

2007 MBG 8680 - Computer Applications in Molecular Genetics  
Center for Molecular Medicine & Genetics

Course Directors: Dr. David D. Womble, CMMG  
Dr. Derek Wildman, CMMG

Faculty: Dr. D. Womble, Center for Molecular Medicine & Genetics  
Ms. E. Marks, Shiffman Medical Library  
Dr. D. Wildman, CMMG  
Mr. Michael Kruger, Ob/Gyn  
Dr. R. Haddad, KCI  
Dr. G. Tromp, CMMG

Text: Introduction to Bioinformatics, Edited by S.A. Krawetz & D.D. Womble, Humana Press, 2003.

Location: Shiffman Medical Library, Room 217 (2<sup>nd</sup> Floor Lab)

Format: 9:30-10:30 Instruction  
10:30-11:00 Break  
11:00-12:00 Instruction

**ALL ASSIGNMENTS ARE DUE BY 1:00 AM THE FOLLOWING DAY**

June 18      The Virtual Library      Instructors      Marks/Ellis-Danquah

Students will be introduced to the digital resources available today including electronic journals, electronic databases, and information management software. Key databases to be introduced include PubMed MEDLINE, Web of Science and Journal Citation Reports. Students will learn how to structure a search, manipulate search results and access journal articles. Services available from the library to assist students will also be discussed. EndNote will be introduced as a tool for managing the literature as well as for citing references during manuscript preparation. The format will include lecture, case-based examples and hands-on activities.

Students are encouraged to obtain EndNote 9.0 X (10.0) or related versions from the university's Software Clearinghouse at:  
<http://clearinghouse.wayne.edu/currentsite/downloads.htm#endnotedownload> There is no charge for this software but students will need to have their Access ID/Pipeline ID to download EndNote.

June 19      UNIX Basics      Instructor      Womble

Introduction to UNIX. Students will learn to login to and use a UNIX server. Tasks will include file and directory management, text editing, command line, X Windows, FTP, and E-Mail.

June 20      GCG Sequence Analysis Suite      Instructor      Womble

Introduction to GCG. GCG is a suite of related programs for analyzing nucleic acid and amino acid sequence data. The GCG programs are installed on the Genetics server. Students will learn how to use the GCG programs for basic analysis of sequence data, as well as more advanced functions such as motif searching, database searching, and multiple sequence analysis. Command line, X Windows, and Web interfaces to GCG will be demonstrated. Students will use GCG to solve problems related to their own dissertation research projects.

June 21, 22      Internet Tools for Genomic Sequence Analysis      Instructor      Wildman

Students will be shown the main features of genome databases and genome gateways. The focus will be on the major genomic databases and browsers available at NCBI: <http://www.ncbi.nlm.nih.gov/>, UCSC: <http://genome.ucsc.edu/>, and Ensembl: <http://www.ensembl.org/index.html>. Students will learn how to access data and assess their quality. In addition to the genome browsers, students will learn about gene ontology classifications (e.g. <http://www.geneontology.org/>, <http://www.pantherdb.org/>), and gene expression databases (e.g. <http://symatlas.gnf.org/SymAtlas/>). Students will also be shown how to find internet-based analytical tools that meet their personal research needs.

June 25      BioStatistics      Instructor      Kruger

Basic principles of probability theory and counting rules will be presented. The use of probability in hypothesis testing and its application to molecular biology and genetics will be discussed. Steps in genetic mapping and localization of disease genes will be discussed. Methods for estimating sample sizes required for gene identification will be presented and the use of SPSS for making these determinations.

June 26, 27, 28      Programming with R and Bioconductor      Instructor      Haddad

Programming with R and Bioconductor for microarray expression analysis and interpretation. R is open source software that implements a variant of the S programming language. The Bioconductor suite of R packages permit the analysis and interpretation of microarray data by providing powerful statistical and graphical methods that integrate biological metadata. This series of lectures will serve as an introduction to the Linux implementation of R and Bioconductor with a heavy emphasis on scripting. Students will first be instructed in the structure of R data types and their manipulation. Subsequent lectures will cover essential Bioconductor packages required for the pre-processing of Affymetrix microarray data including quality control and normalization. Students will then be instructed in the different types of statistical analysis of microarray data available through R and Bioconductor. Last, gene expression changes will be interpreted through the use of Gene Ontology annotations.

June 29 Perl Scripting

Instructor

Tromp

Perl is becoming one of the most popular computer languages for biological data analysis. Perl is widely used to automate repetitive tasks, work with DNA sequences and strings, find motifs in data, parse annotations and interface with relational databases. Popular perl modules such as BioPerl, LWP (Web Automation), and DBI (Database Connectivity) will also be discussed.

July 2 Final Project Reports due by 5 pm.