### Course Description

Over the past decade, the emergence and rapid advance of genomic and proteomic technologies have generated copious amounts of data and Web-based bioinformatics resources. Use of bioinformatics databases and software tools are the key for success in both basic science and translational research as they guide researchers in formulating new hypothesis, designing studies to test their hypothesis, and interpreting and validating their experimental results.

This course will introduce students to real-life examples of the use of bioinformatics tools in research as well as teach efficient identification of appropriate bioinformatics database/software and effective application of the identified resource(s) in order to solve scientific questions. It is offered in a computer classroom, thereby providing hands-on training for solving bioinformatics queries using Web-based tools. Homework will be assigned to reinforce the concepts presented during each class. A final project will promote integration of the bioinformatics resources introduced throughout the course.

### Grading

Performance will be evaluated by homework (50%), class participation (20%) and a final project (30%).

**Final Project**

At the beginning of the course each student will receive a short peptide sequence. Starting from the peptide sequence students will identify its corresponding gene and protein sequence and will write a short research paper on the identified gene/protein. Research papers should contain (1) a review of the literature and (2) a proposal describing a testable hypothesis generated by the use of bioinformatics databases and software.

### Target Audience

designed for graduate students in basic, clinical, and translational science programs.

### Goals

1. Discover the latest innovations in literature searching
2. Explore gene, genome and protein centric databases
3. Learn DNA & protein sequence analysis and similarity searching
4. Understand genetic variations and prediction of their functional consequences
5. Identify common biological functions present in a large set of genes
### Session 1.1 (3/1) Introduction and Overview of HSLS Resources

**Chattopadhyay**

**Topics:**
- Course expectations and explanation of resources and services provided by the HSLS Molecular Biology Information Service.

### Session 1.2 (3/1) Literature Informatics

**Iwema**

**Topics:**
- Medline search: NCBI PubMed, MeSH database
- 2nd generation tools: GoPubMed, Novoseek, DeepDyve
- NIH grant proposal search: RePORTER, Novoseek
- Text similarity search tool: eTBLAST
- Open access publishing: PubMed Central
- Journal impact factors, publication citation index, author H index: ISI Web of Knowledge, Scopus

**Recommended Reading:**

### Session 2 (3/3) Genome Biology

**Chattopadhyay**

**Topics:**
- Organism whole genome sequence database: Entrez genome, Integrated Microbial Genome, Viral genome
- Genome Browsers: NCBI Mapviewer, UCSC Genome Browser

**Recommended Reading:**
2. Using the NCBI Map Viewer to Browse Genomic Sequence Data, Tyra G. Wolfsberg *Current Protocols in Bioinformatics*, Unit 1.5
### Session 3.1 (3/8) Bioinformatics Information Retrieval & Gene-Centric Knowledgebases

**Topics:**
- Search engine: Search.HSLS.MolBio
- NCBI resources: Entrez Gene, OMIM
- HSLS Licensed Resources: Ingenuity IPA, GeneGo Metacore

**Recommended Reading:**

### Session 3.2 (3/8) Gene Regulation

**Topics:**
- Promoter Sequence Analysis: Biobase Knowledge Library, Transfac

**Recommended Reading:**

### Session 4.1 (3/10) Protein-Centric Knowledgebases

**Topics:**
- Protein Database: UniProt, Human Protein Reference Database,
- Motifs and Domains: Pfam, Prosite, NCBI Conserved Domain Database
- Protein Interaction Database: STRING

**Recommended Reading:**
### Session 4.2 (3/10) Protein Structure

**Topics:**
- Databases: Protein Data Bank, Molecular Modeling DataBase
- 3D Structure Viewer: Cn3D, FirstGlance in Jmol

**Recommended Reading:**

### Session 5.1 (3/15) Sequence Similarity Searching

**Topics:**
- NBCI BLAST, PSI BLAST, PHI BLAST, UCSC BLAT

**Recommended Reading:**

### Session 5.2 (3/15) Sequence Analysis

**Topics:**
- PCR Primer Design: Primer3, VectorNTI
- Multiple Sequence Alignment: ClustalW

**Recommended Reading:**
<table>
<thead>
<tr>
<th>Session 6 (3/17)</th>
<th>Genetic Variations</th>
<th>Chattopadhyay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topics:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Databases: dbSNP, Hapmap, Human Mutation Database, dbGAP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• SNP functional analysis: f-SNP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recommended Reading:</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Session 7 (3/22)</th>
<th>Omics Resources</th>
<th>Chattopadhyay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topics:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Gene Expression Database: NCBI GEO, Oncomine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Pathway Informatics: Ingenuity IPA, NIH DAVID</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recommended Reading:</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Session 8 (3/24)</th>
<th>Final Project Presentations</th>
<th>Class Participants</th>
</tr>
</thead>
</table>