics lab sessions is mandatory. Illness is considered an excused absence; students are encouraged to stay home when sick in order to take of themselves and each other. Please see http://www.catalog.gatech.edu/rules/4/ for information about institute expectations and restrictions around attendance, including information about excused absences.

Collaboration, Group Work & Artificial Intelligence (AI)

Bioinformatics is a highly collaborative field. Collaboration and group work are encouraged in order to emulate the environment that students will encounter in the working world. Projects will be conducted in student groups. Students are free to use outside resources to support their project work, much as they would do if they were in the working world. This includes the use of artificial intelligence (AI), which is

becoming an essential part of the bioinformatics toolkit. Intelligent and ethical use of AI is essential. AI should be used to support students' work rather than as a substitute. Students should always verify and validate the output of any AI models used to support their work, and the limitations of AI should be recognized. Disclosure of AI use is mandatory.

Extensions, Late Assignments, & Re-Scheduled/Missed Exams

Projects will be submitted as Jupyter notebooks via the Georgia Tech PACE-ICE cloud HPC environment. Missed project deadlines will result in a grade of zero.

Note that Georgia Tech makes some exceptions for "approved Institute activities" (e.g. field trips and athletic events) and religious observances, which will be handled on a case-by-case basis. See http://www.catalog.gatech.edu/rules/4/ for more information. We will also consider the impact of events like the All-Majors Career Fair, and off-campus interviews, and plan accordingly.

Student-Faculty Expectations Agreement

At Georgia Tech we believe that it is important to strive for an atmosphere of mutual respect, acknowledgement, and responsibility between faculty members and the student body. See http://www.catalog.gatech.edu/rules/22/ for an articulation of some basic expectation that you can have of me and that I have of you. In the end, simple respect for knowledge, hard work, and cordial interactions will help build the environment we seek. Therefore, I encourage you to remain committed to the ideals of Georgia Tech while in this class.

Student Use of Mobile Devices in the Classroom

As research on learning shows, unexpected noises and movement automatically divert and capture people's attention, which means you are affecting everyone's learning experience if your cell phone, tablet, laptop, etc. makes noise or is visually distracting during class. That said, many students find it useful to have a mobile device on hand to access course materials.

With this in mind, students are allowed to take notes or access course materials on mobile devices, with the sound off so that other students' learning is not disrupted. Please refrain from doing anything other than taking notes or looking at course materials on your mobile devices.

Campus Resources for Students

Georgia Tech provides a number of campus resources in support of students' mental health and emotional well-being, including but not limited to:

- Student Life Mental Health and Well-being https://studentlife.gatech.edu/services/mental-health-well-being
- The Center for Mental Health Care & Resources https://mentalhealth.gatech.edu/
- Stamps Health Services Psychiatric Clinic https://health.gatech.edu/psych/

Class Schedule

Week	Day	Topic	Assignment	Points
1	Mon	Introduction & logistics		
	Wed	Reproducible bioinformatics & AI		
	Fri	PACE-ICE & Jupyter Lab	Group formation	
2	Mon	Human genome		
	Wed	Reference genome sequence		
	Fri	PACE-ICE & Jupyter Lab	Group formation	
3	Mon	Labor Day holiday		
	Wed	Genome sequencing		
	Fri	Genome data structure lab	Group contracts	
4	Mon	Data access		
	Wed	Quality control		
	Fri	Data access & Quality control lab	Project #1	100
5	Mon	Alignment & read mapping		
	Wed	Variant calling		
	Fri	Variant calling lab	Project #2	100
6	Mon	Population biobanks		
	Wed	Electronic health records		
	Fri	Phecode lab	Project #3	100
7	Mon	Genetic epidemiology		
	Wed	Disease modeling		
	Fri	Modeling lab	Project #4	100
8	Mon	Fall Break		
	Wed	Make up and review		
	Fri	Review lab		
9	Mon	Genetics of rare, monogenic disease		
	Wed	Clinical genetic testing		
	Fri	Clinical genetics testing lab	Project #5	100
10	Mon	Genetics of complex, common disease		
	Wed	Genome-wide association studies		
	Fri	Genome-wide association study lab	Project #6	100
11	Mon	Disease risk prediction		
	Wed	Polygenic risk scores		
	Fri	Polygenic risk score lab	Project #7	100
12	Mon	Pharmacogenomics		
	Wed	Pharmacogenomic diversity		
	Fri	Pharmacogenomics lab	Project #8	100
13	Mon	Genetic diversity		
	Wed	Principal components analysis		
	Fri	Principal components analysis lab	Project #9	100
14	Mon	Genetic ancestry		
	Wed	Admixture		
	Fri	Admixture lab	Project #10	100
15	Mon	Make up and review		
	Wed	Thanksgiving break		
	Fri	Thanksgiving break		
16	Mon	Course review		