Syllabus

Logistics

Graduate School of Public Health
Department of Human Genetics
HuGen 2070
Bioinformatics for Human Genetics

Tuesdays 1:30 PM–2:55 PM · 3121C Public Health
Fridays 1:30 PM–2:55 PM · 3121C Public Health
3 credit hours
Fall 2019

Contact Information: Please see the ‘Faculty Information’ section in CourseWeb.

University Course Catalog Description

This course focuses on manipulation and management of human genetic data. The course will cover programming approaches to the processing of phenotype and genotype data for genome-wide association analysis, sequence data, and integrated analyses in the R statistical computing language and Unix scripting language. A key component of the course will be hands-on analyses of example data sets.

Course Description

As technology makes it easier to generate large amounts of genetic data, the well-trained human geneticist needs to learn how to efficiently manipulate, manage, and annotate such data. This course will provide training in the use of the R statistical computing language and environment, in the use of genetic databases and tools, and Unix scripting as applied to data for genetic association studies. We will cover the programming and data structures that underlie the bioinformatics for genome-wide association analysis, sequence data, and integrated analyses. We will complement lectures with exercises in the use of publically-available software and intensive hands-on experience in the analyses of example data sets.

Course Goals
Upon completion of this course, the student will:

- Be able to write efficient R code to reformat, summarize, and visualize data.
- Be able to write efficient Unix scripts to reformat and summarize data.
- Have gained programming skills useful for managing and manipulating human genetic data.
- Describe and create files containing phenotypic and genetic data in a variety of common formats.
- Use data management tools written specifically for common formats to merge, filter and edit genetic data.

Faculty Availability

We welcome your questions. Please feel free to drop by our offices, or to set up an appointment. E-mail is also an excellent way to reach us. However, since we get so many e-mails, please use an informative subject line, starting with "HuGen 2070: "

Schedule 2019

Attached Files:

- [2019_Bioinf_HuGen_Schedule_26Aug2019.pdf](https://courseweb.pitt.edu/webapps/blackboard/content/listContent.jsp?course_id=_533796_1&content_id=_26144346_1&mode=reset) (57.008 KB)
<table>
<thead>
<tr>
<th>Num/Date</th>
<th>Day</th>
<th>Lecture Topic</th>
<th>Meeting</th>
<th>Objective 1</th>
<th>Objective 2</th>
<th>Objective 3</th>
<th>Notes</th>
<th>HW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8/25/19</td>
<td>Introduction/Overview</td>
<td>Seminar (2007) Chapter 1</td>
<td>Review the syllabus unpleasant</td>
<td>Describe how to handle homework</td>
<td>will recommend</td>
<td>Assign HW 1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8/25/19</td>
<td>General Information &amp; DeOx</td>
<td>Seminar (2006) Chapter 2 &amp; 1</td>
<td>Describe how to handle homework</td>
<td>Explain how to handle homework</td>
<td>will recommend</td>
<td>Assign HW 1</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>8/5/19</td>
<td>No Class</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>8/12/19</td>
<td>No Class</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>9/12/19</td>
<td>R Character Manipation</td>
<td>Seminar (2008) Chapter 2, 3, 4</td>
<td>Learn how to</td>
<td>Learn how to use regular expressions</td>
<td>Assign HW 2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>9/15/19</td>
<td>R NA 3</td>
<td>Seminar (2008) Chapter 2.5.3</td>
<td>Learn how to</td>
<td>Learn how to use regular expressions</td>
<td>Assign HW 2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>9/11/19</td>
<td>No Class</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>9/11/19</td>
<td>R Graphics &amp; Advanced Graphics</td>
<td>Seminar (2008) Chapter 2.5.3</td>
<td>Learn how to</td>
<td>Learn how to use regular expressions</td>
<td>Assign HW 2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>9/17/19</td>
<td>R Statistics</td>
<td>Seminar (2008) Chapter 2.5.3</td>
<td>Learn how to</td>
<td>Learn how to use regular expressions</td>
<td>Assign HW 2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>9/19/19</td>
<td>R Interactive &amp; Dynamic Graphics</td>
<td>Seminar (2008) Chapter 2.5.3</td>
<td>Learn how to</td>
<td>Learn how to use regular expressions</td>
<td>Assign HW 2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>9/15/19</td>
<td>PUNK &amp; PUNK Format</td>
<td>Seminar (2008) Chapter 2.5.3</td>
<td>Learn how to</td>
<td>Learn how to use regular expressions</td>
<td>Assign HW 2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>9/19/19</td>
<td>No Class</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Course Requirements and Grading

Student Performance Evaluation (Factors and Weights)

Evaluation will be based on the following components:

Homework & Quizzes (70% of final grade)

The homework assignments will comprise problems that extend the in class activities and build the skills needed to complete each of the projects.

Midterm Project (15% of final grade)

A common but not unchallenging task in working with large genetic data sets is cleaning and preparing them for deposition in the National Institutes of Health Database of Genotypes and Phenotypes (dbGaP). All NIH-funded projects that generate large-scale data must place such data into dbGaP once the data are cleaned.

The midterm project of this course asks the students to prepare mock phenotype data for deposition into dbGaP. This includes merging, cleaning, and validating the data, as well as creating a data dictionary that describe the data.

Final Project (15% of final grade)

The final project of this course asks the students to prepare mock genotype data for deposition into dbGaP. This includes merging, cleaning, and validating the data.

Late Assignment Policy

We expect homework to be handed in on time. If you are not going to be able to make the deadline, please let us know the reason why. For valid reasons, we may be able to extend the deadline for you.

Even if your homework is late, we’d prefer that you hand your homework in, as we believe you’ll learn something from attempting to do it. Reduced points may be awarded for late homework if it is handed in before we have handed out or discussed the answers and before we’ve completed correcting the homework.

Course Prerequisites

Hugen 2022 · Population Genetics

Biostat 2041 · Biostats Methods I

or approval of the instructor

Students should also have basic computing and programming skills.

Grading Scale

<table>
<thead>
<tr>
<th>Grades scored between</th>
<th>will equal</th>
</tr>
</thead>
<tbody>
<tr>
<td>97% and 100%</td>
<td>A+</td>
</tr>
<tr>
<td>94% and less than 97%</td>
<td>A</td>
</tr>
<tr>
<td>90% and less than 94%</td>
<td>A−</td>
</tr>
<tr>
<td>87% and less than 90%</td>
<td>B+</td>
</tr>
<tr>
<td>84% and less than 87%</td>
<td>B</td>
</tr>
<tr>
<td>80% and less than 84%</td>
<td>B−</td>
</tr>
</tbody>
</table>
CourseWeb/BlackBoard Instruction

This course will extensively use the University's CourseWeb site [http://courseweb.pitt.edu/](http://courseweb.pitt.edu/) [also known as BlackBoard]. To login, you must have a Pitt account. Your login ID is the same as your login ID for your Pitt account and your password is the same as for your Pitt account. 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. Discussion topics related to the course may also be posted on CourseWeb, and, for the purpose of determining a student's grade, participation in these discussions will be considered as equivalent to participation in class discussion. The instructors will use the CourseWeb site as the primary means of communicating with the students, who are expected to check the site on a regular basis throughout the semester.

Course Materials

Readings (Available online):

- **Data manipulation with R**
  *Author*: Spector, Phil.
  *Permanent URL*: [http://pittcatplus.pitt.edu/?itemid=|library/marc/voyager|5689739](http://pittcatplus.pitt.edu/?itemid=|library/marc/voyager|5689739)

- **ggplot2: Elegant Graphics for Data Analysis**
  *Author*: Wickham, Hadley Author
  *Web access*: [http://dx.doi.org/10.1007/978-0-387-98141-3](http://dx.doi.org/10.1007/978-0-387-98141-3)
  *Permanent URL*: [http://pittcatplus.pitt.edu/?itemid=|library/marc/voyager|6140635](http://pittcatplus.pitt.edu/?itemid=|library/marc/voyager|6140635)

- **Bioinformatics Data Skills**
  *Editor*: Vince Buffalo
  *Publisher*: O'Reilly

- **Bioinformatics for Geneticists (Second Edition)**
  *Editor*: Michael R. Barnes
  *DOI*: 10.1002/9780470059180

Supplemental Readings/Bibliography (Optional)

- **Introductory statistics with R**
  *Author*: Dalgaard, Peter.

- **Current Protocols in Bioinformatics**
Editor: Baxevanis AD, Stein LD, Stormo GD, Yates JR  
Publisher: John Wiley and Sons, Inc., c2010  
Web access: http://www.mrw.interscience.wiley.com/emrw/9780471250951/cp/cpbi/toc

**Single Nucleotide Polymorphisms: Methods and Protocols**  
Editor: Anton A. Komar  
Series: Methods in Molecular Biology  
Volume: 578  
Print ISBN: 978-1-60327-410-4  

**Searching by using the MeSH Database**  
Source: NCBI Handbook (PubMed help page)  

**GoPubMed: exploring PubMed with the Gene Ontology**  
Authors: Doms A, Schroeder M.  

**A quick guide for developing effective bioinformatics programming skills**  
Authors: Dudley and Butte.  
Web access: http://www.ploscompbiol.org/article/info:doi/10.1371/journal.pcbi.1003789

**R programming for bioinformatics**  
Author: Robert Gentleman  
Publisher: Boca Raton: CRC Press, c2009.

**Interactive and Dynamic Graphics for Data Analysis: with R and GGobi**  
Authors: Dianne Cook and Deborah F. Swayne  
Publisher New York: Springer Verlag 2007.

**Bioinformatics and computational biology solutions using R and Bioconductor**  
Editors: Robert Gentleman et al.  

---

**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.

All students are expected to adhere to the school’s standards of academic honesty. Cheating/plagiarism will not be tolerated. The Graduate School of Public Health’s policy on academic integrity, which is based on the University policy, is available online in the Pitt Public Health Academic Handbook. The
Plagiarism

University policy:

Integrity of the academic process requires that credit be given where credit is due. Accordingly, it is unethical to present as one's own work the ideas, representations, words of another, or to permit another to present one's own work without customary and proper acknowledgement of sources.

A student has an obligation to exhibit honesty and to respect the ethical standards of the profession in carrying out his or her academic assignments. Without limiting the application of this principle, a student may be found to have violated this obligation if he or she:

10. Presents as one's own, for academic evaluation, the ideas, representations, or words of another person or persons without customary and proper acknowledgment of sources.

11. Submits the work of another person in a manner which represents the work to be one's own.

Source: http://www.bc.pitt.edu/policies/policy/02/02-03-02.html

To avoid plagiarism, you must give “customary and proper acknowledgment of sources” by appropriately and clearly identifying which thoughts are yours and which are others, and appropriately citing your sources.

Sophisticated plagiarism detection software will be used in this course. If plagiarism is detected, you will automatically receive a grade of zero for that assignment and the incident will be reported, as required, to your Dean.

Disability Services

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, drsrecp@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.

Accessibility

Blackboard is ADA Compliant and has fully implemented the final accessibility standards for electronic and information technology covered by Section 508 of the Rehabilitation Act Amendments of 1998. Please note that, due to the flexibility provided in this product, it is possible for some material to inadvertently fall outside of these guidelines.

Copyright Notice

These materials may be protected by copyright. United States copyright law, 17 USC section 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.

Statement on Classroom Recording

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.
Diversity

The University of Pittsburgh Graduate School of Public Health considers the diversity of its students, faculty, and staff to be a strength and critical to its educational mission. Pitt Public Health is committed to creating and fostering inclusive learning environments that value human dignity and equity. Every member of our community is expected to be respectful of the individual perspectives, experiences, behaviors, worldviews, and backgrounds of others. While intellectual disagreement may be constructive, no derogatory statements, or demeaning or discriminatory behavior will be permitted. If you feel uncomfortable or would like to discuss a situation, please contact any of the following:

- the course instructor;
- the Pitt Public Health Associate Dean for Diversity at 412-624-3506 or nam137@pitt.edu;
- the University’s Office of Diversity and Inclusion at 412-648-7860 or via this anonymous reporting form.

Sexual Misconduct, Required Reporting and Title IX

The University is committed to combating sexual misconduct. As a result, you should know that University faculty and staff members are required to report any instances of sexual misconduct, including harassment and sexual violence, to the University's Title IX office so that the victim may be provided appropriate resources and support options. What this means is that as your professor, I am required to report any incidents of sexual misconduct that are directly reported to me, or of which I am somehow made aware.

There are two important exceptions to this requirement about which you should be aware:

A list of the designated University employees who, as counselors and medical professionals, do not have this reporting responsibility and can maintain confidentiality, can be found here.

An important exception to the reporting requirement exists for academic work. Disclosures about sexual misconduct that are shared as 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 are the victim of sexual misconduct, Pitt encourages you to reach out to these resources

- Title IX Office: 412-648-7860
- The University Counseling Center: 412-648-7930 (8:30 A.M. TO 5 P.M. Monday–Friday) and 412-648-7856 (after business hours)

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

Other reporting information is available here.