Course Title: Bioinformatics, 3 credits, 14 weeks + exam
Meeting: 3-4:30 PM Tues, 10-11:30 AM Wed, Translational Res. Bldg 227 E 30th St, Rm 718
Instructor: Stuart Brown, Associate Professor NYUMC
Contact: stuart.brown@nyumc.org
212-263-7689

Prerequisites: Graduate level courses in genetics, cellular and molecular biology, biostatistics

Textbook: Understanding Bioinformatics (Zvelebil and Baum)
Additional weekly readings (see lecture schedule)

Course overview: This course is intended for any biomedical scientist with a need to sharpen bioinformatics skills and understand the algorithms that are used in current methods. It provides practical hand-on training with an emphasis on data structures and algorithms for the most commonly used molecular biology and genomics applications including sequence alignment, database searching, DNA and protein motifs, phylogenetics, and gene expression.

Weekly readings will include book chapters, key journal articles, and software tutorials. Weekly exercises will utilize basic computing skills in Linux, Python, R, and the use of both command line and web-based bioinformatics tools.

Course Format, Requirements, Assignments

Course schedule: The course will meet once per week for 1.5 hours of lecture and an additional 1.5 hour meeting for a computer practical session.

Readings and participation: Students are required to attend class, to complete reading and hands-on assignments and to participate in discussions and engage in healthy exchange of ideas. Students will require a personal laptop computer in class for interactive laboratory sessions to complete tutorials and exercises. Weekly assignments will be introduced during lectures. Hands-on work will begin during the lab session, and will be completed on the students own time. Assignments will be submitted and graded each week. Weekly readings should be completed before lecture to allow for clarification and in depth discussion. One problem set will be graded as a midterm and one objective exam (or project, TBD) will be held that will focus on concepts, skills, and readings.

Exam
There will be one exam in this class. It will primarily focus on practical skills.

General Policies
Grading: class participation (20%), weekly assignments (40%), final (40%)

Late/missed work
You must adhere to the due dates for all required submissions. If you miss a deadline, then you will not get credit for that assignment/post. Try to avoid last minute submissions.
I n c o m p l e t e s
No “Incompletes” will be assigned for this course unless we are at the very end of the course and you have an emergency.

R e s p o n d i n g t o M e s s a g e s
I will check e-mails daily during the week, and I will respond to course related questions within 48 hours.

A n n o u n c e m e n t s
I will make announcements throughout the semester, so make sure you check updated messages in the 'Announcement' section. Meeting locations are subject to change. In most cases, I will also email the announcement to you. Make sure that your email address is updated; otherwise you may miss important emails from me.

S a f e g u a r d s
Always back up your work on a safe place (electronic file with a backup is recommended) and make a hard copy. Do not wait for the last minute to do your work. Allow time for deadlines.

P l a g i a r i s m
Plagiarism, the presentation of someone else's words or ideas as your own, is a serious offense and will not be tolerated in this class. The first time you plagiarize someone else's work, you will receive a zero for that assignment. The second time you plagiarize, you will fail the course with a notation of academic dishonesty on your official record.

M i s s e d E x a m s a n d G r a d e A p p e a l s
Make-up examinations (for final only) will be given under special circumstances. If a make-up exam is permitted, a different exam will be written for that student and may have a different format than the regular examination. The projects must be turned in on time and no late projects will be accepted.

If you believe that there is a mistake in grading of an assignment/exam, you may appeal your grade within a week after you receive your grade. You must write a note describing the error, attach it to the original exam, and give it to me within a week of the return of your exam. I will review your argument and my initial grading, and then return your exam with a decision to you in a timely manner.
Weekly Course Schedule

Lecture 1: Bioinformatics Algorithms & Data Structures (Tues, Sept 6, TRB 718)
- Overview of the challenges of Bioinformatics in the era of Genomics
- What is an algorithm
- What is a data structure, how is genomic data structured
- "Data Skills"
Reading: Why Biologists Want to Program Computers, by James Tisdal

Practical Session 1: Linux basics and Intro to cluster computing (Wed, Sept 7, TRB 718)
- Intro to the Linux shell, basic UNIX commands, shell scripting
- Emacs Guide, tutorial
- Intro to the HPC cluster, modules
- SGE scripts, data transfer

Lecture 2: Bioinformatics Databases, BioPython (Tues, Sept 13, TRB 718)
- Using ENTREZ to find and retrieve sequences and MEDLINE references
- Genome databases
- Other online genetic databases: mutations, SNPs, Pathways, etc
Reading: The NCBI Handbook: Genbank
Entrez Tutorial

Practical Session 2: Database skills, BioPython tutorial (Wed, Sept 14, TRB 718)
- Using the browser URL and the command line to access databases directly
- BioPython for database access and parsing sequence records
Reading: Python for Biologists: A complete programming course for beginners by Martin L. B. J. N. E. S.
http://pythonforbiologists.com/

Lecture 3: Alignment (Tues, Sept 20, TRB 718)
- Pairwise similarity (manual, dot plots)
- Pairwise alignment algorithms: Needleman-Wunch, Smith-Waterman
- Dynamic programming
- Scoring systems, optimization
- Gap penalties

Practical Session 3: Pairwise Alignment exercises (Wed, Sept 21, TRB 718)

Lecture 4: Similarity Searching of Sequence Databases (BLAST) (Tues, Sept 27, TRB 718)
- Using alignment for database searching
- Word/seed/hash based methods:
- FASTA alignment algorithm
- BLAST alignment algorithm
- BLAST2 alignment algorithm (gapped BLAST)
- Using BLAST to search databases
Reading:
Altshul et al, (1990) Basic local alignment search tool
Korf, BLAST (chap 4).
Practical Session 4: BLAST exercises (Wed, Sept 28, TRB 718)
- Customizing BLAST searches
- Interpreting BLAST output (the deeper meanings in sequence alignments)
- Building custom BLAST databases
- Making SGE scripts and submitting cluster jobs

Lecture 5: Multiple Alignment (Tues, Oct 4, TRB 718)
- Progressive Pairwise Alignment
- Formatting alignments
- Evaluating alignments for sequence conservation
Reading:
  Thompson et al. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment

Practical Session 5: Multiple Alignment exercises (Wed, Oct 5, TRB 718)

Lecture 6: Motifs (Tues, Oct 11, TRB 718)
- Patterns in DNA: promoters, transcription factor binding sites
- Regular expressions
- Patterns in proteins: structural motifs
- Hidden Markov models
- Databases of biological motifs and profile searches
- Evaluating motifs – log odds scores and the use of proper controls
Reading
  - Sonnhammer, Eddy, Durbin (1997), Pfam: a comprehensive database of protein domain families based on seed alignments.
  - Altschul & Koonin (1998) Iterated profile searches with PSI-BLAST

Practical Session 6: Using BioPython for Motif analysis (Wed, Oct 12, TRB 717)
- Working with the PFam database
- Using BioPython for sequence motif analysis
- Find transcription factor binding motifs in ChIP-seq data
- Simulation and significance testing
Reading
  Python for Biologists: A complete programming course for beginners by Martin Jones
  http://pythonforbiologists.com/

Lecture 7: Computing Evolution - Phylogenetic Analysis (Tues, Oct 18, TRB 718)
- From Multiple Alignment to Phylogeny
- Sequence-based taxonomy: Overview and Assumptions
- Distance methods
- Parsimony & Maximum likelihood methods
- Detecting selection in sequences
Reading
  - MEGA tutorial

Practical Session 7: Building Phylogenetic Trees from public data (Wed, Oct 19, TRB 717)
- From Multiple Alignment to Phylogeny
- Sequence-based taxonomy: Overview and Assumptions
Lecture 8: Gene Expression analysis (Tues, Oct 25, TRB 718)

- Microarray technology
- RNA-seq technology
- Normalization in gene expressions studies
- Computing differential expression
- Functional enrichment in gene lists

Reading:
- R tutorial
- Golub et al (1999), Molecular Classification of Cancer

Practical Session 8: Intro to R, Using R for microarray analysis (Wed Oct 26, TRB 718)

Lecture 9: Metagenomics (Tues, Nov 1, TRB 718)

- Microbiology
- PCR assays & 16S rDNA sequencing
- Taxonomic assignment
- OTU clustering
- Species abundance & Diversity measures
- Computing Differences between samples
- Shotgun Metagenomic data (WGS)

Practical Session 9: Qiime tutorial (Wed Nov 2, TRB 718)

Lecture 10: Sequence Variation: SNPs and Haplotypes (Tues, Nov 8, TRB 718)

- What is a SNP
- Genomics assay platforms for sequence variants
- Large scale SNP detection by microarray and by Sequencing
- Haplotypes (HapMap)
- GWAS

Reading
- 1,42 Million SNPs
- User’s Guide to the HapMap
- Hapmap Tutorial
- Comparative Genomics of Regulatory Regions

Practical Session 10: Using Galaxy to analyze SNPs (Wed, Nov 9, TRB 717)

- Galaxy is a web-based toolkit for genome analysis and workflows

Lecture 11. Next Generation Sequencing(Tues, Nov 15, TRB 718)

- High-Throughput sequencing technologies
- Quality control of NGS data
- Genome re-sequencing – find variants
- ChIP-seq – find protein-DNA interactions
- RNA-seq – measure gene expression
- De novo sequencing – study new genomes

Practical Session 11: Working with NGS data (Wed, Nov 16, TRB 717)

- Process NGS data with FastQC
- Visualize Sequence data in IGV and on the UCSC Genome Browser
Thanksgiving Break

**Lecture 12. Chip-Seq & RNA-seq (Tues, Nov 29, TRB 718)**
- Align NGS data to a reference genome
- Use a GTF file for gene annotation
- Count read depth in genome intervals
Reading
  Zhang et al Model-based Analysis of ChIP-Seq
  Computational Analysis of ChIP-seq Data

**Practical Session 11: ChIP-seq practical (Wed, Nov 30, TRB 717)**
- Process ChIP-seq data with Bowtie & MACS

**Lecture 13: Gene Regulation and Epigenetics (Tues, Dec 6, TRB 718)**
- Gene Structure
- Transcription factors and binding sites
- Enhancers
- Chromatin Structure (Histones, DNA methylation, DNase sensitivity, etc)
- ChIP-seq

**Practical Session 13: Use ENCODE data to define gene regulatory modules (Wed, Dec 7, TRB 717)**
- Use BEDTools to find overlaps between TF binding sites, DNase sensitive sites, and modified histone sites

**Project work week – One on one meetings with each student.**

**Final Project Presentations: (Tues, Dec 20, TRB 718)**