MOLGEN 5280/PUBH-BIO 5280
Introduction to Genomic Data Analysis (2 credit hours)
7 Week Term Spring 2015 (every year thereafter)

Instructors (50% time each)
Grzegorz A. Rempala, Ph.D. D.Sc. Email: rempala.3@osu.edu
Office: 226 Cunz Hall
Office hours: M/W 12pm–2pm, or by appointment

Helen Chamberlin, Ph.D. Email: chamberlin.27@osu.edu
Office: 903 Biological Sciences Building
Office hours: TBA

Class meeting times
The class meets for a two hour period, two times a week.
Days and times to be determined if the course is approved.

Website
Carmen: http://carmen.osu.edu
Login with OSU internet username (name.#) and password, then go to the appropriate course.
When available, course notes and links to reading material will be posted on the course website prior to each lecture.

Textbook
Introduction to Statistical Learning: with Applications in R. James, Witten, Hastie, Tibshirani. This book is available electronically through OSU libraries. Other materials will be distributed throughout the course via the website as appropriate.

Software
R (http://www.r-project.org/) R is available free online.

Prerequisites
Undergraduate students must be of rank 3 (junior standing) or higher. Undergraduate Calculus (Math 1151 or 1156), a Statistics course at the 2000 or above level (e.g., Stats 2450, 2480, or MolGen 5650) and an Intermediate Biology class (Biology 1113 or MolGen 5660), or graduate standing, or permission of instructor. Some experience with programming is recommended.

Description
The course is aimed at both graduate and advanced undergraduate students.
This course provides an in-depth analysis of a specific question to which genomic methods are applied. The course will intersperse experimental methods and statistical analysis of the biological data.

Learning and assessment activities

Homework (60% of final grade)

There will be 7 homework assignments. Late homework will NOT be accepted without a reasonable and advance notice. Students are permitted (indeed encouraged) to work together on homework, but submitted assignments must be written independently.

Final Exam (20% of final grade)

There will be one in-class written exam. The exam will be closed book and closed notes.

Final Project (20% of final grade)

The final project in this course will account for a significant portion (20%) of the grade and will showcase students understanding of methods in genomic data analysis in their appropriate biological context. Each student will be assigned a student team to write a short project report and project teams should be prepared to give a 15-minute in-class presentation where they describe their proposed project, and get feedback from other class members.

Grading

Grading Scale: 90-100 A, 78-89 B, 66-77 C, 50-65 D, below 50 E. (”+” for exceeding mid-range, no ”-”)

For graduate credit additional work will be required in the homework assignments and exam problems, showcasing the more in-depth mastering of key concepts and statistical skills. The graduate credit projects will be expected to demonstrate the appropriate level of sophistication beyond what would be expected of undergraduate students.

Learning Objectives

Upon successful completion of the course, students will have the knowledge, comprehension and/or skills to be able to apply commonly used statistical analysis methods for genomic data. Particularly, the students will be able to

• perform with software the signal/noise extraction, (including testing for signal), p-value adjustment, dimension reduction and enrichment analysis for general genomic datasets
• proficiently apply the statistical methods for summary, visualization and analysis of data from high-throughput platforms, including DNA-microrray, RNA-seq, ChIP-seq and DNase-seq
• describe the basic principles of genetic data annotation, linkage and storage
• be proficient in using the USC genome browser web tool
• identify the biological context of data collection and experimental design
• interpret the results of a statistical analysis in the context of relevant biological questions.

• describe the chromatin and gene expression factors important in promoting cell differentiation, and in inducing cellular pluripotency
• propose experimental approaches to address unanswered biological questions using genomic methods

Competencies (for College of Public Health)

• Core MPH Competencies

1. Apply descriptive and graphical techniques commonly used to summarize public health data.
2. Distinguish among the different measurement scales and the implications for selection of statistical methods to be used based on these distinctions.
3. Apply descriptive and graphical techniques commonly used to summarize public health data.
4. Describe basic concepts of probability, random variation and commonly used statistical probability distributions.
5. Describe alternatives to standard statistical methods when assumptions are not met.
6. Interpret results of statistical analyses found in public health studies.
7. Work in an interdisciplinary setting.

• Core MPH in Biostatistics Competencies

1. Conduct statistical procedures and data analysis methods appropriate for analyzing data obtained from health-related research studies.
2. Use at least one major statistical data analysis package (R).
3. Apply appropriate statistical techniques for analyzing public health-related data with specific characteristics, including:

    a. Continuous data
    b. Categorical data
    c. Time-to event data
    d. Repeated measurements data
    e. Clustered data

• Core MS Competencies

1. Read the scientific literature in the field and critique the methods and results.

2. Conduct a brief literature review to evaluate the state of the science regarding a specific topic in the student's area of interest.

• Core PHD Competencies

1. Conduct a thorough literature review to summarize and evaluate the state of the science regarding a new topic in the student's general area.

2. Outline a study to address one of the relevant scientific questions in the field of interest using the appropriate research design.

3. Prepare and deliver lectures or other appropriate class sessions in the student's area of expertise.

4. Demonstrate advanced knowledge in at least one area of subspecialty within the discipline of specialization.

Class Schedule

Date	Topic
W1	Bio: An overview of transcriptional gene regulation Reading: Brown, 2002 http://www.ncbi.nlm.nih.gov/books/NBK21121/
W1	Stat: an introduction to R and bioconductor Reading: ISLAR , Chapter 2
W2	Bio: Genomic analysis of gene expression: methods and experimental approaches
W2	Stat: retrieving data, quality control, and summary statistics Reading: ISLAR, Chapter 5 Sec 3, and selected fragments of http://master.bioconductor.org/help/course-materials/2007/seattle_bioc_intro_nov_07/affy_a_to_z/affyNov07.pdf
W3	Bio: Multicellular organisms and the relationship between cellular specialization and gene expression Reading: Sha and Boyer, 2009 http://www.ncbi.nlm.nih.gov/books/NBK27041/
W3	Stat: analysis of gene expression microarray data: affy, lumi, limma packages Reading: selected parts of http://www.bioconductor.org/help/workflows/arrays/
W4	Bio: Transcriptional regulation in the differentiation process Reading: Lee et al., 2013, Nature, 503, p 360
W4	Stat: analysis of RNA seq data: easyRNAseq Reading:http://www.bioconductor.org/packages/release/bioc/vignettes/easyRNASeq/inst/doc/easyRNASeq.pdf
W5	Bio: The relationship between chromatin structure and gene transcription Reading: Zhu et al., 2013 Cell 152, p 642
W5	Stat: analysis of ChIP seq data Reading: http://www.bioconductor.org/help/course-materials/2011/RNASeqChIPSeq/Lectures/Marioni_ChIPseq_course.pdf
W6	Bio: Reversing the trend: inducing pluripotency in differentiated cells Reading: Soufi et al., 2012, Cell 151, p 994
W6	Stat: DNAse-seq and footprinting analysis http://www.bioconductor.org/packages/release/bioc/vignettes/DNaseR/inst/doc/DNaseR.pdf
W7	Team project presentations
W7	Stat: Combining data from different sources: Rcade Reading: http://www.bioconductor.org/packages/release/bioc/vignettes/Rcade/i

	nst/doc/Rcade.pdf
Final exam	
Written project due date	

Office for Disability Services

Any student who feels s/he may need an accommodation based on the impact of a disability should contact me privately to discuss your specific needs. Please contact the Office for Disability Services at 614-292-3307 in room 150 Pomerene Hall to coordinate reasonable accommodations for students with documented disabilities.

Student Support

Students experiencing personal problems or situational crises during the semester are en- couraged to contact OSU Counseling and Consultation Services (292-5766; http://www.ccs.ohio-state.edu) for assistance, support and advocacy. This service is free to students and is confidential.

Academic Integrity

The Ohio State University, the College of Public Health, and the Committee on Academic Misconduct (COAM) expect that all students have read and understood the University's Code of Student Conduct and the College's Student Handbook, and that all students will complete all academic and scholarly assignments with fairness and honesty.

Academic integrity is essential to maintaining an environment that fosters excellence in teaching, research, and other educational and scholarly activities. Thus, The Ohio State University, the College of Public Health, and the Committee on Academic Misconduct (COAM) expect that all students have read and understood the Universitys Code of Student Conduct and the Colleges Student Handbook, and that all students will complete all academic and scholarly assignments with fairness and honesty. The Code of Student Conduct and other information on academic integrity and academic misconduct can be found at the COAM web pages (http://oaa.osu.edu/coam/home.html). Students must recognize that failure to follow the rules and guidelines established in the Universitys Code of Student Conduct, the Student Handbook, and in the syllabi for their courses may constitute "Academic Misconduct". The Ohio State Universitys Code of Student Conduct (Section 3335-23-04) defines academic misconduct as: "Any activity that tends to compromise the academic integrity of the University, or subvert the educational process". Examples of academic misconduct include (but are not limited to) plagiarism, collusion (unauthorized collaboration), copying the work of another student, and possession of unauthorized materials during an examination. Please note that the use of material from the Internet without appropriate acknowledgement and complete citation is plagiarism just as it would be if the source were printed material. Further examples are found in the Student Handbook. Ignorance of the Code of Student Conduct and the Student Handbook is never considered an excuse for

academic misconduct. If we suspect a student of academic misconduct in a course, we are obligated by University Rules to report these suspicions to the University Committee on Academic Misconduct. If COAM determines that the student has violated the University's Code of Student Conduct (i.e., committed academic misconduct), the sanctions for the misconduct could include a failing grade in the course and suspension or dismissal from the University. If you have any questions about the above policy or what constitutes academic misconduct in this course, please contact me.