

**Graduate School of Public Health
Department of Human Genetics**

**HUGEN 2072
Genomic Data Pipelines and Tools**

Spring 2023

Mondays · 10:30 AM–11:50 AM
Wednesdays · 10:30 AM–11:50 AM

Break · 11:05 AM–11:10 AM

Public Health 3121C
3 Credits

COURSE DESCRIPTION

This course will teach the analytical methods and tools used for genotype data quality control, sequencing read alignment, base calling, genotype calling, quantitative sequencing methods, data harmonization, genotype imputation, and statistical analysis and meta-analysis through the development and implementation of next-generation whole genome sequencing and RNA-Seq pipelines.

COURSE OBJECTIVES

Upon completion of this course, the student will be able to:

- Implement the necessary elements of a genotype quality control pipeline.
- Implement the necessary elements of a genome-wide associate study and genome-wide association meta-analysis
- Evaluate the qualitative and quantitative output from such pipelines.
- Implement the necessary elements of a next-generation whole-genome sequencing analysis pipeline.
- Implement the necessary elements of an RNA-Seq pipeline.
- Analyze the qualitative and quantitative output from such pipelines for genetic associations and expression analyses.

COURSE PREREQUISITES

HUGEN 2071 · Genomic Data Processing & Structures
HUGEN 2022 · Human Population Genetics

or approval of the instructor

FACULTY

Ryan L. Minster, Ph.D., M.S.I.S.

3118 Public Health

412-624-6928

rminster@pitt.edu

Office hours Wednesdays 3–4 PM
and upon request

Hyun Jung “H.J.” Park, Ph.D.

3122 Public Health

412-383-0520

hyp15@pitt.edu

Office hours TBD and upon request

EVALUATION AND GRADING

Evaluation will be based on the following components:

Syllabus Review (1)

There will be one online syllabus review. It will consist of 10 questions posed online through Canvas about the syllabus. The online syllabus review is passed by correctly answering 80% of its questions. *The deadline for the first attempt of this assignment is Thu Jan 12. It can be retaken an unlimited number of times until passed up to Wed Jan 18 at 11:59 PM EST.*

Coding Review (1)

This assessment consists of 20 coding tasks with R, Unix scripting, PLINK, and bcftools.

Five-Minute Papers (25)

At the end of each class session—except those devoted solely to a case study—students will submit responses to a few questions about the session’s content. If a student is unable to attend a class session and needs to leave early, they can submit their five-minute paper up to forty-eight hours after the session (facilitated by viewing the recorded session via Canvas).

Case Study Engagement (6)

There will be six case studies examined during the course. For each case study, students will submit two engagement questions or comments on the case study to facilitate discussion. When a case study engagement is submitted late (with a mulligan), it must be submitted as a paper critique (see below). *Such late-submitted case studies do not count toward an A+.*

Paper Critiques (2)

Two current papers from *bioRxiv* or *medRxiv* will be provided for critique. Each critique will be a 1000-word review of the paper’s methods. *Only required if you want to earn an A+.*

Projects (6)

There will be five course projects: (1) genotype quality control, (2) genome-wide analysis and meta-analysis, (3) sequencing for genotyping and variant discovery, (4) sequencing for structural variation discovery, (5) RNA-seq for differential expression, and (6) RNA-seq for lncRNA identification and analysis.

The outcomes of each project will be (1) the programming scripts necessary to conduct the analyses, (2) written narrative of the methods and results, and (3) a ten-minute recorded presentation of the methods and results and responses to questions from the instructors.

Late Policy & Revision Policy

Each student begins the term with four mulligans. Each mulligan can be used for a forty-eight-hour extension on an assignment; an opportunity to retake the coding review; or an opportunity to revise a project or paper critique. You can use additional mulligans for further extensions on a project, so that you could spend two mulligans for a 96-hour extension.

Grading

Letter grades for the course are assigned based on the number of items in each component for which the student earns a 'satisfactory.'

Course Grade	D	C	B	A	A+	
Syllabus Review	8/10	8/10	8/10	8/10	8/10	
Coding Review Tasks	16/20	16/20	16/20	16/20	16/20	
Five-Minute Papers	15/25	15/25	19/25	22/25	25/25	Each five-minute paper, each project component, and each paper critique will be graded satisfactory/unsatisfactory.
Case-Study Engagement		4/6	4/6	5/6	6/6	
Project <i>Scripts</i>		6/6	6/6	6/6	6/6	
Project <i>Narratives</i>			5/6	5/6	6/6	
Project <i>Presentations</i>				5/6	6/6	
Paper Critiques					2/2	

For example, to earn a **B**, a student must

- satisfactorily answer 8 out of 10 questions on the syllabus review,
- satisfactorily complete 16 out of 20 coding review tasks,
- satisfactorily complete 19 out of 25 five-minute papers,
- satisfactorily complete 4 out of 6 case-study engagements,
- submit and satisfactorily complete 6 of 6 project scripts, and
- submit and satisfactorily complete 5 of 6 project narratives.

To earn an **A**, a student must

- do all over the above *as well as*
- satisfactorily complete 3 more five-minute papers,
- satisfactorily complete 1 more case-study engagement, and
- submit and satisfactorily complete 5 of 6 project presentations.

To earn an **A+**, a student must

- do all over the above *as well as*
- satisfactorily complete 3 more five-minute papers (that is, all five-minute paper),
- satisfactorily complete 1 more case-study engagement (that is, all case studies), *and also must*
- submit and satisfactorily complete 2 of 2 paper critiques.

Remember that unsatisfactory five-minute papers, projects components, and critiques can be revised by using a mulligan.

CANVAS INSTRUCTION

This course will use the University's Canvas site (canvas.pitt.edu). Each lecture will be accompanied by supporting material and further reading, all of which will be made available around the time of the lecture. It is the student's responsibility to check for, and read, this material. The instructors will use Canvas as the primary means of communicating with the students, who are expected to check the site on a regular basis throughout the semester.

Accessibility

Ensuring an accessible and pleasant experience to all users, regardless of disability, is a key focus of Canvas. The Canvas platform was built using the most modern HTML and CSS technologies and is committed to W3C's Web Accessibility Initiative and §508 (www.section508.gov) guidelines.

COURSE MATERIALS

Required Software (All available free online)

Web Browser

R
www.r-project.org

R Studio
www.rstudio.com

Pulse Secure
www.pulsesecure.com

Required Readings

Genomics in the Cloud: Using Docker, GATK, and WDL in Terra
by Geraldine A. Van der Auwera and Brian D. O'Connor
Online free through Pitt Libraries

Selected papers from the current literature.

Optional Readings

Current Protocols in Bioinformatics — Online free through Pitt Libraries
[dx.doi.org/10.1002/0471250953](https://doi.org/10.1002/0471250953)

SCHEDULE

Date	Activity or Deadline
Mon Jan 9	Lecture: Introduction; Review of R, R Studio, R Markdown Due: Five-Minute Paper 1
Wed Jan 11	Read: <i>Genomics in the Cloud</i> , Chapter 3, pp. 53–77 Lecture: Review of Unix Due: Five-Minute Paper 2
Thu Jan 12	Due: Syllabus Review
Mon Jan 16	Lecture: No Class
Wed Jan 18	Lecture: Review of PLINK & GDS Due: Five-Minute Paper 3
Mon Jan 23	Read: Gunderson <i>et al. Nat Genet</i> 2005 https://dx.doi.org/10.1038/ng1547 Ha <i>et al. Eur J Hum Genet</i> 2014 https://doi.org/10.1038/ejhg.2013.304 Bien <i>et al. PLoS One</i> 2016 https://dx.doi.org/10.1371/journal.pone.0167758 Verlouw <i>et al. Eur J Hum Genet</i> 2021 https://dx.doi.org/10.1038/s41431-021-00917-7 Lecture: Genotyping Array Data Generation Due: Five-Minute Paper 4
Mon Jan 25	Read: Laurie <i>et al. Genet Epidemiol</i> 2010 https://dx.doi.org/10.1002/gepi.20516 Lecture: Genotyping: Missingness & Sex Chromosome Checks Due: Five-Minute Paper 5
Sun Jan 29	Due: Coding Review
Mon Jan 30	Read: Gogarten <i>et al. Bioinformatics</i> 2012 https://dx.doi.org/10.1093/bioinformatics/bts610 Lecture: Genotyping: B Allele Frequencies & LRR Due: Five-Minute Paper 6
Wed Feb 1	Read: Novembre <i>et al. Nature</i> 2008 https://dx.doi.org/10.1038/nature07331 Conomos <i>et al. Genet Epidemiol</i> 2015 https://dx.doi.org/10.1002/gepi.21896 Rosenberg. <i>Ann Hum Genet</i> 2006 https://dx.doi.org/10.1111/j.1469-1809.2006.00285.x Conomos <i>et al. Am J Hum Genet</i> 2016 https://dx.doi.org/10.1016/j.ajhg.2015.11.022 Lecture: Genotyping: Relatedness & Population Substructure Due: Five-Minute Paper 7
Mon Feb 6	Lecture: Genotyping: Mendelian Errors & Duplicate Errors Due: Five-Minute Paper 8

Date	Activity or Deadline
Wed Feb 8	<p>Read: Marchini & Howie <i>et al. Nat Rev Genet</i> 2010 https://dx.doi.org/10.1038/nrg2796 Loh <i>et al. Nat Genet</i> 2016 https://dx.doi.org/10.1038/ng.3679 Das <i>et al. Nat Genet</i> 2016 https://dx.doi.org/10.1038/ng.3656 McCarthy <i>et al. Nat Genet</i> 2016 https://dx.doi.org/10.1038/ng.3643</p> <p>Lecture: Genotype Imputation</p> <p>Due: Five-Minute Paper 9</p>
Sun Feb 12	Due: Project 1 · Genotype Quality Control
Mon Feb 13	<p>Read: Risch & Merikangas <i>Science</i> 1996 https://dx.doi.org/10.1126/science.273.5281.1516 Uffelmann <i>et al. Nat Rev Methods Primers</i> 2021 https://doi.org/10.1038/s43586-021-00056-9 Tam <i>et al. Nat Rev Genet</i> 2019 https://dx.doi.org/10.1038/s41576-019-0127-1 Gogarten <i>et al. Bioinformatics</i> 2019 https://dx.doi.org/10.1093/bioinformatics/btz567</p> <p>Lecture: Genome-Wide Association & GWAS Quality Control</p> <p>Due: Five-Minute Paper 10</p>
Wed Feb 15	<p>Read: Willer <i>et al. Bioinformatics</i> 2010 https://dx.doi.org/10.1093/bioinformatics/btq340 Winkler <i>et al. Nat Protoc</i> 2014 https://dx.doi.org/10.1038/nprot.2014.071</p> <p>Lecture: Meta-Analysis Data Processing & Quality Control</p> <p>Due: Five-Minute Paper 11</p>
Sun Feb 19	Due: Case Study 1 Engagement Questions & Comments
Mon Feb 20	Activity: Case Study 1 · GWAS / GWAMA
Wed Feb 22	<p>Read: Schuster <i>et al. Nat Methods</i> 2008 https://dx.doi.org/10.1038/nmeth1156 Goodwin <i>et al. Nat Rev Genet</i> 2016 https://dx.doi.org/10.1038/nrg.2016.49 Gawad <i>et al. Nat Rev Genet</i> 2016 https://dx.doi.org/10.1038/nrg.2015.16 De Coster <i>et al. Nat Rev Genet</i> 2021 https://dx.doi.org/10.1038/s41576-021-00367-3</p> <p>Lecture: Sequencing: Data Generation</p> <p>Due: Five-Minute Paper 12</p>
Sun Feb 26	Due: Project 2 · Genome-Wide Analysis & Meta-Analysis
Mon Feb 27	<p>Read: <i>Genomics in the Cloud</i>, Chapter 2, pp. 13–52</p> <p>Lecture: Sequencing: Read Quality Control</p> <p>Due: Five-Minute Paper 13</p>
Wed Mar 1	Read: Li & Durbin. <i>Bioinformatics</i> 2009

Date	Activity or Deadline
	<p>https://dx.doi.org/10.1093/bioinformatics/btp324 Reinert <i>et al. Annu Rev Genomics Hum Genet</i> 2015 https://dx.doi.org/10.1146/annurev-genom-090413-025358 Schneider <i>et al. Genome Res</i> 2017 https://dx.doi.org/10.1101/gr.213611.116</p> <p>Lecture: Sequencing: Alignment & Assembly</p> <p>Due: Five-Minute Paper 14</p>
Mon Mar 6	Lecture: No Class · Spring Break
Wed Mar 7	Lecture: No Class · Spring Break
Mon Mar 13	<p>Read: <i>Genomics in the Cloud</i>, Chapters 5–6, pp. 115–182</p> <p>Lecture: Sequencing: Genotype Calling/Variant Discovery</p> <p>Due: Five-Minute Paper 15</p>
Tue Mar 14	Due: Case Study 3 Engagement Questions & Comments
Wed Mar 15	<p>Read: Salk <i>et al. Nat Rev Genet</i> 2018 https://dx.doi.org/10.1038/nrg.2017.117</p> <p>Lecture: Sequencing: Genotypes Quality Control</p> <p>Activity: Case Study 2 · Sequencing</p>
Sun Mar 19	Due: Paper Critique 1
Mon Mar 20	<p>Read: Zhao <i>et al. Am J Hum Genet</i> 2021 https://doi.org/10.1016/j.ajhg.2021.03.014 Ho <i>et al. Nat Rev Genet</i> 2020 https://dx.doi.org/10.1038/s41576-019-0180-9</p> <p>Lecture: Sequencing: Structural Variation Discovery</p> <p>Due: Five-Minute Paper 16</p>
Tue Mar 21	Due: Case Study 4 Engagement Questions & Comments
Wed Mar 22	<p>Lecture: Sequencing: Structural Variation Discovery</p> <p>Due: Five-Minute Paper 17</p> <p>Activity: Case Study 3 · Structural Variation Discovery</p>
Sun Mar 26	Due: Project 3 · Sequencing for Genotyping & Variant Discovery
Mon Mar 27	<p>Read: Sherman <i>et al. Nat Genet</i> 2018 https://dx.doi.org/10.1038/s41588-018-0273-y Sherman & Salzberg, <i>Nat Rev Genet</i> 2020 https://dx.doi.org/10.1038/s41576-020-0210-7</p> <p>Lecture: Alignment against Pangenomes</p> <p>Due: Five-Minute Paper 18</p>

Date	Activity or Deadline
Wed Mar 29	<p>Read: <i>Genomics in the Cloud</i>, Chapter 7, pp. 183–208 Xu. <i>Comput Struct Biotechnol J</i> 2018 https://dx.doi.org/10.1016/j.csbj.2018.01.003 Gonzalez Casto <i>et al.</i> <i>Cancer Discov</i> 2021 https://dx.doi.org/10.1158/2159-8290.cd-20-1376 Cortés-Ciriano <i>et al.</i> <i>Nat Rev Genet</i> 2021 https://dx.doi.org/10.1038/s41576-021-00431-y</p> <p>Lecture: Sequencing: Somatic Variation Discovery</p> <p>Due: Five-Minute Paper 19</p>
Mon Apr 3	<p>Read: Stark <i>et al.</i> <i>Nat Rev Genet</i> 2019 https://dx.doi.org/10.1038/s41576-019-0150-2 Buccitelli & Selbach. <i>Nat Rev Genet</i> 2020 https://dx.doi.org/10.1038/s41576-020-0258-4</p> <p>Lecture: RNA-Seq: Quality Control and Mapping</p> <p>Due: Five-Minute Paper 20</p>
Wed Apr 5	<p>Lecture: RNA-Seq: messenger RNA Analysis</p> <p>Due: Five-Minute Paper 21</p>
Sun Apr 9	<p>Due: Project 4 · Sequencing for Structural Variation Discovery</p>
Mon Apr 10	<p>Lecture: RNA-Seq: non-coding RNA Analysis</p> <p>Due: Five-Minute Paper 22</p>
Wed Apr 12	<p>Read: Klemm <i>et al.</i> <i>Nat Rev Genet</i> 2019 https://doi.org/10.1038/s41576-018-0089-8</p> <p>Lecture: ChIP-Seq</p> <p>Due: Five-Minute Paper 23</p>
Wed Apr 17	<p>Read: Haque <i>et al.</i> <i>Genome Medicine</i> 2017 https://doi.org/10.1186/s13073-017-0467-4</p> <p>Lecture: RNA-Seq: Single-cell RNA Analysis</p> <p>Due: Five-Minute Paper 24</p>
Tue Apr 18	<p>Due: Case Study 4 Engagement Questions & Comments</p>
Wed Apr 19	<p>Lecture: Case Study 4 · Single-cell RNA-Seq Analysis</p> <p>Due: Project 5 · RNA-Seq for Differential Expression Analysis</p>
Mon Apr 24	<p>Read: Albert & Kruglyak. <i>Nat Rev Genet</i> 2015 https://dx.doi.org/10.1038/nrg3891</p> <p>Lecture: Integrative Analysis of Genotype & RNA-seq Data</p> <p>Due: Five-Minute Paper 25</p>
Tue Apr 25	<p>Due: Paper Critique 2</p>
Wed Apr 26	<p>Lecture: No Class</p>
Thu Apr 27	<p>Due: Project 6 · RNA-Seq for lncRNA Identification and Analysis</p>

ACADEMIC POLICIES

Academic Integrity

Students in this course will be expected to comply with the [University of Pittsburgh's Policy on Academic Integrity](#). Any student suspected of violating this obligation for any reason during the semester will be required to participate in the procedural process, initiated at the instructor level, as outlined in the University Guidelines on Academic Integrity. This may include, but is not limited to, the confiscation of the examination of any individual suspected of violating University Policy. Furthermore, no student may bring any unauthorized materials to an exam, including dictionaries and programmable calculators.

To learn more about Academic Integrity, visit the [Academic Integrity Guide](#) for an overview of the topic. For hands-on practice, complete the [Understanding and Avoiding Plagiarism tutorial](#).

Course Recording

This class or portions of this class will be recorded by the instructors for educational purposes. These recordings will be shared only with students enrolled in the course via Canvas.

To ensure the free and open discussion of ideas, students may not record classroom lectures, discussion and/or activities without the advance written permission of the instructor, and any such recording properly approved in advance can be used solely for the student's own private use.

Disability Resources

If you have a disability for which you are or may be requesting an accommodation, you are encouraged to contact both your instructor and [Disability Resources and Services](#) (DRS), 140 William Pitt Union, (412) 648-7890, [dresrecep@pitt.edu](mailto:drsrecep@pitt.edu), (412) 228-5347 for P3 ASL users, as early as possible in the term. DRS will verify your disability and determine reasonable accommodations for this course.

Sexual Misconduct, Required Reporting, & Title IX

If you are experiencing sexual assault, sexual harassment, domestic violence, and stalking, please report it to me and I will connect you to University resources to support you.

University faculty and staff members are required to report all instances of sexual misconduct, including harassment and sexual violence to the Office of Civil Rights and Title IX. When a report is made, individuals can expect to be contacted by the Title IX Office with information about support resources and options related to safety, accommodations, process, and policy. I encourage you to use the services and resources that may be most helpful to you.

As your professor, I am required to report any incidents of sexual misconduct that are directly reported to me. You can also report directly to Office of Civil Rights and Title IX: 412-648-7860 (M-F 8:30 AM-5:00 PM) or via the Pitt Concern Connection at: [Make A Report](#)

An important exception to the reporting requirement exists for academic work. Disclosures about sexual misconduct that are shared as a relevant part of an academic project, classroom discussion, or course assignment, are not required to be disclosed to the University's Title IX office.

If you wish to make a confidential report, Pitt encourages you to reach out to these resources:

- The University Counseling Center:
412-648-7930 (M-F 8:30 AM-5:00 PM) and
412-648-7856 (*after business hours*)
- Pittsburgh Action Against Rape (community resource): 866-363-7273 (24/7)

If you have an immediate safety concern, please contact the University of Pittsburgh Police, 412-624-2121.

Any form of sexual harassment or violence will not be excused or tolerated at the University of Pittsburgh.

Equity, Diversity & Inclusivity

The University of Pittsburgh does not tolerate any form of discrimination, harassment, or retaliation based on disability, race, color, religion, national origin, ancestry, genetic information, marital status, familial status, sex, age, sexual orientation, veteran status or gender identity or other factors as stated in the University's Title IX policy. The University is committed to taking prompt action to end a hostile environment that interferes with the University's mission. For more information about policies, procedures, and practices, visit the [Civil Rights & Title IX Compliance web page](#).

I ask that everyone in the class strive to help ensure that other members of this class can learn in a supportive and respectful environment. If there are instances of the aforementioned issues, please contact the Title IX Coordinator, by calling 412-648-7860, or e-mailing titleixcoordinator@pitt.edu. Reports can also be [filed online](#). You may also choose to report this to a faculty/staff member; they are required to communicate this to the University's Office of Diversity and Inclusion. If you wish to maintain complete confidentiality, you may also contact the University Counseling Center (412-648-7930).

Copyright Notice

These materials may be protected by copyright. United States copyright law, 17 USC § 101, *et seq.*, in addition to University policy and procedures, prohibit unauthorized duplication or retransmission of course materials.

See [Library of Congress Copyright Office](#) and the [University Copyright Policy](#).