GCB537 Advanced Computational Biology

Term: Spring 2013
Instructor: Li-San Wang

Objectives:
1. Review important concepts for computer science and computational biology
2. Discuss classic papers on important topics in genomics and computational biology
3. Students learn to evaluate and criticize research papers in genomics and computational biology
4. Students learn to survey and summarize literature in the final presentation projects

Requirement: Background in statistics, biology, genetics and genomics, and computer science. This is NOT a course for a bioinformatics lab. Please contact the instructor if you are not a GCB graduate student.

Time and Location: Tuesdays and Thursdays 1:30-3:00pm, Blockley Hall 1311

Schedule:

<table>
<thead>
<tr>
<th>Date</th>
<th>Type</th>
<th>Topic</th>
<th>Date</th>
<th>Type</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>8-Jan</td>
<td>Winter break</td>
<td>No class</td>
<td>10-Jan</td>
<td>Organizational meeting</td>
<td>Organizational meeting</td>
</tr>
<tr>
<td>15-Jan</td>
<td>Lecture</td>
<td>Review on data structures: list, tree, stack, queue, hash table; time complexity</td>
<td>17-Jan</td>
<td>Lecture</td>
<td>Review on Time complexity and NP completeness</td>
</tr>
<tr>
<td>22-Jan</td>
<td>Lecture</td>
<td>Review on Basic algorithms (Spanning tree and shortest path; graph traversal)</td>
<td>24-Jan</td>
<td>Lecture</td>
<td>Review on Sequence alignment (Dynamic Programming; Smith-Waterman algorithm; BLAST)</td>
</tr>
<tr>
<td>29-Jan</td>
<td>Paper discussion</td>
<td>NGS read mapping and variant calling</td>
<td>31-Jan</td>
<td>Paper discussion</td>
<td>RNA-seq</td>
</tr>
<tr>
<td>5-Feb</td>
<td>Paper discussion</td>
<td>DNA-seq</td>
<td>7-Feb</td>
<td>Lecture</td>
<td>Review on Phylogenetics</td>
</tr>
<tr>
<td>12-Feb</td>
<td>Paper discussion</td>
<td>Phylogenetics</td>
<td>14-Feb</td>
<td>Paper discussion</td>
<td>PhyloHMM</td>
</tr>
<tr>
<td>19-Feb</td>
<td>Lecture</td>
<td>Review on Gene expression regulation; motif discovery</td>
<td>21-Feb</td>
<td>Paper discussion</td>
<td>Motif discovery</td>
</tr>
<tr>
<td>26-Feb</td>
<td>Paper discussion</td>
<td>ChIP-seq</td>
<td>28-Feb</td>
<td>Paper discussion</td>
<td>Modeling gene expression regulation</td>
</tr>
<tr>
<td>5-Mar</td>
<td>Spring break</td>
<td></td>
<td>7-Mar</td>
<td>Spring break</td>
<td></td>
</tr>
<tr>
<td>19-Mar</td>
<td>Paper discussion</td>
<td>Microarray analysis</td>
<td>21-Mar</td>
<td>Paper discussion</td>
<td>Biomarker discovery</td>
</tr>
<tr>
<td>2-Apr</td>
<td>Paper discussion</td>
<td>CNV analysis</td>
<td>4-Apr</td>
<td>Paper discussion</td>
<td>Rare variant discovery</td>
</tr>
<tr>
<td>9-Apr</td>
<td>Presentation</td>
<td>Student presentation</td>
<td>11-Apr</td>
<td>Presentation</td>
<td>Student presentation</td>
</tr>
<tr>
<td>16-Apr</td>
<td>Presentation</td>
<td>Student presentation</td>
<td>18-Apr</td>
<td>Presentation</td>
<td>Student presentation</td>
</tr>
<tr>
<td>23-Apr</td>
<td>Presentation</td>
<td>Student presentation</td>
<td>25-Apr</td>
<td>Presentation</td>
<td>Student presentation</td>
</tr>
</tbody>
</table>
Course format:

Some classes are lectures to help students review important concepts in computer science and computational biology. In paper discussion classes, papers are selected to cover a range of important topics, with emphasis on algorithm design and best practice for data interpretation and presentation. Towards the end of the semester students will team up and present a topic they select.

For paper discussion courses, two students will be in charge of presenting the background and prior arts, and leading the discussion for each class, including whiteboard presentations and discussions. To ensure the quality of the presentation, the students will send powerpoint files to the instructor or discuss with the instructor at least two days before the scheduled presentation. After the class, all other students at the presentation will send a grade (between 1=unprepared and 5=excellent) or any constructive comments to the instructor by email; comments will be forwarded to the presenters anonymously.

Each student will select three papers/paper sets and hand in a 1 page report (Times New Roman, 12pt, single-spaced, 1” margin) for each paper before the discussion class that cover three areas of assessment: Impact, Innovation, Shortcoming.

Term project -- literature review:

The students will team up (two students per team) and:
(a) Select a specific topic on bioinformatics and computational biology by the end of February. In addition to the areas we cover in this course, possible topics include but are not limited to: protein/RNA folding, proteomics, text mining, systems biology, epigenetics, etc.
(b) Check with the instructors the list of papers they intends to summarize by the end of March.
(c) Make a 50-minute presentation in April.

Grading: 30% paper presentation (20% graded by other students and 10% graded by instructor), 20% participation in paper discussion during the class, 20% written critiques, 30% term presentation.

Paper List:

Next generation sequencing technology

Read mapping and variant calling


**RNA-seq**

Anders and Huber, Differential expression analysis for sequence count data. Genome Biology 11:R106, 2010. (DESeq)


**DNA-seq**


**Phylogenetics**

**Phylogeny reconstruction**


**PhyloHMM**

Adam Siepel & David Hausseler, Computational Identification of Evolutionarily Conserved Exons, RECOMB 2004. (PhastCons)

**Gene expression regulation**

**Motif discovery**


**ChIP-seq**

Modeling gene expression regulation


Biomarker discovery

Microarray analysis

Tibshirani et al., Diagnosis of multiple cancer types by shrunken centroids of gene expression. PNAS 99:6567-72, 2002. (PAM)


Biomarker discovery


Genome-Wide Association Studies

Population stratification correction


CNV Analysis

Wang K et al., PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Research, 17:1665-1674, 2007. (PennCNV)

Rare variant discovery

Wu et al., Rare Variant Association Testing for Sequencing Data Using the Sequence Kernel Association Test (SKAT). American Journal of Human Genetics, 89:82-93, 2011. (SKAT)