**Course Title:** Bioinformatics, 2 credits, 8 weeks + exam  
**TBD:** 1-2:30 PM, Tues & Thurs, Translational Research 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; Sackler Biostatistics & Bioinformatics (BMSC-GA 4451) or equivalent

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

**Course overview:** This course is designed as a followup to the Introductory Biostatistics and Bioinformatics course (BMSC-GA 4451), but is also appropriate for any student with a need to sharpen bioinformatics skills and understand the algorithms that are used in current methods. It provides practical training in bioinformatics for biomedical scientists, 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 Unix, Python, R, and the use of command line, GUI, 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 exercises in 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 will be held that will focus on concepts 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.

**Incompletes**
No “Incompletes” will be assigned for this course unless we are at the very end of the course and you have an emergency.

**Responding to Messages**
I will check e-mails daily during the week, and I will respond to course related questions within 48 hours.

**Announcements**
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.

**Safeguards**
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.

**Plagiarism**
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.

**Missed Exams and Grade Appeals**
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:** Unix for Biologists (*Tues, Oct 27*)
- Intro to the Linux shell, basic UNIX commands, shell scripting
- Reading: 
  - *Why Biologists Want to Program Computers*, by James Tisdal
  - *Emacs Guide*, tutorial

**Practical Session 1:** Intro to cluster computing (*Thurs, Oct 29*)
- Intro to the HPC cluster, modules
- SGE scripts, data transfer

**Lecture 2:** Bioinformatics Algorithms, Data structures, and Databases (*Tues, Nov 3*)
- What is an algorithm
- What is a data structure, how is genomic data structured
- Using ENTREZ to find and retrieve sequences and MEDLINE references
- Genome databases
- Other online genetic databases: mutations, SNPs, etc
- Reading: 
  - *The NCBI Handbook: Genbank*
  - *Entrez Tutorial*

**Practical Session 2:** Using BioPython (*Thurs, Nov 5*)
- Using BioPython to automate access and manipulation of sequences from GenBank
- Intro to Rosalind: learning by problem solving
- Reading: 
  - *Python for Biologists: A complete programming course for beginners* by Martin Jones

**Lecture 3:** Advanced Alignment Methods & Multiple Alignment (*Tues, Nov 10*)
- Translated BLAST, psi-BLAST
- BLAT
- large scale alignment: synteny (MAUVE)
- Progressive pairwise algorithms
- Reading: 
  - Altshul & Koonin (1998) Iterated profile searches with PSI-BLAST
  - Searching Sequence Databases A novel Src kinase in the C.elegans genome
  - Thompson et al. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment

**Practical Session 4:** Find and align insect cytochrome p450 genes. (*Thurs, Nov 12*)

**Lecture 5:** Patterns and Motifs in biological sequences (*Tues, Nov 17*)
- Patterns in DNA: promoters, transcription factor binding sites
- Regular expressions
- Patterns in proteins: structural motifs
- Hidden Markov models
- Databases of biological motifs
- Reading: 
  - Sonnhammer, Eddy, Durbin (1997), Pfam: a comprehensive database of protein domain families based on seed alignments.
  - *Protein structure tutorial with Cn3D*
Practical Session 5: Motif analysis (Thurs, Nov 19)
• Find transcription factor binding motifs in ChIP-seq data
• Simulation and significance testing
• https://www.bioconductor.org/help/workflows/gene-regulation-tfbs/

Lecture 6: Computing Evolution -- Phylogenetic Analysis (Tues, Nov 24)
• From Multiple Alignment to Phylogeny
• Sequence-based taxonomy: Overview and Assumptions
• Distance methods, Parsimony methods, Maximum likelihood methods
• Detecting selection in sequences

Reading:
• MEGA tutorial

Practical Session 6: Seals & Whales, building phylogenies from public data (Thurs, Nov 26)

Lecture 7: Gene Expression (Tues, Dec 1)
• Introduction to the ‘R’ programming language for bio-statistical computing
• Introduction to the microarray platform for gene expression
• Data normalization
• Differential expression: the t-test, non-parametric approaches
• Multiple testing correction
• RNA-seq vs. microarray
• Gene Ontology and functional enrichment analysis

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

Practical Session 7: Use R to find Differential Expression in microarray data (Thurs, Dec 3)

Lecture 8, Sequence Variation: SNPs and Haplotypes (Tues, Dec 10)
• Genomics assay platforms for sequence variants
• Large scale SNP detection, GWAS
• Haplotypes (HapMap)
• Intro to Galaxy platform for genomic analysis

Reading:
• 1.42 Million SNPs
• Users Guide to the HapMap
• Hapmap Tutorial

Practical Session 8: Characterize variants in a cancer and a normal genome (Thurs, Dec 12)
• Galaxy Bushman analysis: https://usegalaxy.org/u/aun1/p/bushman

Lecture 9: Gene Regulation and Epigenetics (Tues, Dec 15)

Practical Session 9: Use ENCODE data to define gene regulatory modules (Thurs, Dec 17)

Final Exam: (Tues, Dec 22)