SYBB 311/411B: Data Integration in Bioinformatics

Course Syllabus

Instructor: Gürkan Bebek, Ph.D.
Office: BRB 921
Phone: 216-368-4541
E-Mail: gurkan.bebek@case.edu
Website: http://gurkan.case.edu/teaching.html
Office Hours: by appointment only: Thu, 2:00-3:00 PM

Term: Fall 2015
Class Meeting Days: Sept. 24th - Oct 27th
Class Meeting Hours: Tue, Thu 10:00-11:15 AM
Class Location: Thwing 101

Course Overview

Surveys of Bioinformatics
The SYBB Survey Series is composed of the following course sequence: (A) Technologies in Bioinformatics, (B) Data Integration in Bioinformatics, (C) Translational Bioinformatics, and (D) Programming for Bioinformatics. Each standalone section of this course series introduces students to an aspect of a bioinformatics project - from data collection (SYBB 311/411A), to data integration (SYBB 311/411B), to research applications (SYBB 311/411C), with a fourth module (SYBB 311/411D) introducing basic bioinformatics programming skills. Each semester SYBB 311/411A, SYBB 311/411B, and SYBB 311/411D offered as blocks meeting back to back. SYBB 311/411C is a longitudinal class throughout the semester. Each class is graded separately.

Course Description
SYBB 311/411B is a five-week course that surveys the conceptual models and tools used to analyze and interpret data collected by high-throughput technologies, providing an entry point for students new to the field of bioinformatics. The topics covered include biomedical ontologies, tools for genome/proteome exploration and analysis; large databases used for housing signaling pathways and interaction networks.

Course Objectives

<table>
<thead>
<tr>
<th>Big Idea</th>
<th>Enduring Understandings</th>
<th>Learning Outcomes</th>
</tr>
</thead>
</table>
| **Data Integration in Bioinformatics** | Understand Bioinformatics tools and models used for analyzing large datasets | • Can identify the appropriate tool, algorithm or method for bioinformatics analyses  
• Can assess tools provided for bioinformatics analysis  
• Can utilize tools for genome exploration and analysis. |
| | Interpret data collected by high-throughput technologies | • Can summarize and represent datasets utilizing bioinformatics tools  
• Can organize biomedical knowledge with ontologies  
• Can utilize biomedical databases to augment and integrate with -omic measurements |
Course Prerequisites

Undergraduate  
**Prerequisites:** BIOL 214 AND 215; OR 250  
**Co-requisites:** SYBB311A, SYBB311B and SYBB311D

Graduate  
**Prerequisites:** Graduate standing OR Prerequisites not met permission  
Graduate students have the option of enrolling in all four courses or choosing the individual modules most relevant to their background and goals with the exception of SYBB411D, which must be taken with SYBB411A. Based on experience in the field, instructor can waive this co-requisite.

Required Texts and Materials

**Online Textbook(s)***  
This is an e-book that is free to access. Please make sure you can access to this resource.

**Classroom Response Clickers:**  
We will be using iClicker+ in class on a regular basis. You will need to purchase an iClicker+ pad (commonly called a “clicker”) from the bookstore or online and bring it with you to every class session. It would be wise to bring extra batteries as well, as we will be using the pads in activities that count for class points. The purchase of a clicker is required (NOT optional). It will be used as an integral part of this course. I will provide a short demonstration of how to use it in class. [Note: the clicker can be used in other classes]. After you purchase your clicker, you must [register your clicker](http://blackboard.case.edu) on the course Blackboard site before the first class. It is imperative that every student registers their unit before the first week of class. Instructions for the registration process can be found on Blackboard. Make sure you buy the clicker that looks like the image on the right.

**Course Website:**  
The course website is available on Blackboard. The URL is: [http://blackboard.case.edu](http://blackboard.case.edu). Students will need their CWRU network ID and password to access the site. Important course information and announcements will be posted on Blackboard including the course syllabus, assignments, and grades. The students are responsible to refer to this site regularly. Additional information on the SYBB 311/411 Series can be found on Instructors web page [http://gurkan.case.edu/teaching.html](http://gurkan.case.edu/teaching.html).

**Course format**  
This is an active-learning class. The focus is on understanding and critically evaluating the current bioinformatics methods and approaches in diverse biomedical fields. Readings and in class exercises will be assigned on a weekly basis. Students are expected to complete these assignments before coming to class.  
Class participation, weekly quizzes (via iClicker+) and homework will be the primary mode of assessing students’ grasp of the material. Graduate students are also required to review the literature in one of the bioinformatics topics discussed in class and turn in a short paper within two days of last lecture.  
In class exercises can be assigned before class. These will be shared on blackboard. Students should follow the directions, and provide written evidence of completion via Blackboard when asked.
## Topic Schedule

<table>
<thead>
<tr>
<th>Lecture</th>
<th>Lecture Topic</th>
<th>Readings</th>
</tr>
</thead>
</table>
| 1       | Bioinformatics databases, I  
| 2       | Bioinformatics databases, II  
- Discuss ontologies for organizing knowledge  
- ID Mapping & summary measures | 1. * IntAct: Molecular Interactions at the EBI. [http://www.ebi.ac.uk/training/online/course/intact-molecular-interactions-ebi](http://www.ebi.ac.uk/training/online/course/intact-molecular-interactions-ebi)  
| 3       | Genomic analysis, I  
- Genomic data analysis  
- Genome browser  
3. [http://homer.salk.edu/homer/motif/](http://homer.salk.edu/homer/motif/) |
| 4       | Genomic analysis, II  
- Genomic structural variation  
| 5       | Pattern recognition  
- Regression-based models  
- High-dimensional statistics  
| 6       | Computational Biology  
- Algorithms in bioinformatics  
- Public software tools | 1. [http://meme.nbcr.net/meme](http://meme.nbcr.net/meme) |
| 7       | - Network-view of molecular biology  
- Constructing biological networks  
| 8       | Ontologic and Set Analysis  
- Analysis of mRNA expression  
- Gene set analyses  

## Homework

The homework schedule is shown in the table below. After the students complete the assigned tasks,
they will be required to turn in the results of their task/analysis, as well as a short description and analysis of their findings. The instructor will make the content of the written portion of each homework explicit in class. Assignments will be due in a week at the beginning of class.

<table>
<thead>
<tr>
<th>Assigned Lecture</th>
<th>Assignment</th>
<th>Evaluation (UG)</th>
<th>Evaluation (G)</th>
</tr>
</thead>
</table>
| 1                | 1. Hand out list of genes  
  2. Use specific databases to generate hypotheses regarding:  
     protein interactions, transcriptional regulation             | 20 pts          | 20 pts         |
| 3                | 1. Hand out dataset of sequencing experiment  
  2. Map reads onto genome using genome browser tools  
  3. Use mapped reads to make inference about structural variation  | 20 pts          | 20 pts         |
| 5                | 1. Run a hierarchical clustering on a set of gene expression data  
  2. Use Cluster 3.0  
  3. Use http://meme.nbcr.net/meme                             | 20 pts          | 20 pts         |
| 7                | 1. Extract protein interactions from IntAct  
  2. Visualize gene list interactions in Cytoscape  
  3. Use clustering algorithm to identify topologic groups   | 20 pts          | 20 pts         |
| 9                | 1. Use ontologic enrichment tools to generate hypotheses about common functionality amongst candidate genes  
  2. Identify significant gene functions of the network  
  3. Estimate expression values of genes based on read frequency  
  4. Use expression values to calculate gene set enrichment scores | 20 pts          | 20 pts         |

**Final Paper (Graduate students only)**

Graduate students will be expected to write a 3-4 page, double spaced article reviewing the literature in one of the bioinformatics topics discussed in class. The purpose of the paper is to (1) provide an overview of a subfield of bioinformatics and (2) identify potential avenues for research, which will help students formulate dissertation projects.

These papers must be submitted on Blackboard using SafeAssign. If an issue arises, email the paper to Dr. Bebek before the deadline. Late papers will be penalized a full grade or more depending upon the time until the paper is submitted. The papers are due two days after the last lecture.

**Grading**

Grading will be based primarily of the homework assignments, of which there will be 2 per week. An additional piece of your grade will come from participation in Blackboard and in-class discussions. Though attendance is not enforced, class participation will be evaluated as follows:

<table>
<thead>
<tr>
<th>Assignments</th>
<th>Undergraduate %</th>
<th>Graduate %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework</td>
<td>80</td>
<td>65</td>
</tr>
<tr>
<td>In-class participation</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>Final paper</td>
<td>0</td>
<td>25</td>
</tr>
</tbody>
</table>

Final letter grades will correspond to the percentage of points you earn out of the possible 100 point.
The tentative scale is $\geq 90\% = A$, 80-89.4\% = B, etc. However, these ranges may be adjusted at the discretion of the instructor.

**Make-up Policy:**
Schedule of classes are listed on this document. Requests for make-up will only be considered in the event of an emergency that can be verified (e.g., a doctor’s note must be provided & possible verification from Undergraduate Studies) or a pre-scheduled academic conflict (presenting at a seminar series, presenting at a conference (Students have to notify instructor at least 2 weeks prior to this conflict).

If a student misses final paper submission due to an illness or other emergency, students are responsible to contact right away to report their emergency (48 hours). If the student fails to do this, they will not be granted a make-up and they will earn a zero for the paper.

**Disabilities**
If you have a disability and anticipate needing accommodations for this course, we are willing to work with you and Disability Resources (http://students.case.edu/education/disability) to help provide the accommodations you need.

**Computers**
Computer with Internet access is required for homework assignments. In class use of computers area allowed as long as students stay on task (no Facebook, texting, etc.). Any misuse will be penalized.

**Academic Integrity**
No form of academic dishonesty including cheating, plagiarism, misrepresentation, or obstruction will be tolerated in this class. Plagiarism in any form will result in a failing grade. Please see: http://studentaffairs.case.edu/office/judicial

Students are expected to abide by the academic integrity policy of the university, which can be found at: http://studentaffairs.case.edu/handbook/policy/integrity.html. It is the student’s responsibility to read and familiarize him/herself with the academic integrity policy.

Evidence of academic misconduct will be reported to the Dean and could result in a judicial action. If a student is caught cheating, they may fail the assignment or the entire course depending on the severity of the offense. Examples of academic integrity violations include:

1. Possession of a cell phone, calculator or tablet during an exam.
2. Copying from a student’s exam.
3. Lying about an illness or other emergency to get out of taking an exam.
4. Participating in class with another student’s clicker.