**STAT115 / 215 / BIO / BIST 282**

**Introduction to Computational Biology and Bioinformatics**

**Lectures:** Tu/Th 11:30am – 1:00pm, Sever Hall 203, Cambridge, MA

**Computer Labs:**

* Tuesdays Kresge LL6 5:30-7:00
* Thursdays Northwest B129 6:00-8:00
* Thursday 2/15 lab rescheduled to Wed 2/14 Northwest B129, 6:00-8:00

**Instructor:**

X. Shirley Liu 617-632-2472 [xsliu.dfci@gmail.com](mailto:xsliu.dfci@gmail.com)

**Office hours:**

Tue / Thu 1-2:30pm Science Center 300a

**Teaching fellows:**

Matt Ploenzke [ploenzke@g.harvard.edu](mailto:ploenzke@g.harvard.edu)

Andy Shi [andyshi@g.harvard.edu](mailto:andyshi@g.harvard.edu)

**Course website:**

<https://canvas.harvard.edu/courses/39391>

**Grading:**

HW: 90 points (HW0 is 10 points; HW1-5 are 15 points each for STAT115 and 20 points each for graduate level)

Quizzes: 3 points; Final: 20 points

Students taking the graduate level of the course will be asked to perform additional reading, analysis exploration, or coding in every homework and make a final algorithm video in addition to the STAT115 requirement. Students taking the graduate version will be graded at a different curve from those taking STAT115. Graduate students and postdocs can take STAT115 if it satisfies your credit requirement, and undergraduate students can also take the graduate level course if you like to do the additional work.

We will teach programming in R and python in this class. Each homework will consist of problems to be solved by hand, running some algorithms on the computer to obtain results, and doing some coding (in python or R). Students could discuss course material with each other and with the TA, but the submitted work must be your own work.

Homework is due at midnight on specific Sundays at the end of the week. All HW should be submitted through the Harvard Canvas web interface. A total of 6 late days is granted, which you can use towards any HW, although any single HW could not be late for more than 3 days. After the 6 days are used up, 10% will be deducted from each additional late day on a HW, respectively.

**Lecture schedules:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Week | Tue Lecture | Labs | Thur Lecture | Sun HW Due |
| 1 (1/22) | Introduction to computational biology | Git, Odyssey, LINUX, R & Bioconductor | Microarray normalization,  Expression index |  |
| 2  (1/29) | Differential expression, False discovery rate | Python, RMA, LIMMA | Gene Ontology,  Microarray clustering,  Batch effect | HW1 (Basics) |
| 3  (2/5) | Dimension reduction, Classification | DAVID, Clustering, COMBAT | NGS,  Read mapping |  |
| 4 (2/12) | RNA-seq analysis | PCA, SVM, BWA | Single-cell RNA-seq | HW2  (Microarrays) |
| 5 (2/19) | Flexible | Salmon, DESeq2, Python 2 | Motif finding |  |
| 6  (2/26) | ChIP-seq | scRNA-seq | Integrating ChIP-seq with gene expression | HW3  (RNA-seq, classification) |
| 7  (3/5) | Epigenetics,  DNA methylation | MACS2, UCSC, BedTools, Python 3 | Histone modifications,  Chromatin accessibility |  |
| 8 (3/12) | Spring Recess |  | Spring Recess |  |
| 9 (3/19) | Hidden Markov Model | HMM, Python 4 | Long range chromatin interactions | HW4  (scRNA-seq, ChIP-seq) |
| 10 (3/26) | GWAS interpretation using epigenetics | BETA, Python 5 | Genome sequencing, TCGA |  |
| 11 (4/2) | Targeted therapy and drug resistance | HaploReg, RegulomeDB | High throughput screening | HW5 (Epigenetics and integration) |
| 12 (4/9) | Cancer immunotherapy | TCGA data access,  Survival analysis | Epigenetic therapy, Combination therapy |  |
| 13 (4/16) | Reproducible research  (video) | Flexible | Flexible | HW6 (Cancer Genomics) |
| 14 (4/23) | Final review |  | Reading Period |  |
| 15 (4/30) | Reading Period | Screencast video due |  |  |