Sample Program for Bioinformatics Students with Interests in Population, Statistical, and Quantitative Genetics

Year 1, Fall:

Human Genetics 544 – Population and Statistical Genetics	3
Bioinformatics 527 – Introduction to Bioinformatics and Computational Biology*	4
Statistics 425 – Introduction to Probability†	3
PIBS 501/503 – Research Responsibility and Ethics*	1
PIBS 600 – Biomedical Sciences Independent Study*	3-6

Year 1, Winter:

Human Genetics 542 – Human Genetics	3
Epidemiology 516 – Genetics in Epidemiology	3
Statistics 426 – Introduction to Theoretical Statistics†	3
Bioinformatics 602 – Journal Club*	1
PIBS 600 – Biomedical Sciences Independent Study*	3-6

Year 2, Fall:

Biostatistics 601 – Probability and Distribution Theory	4
Electrical Engineering and Computer Science 484 – Databases Management Systems	4
Statistics 500 – Applied Statistics I	3
Bioinformatics 990 – Research Rotation*	3-6

Year 2, Winter:

Biostatistics 666 – Statistical Models and Numerical Methods in Human Genetics	3
Statistics 501 – Applied Statistics II	3
Biostatistics 602 – Biostatistical Inference	4
Bioinformatics 990 – Research Rotation*	3-6

*Courses required for the PhD in Bioinformatics.

†Depending on their previous mathematical background and training, students might consider taking Biostatistics 601 (Probability and Distribution Theory) & 602 (Biostatistical Inference) in place of Statistics 425 & 426 in their first year.

Electives to consider by core area:

Molecular Biology
Human Genetics 541 – Molecular Genetics (3)
Human Genetics 542 – Human Genetics (3)
Pathology/Micro 553 – Cancer Biology (3)

Databases & Computing
Statistics 606 – Statistical Computing (3)
Electrical Engineering and Computer Science 484 – Databases Management Systems (4)

Probability & Statistics
Statistics 425 – Introduction to Probability (3)
Statistics 426 – Introduction to Theoretical Statistics (3)
Statistics 500 – Applied Statistics I (3)
Statistics 503 – Applied Multivariate Modeling (3)
Biostatistics 601 – Probability and Distribution Theory (4)
Biostatistics 666 – Statistical Models and Numerical Methods in Human Genetics (3)
Biostatistics 695 – Analysis of Categorical Data (3)
Biostatistics 866 – Advanced Topics in Genetic Modeling (3)

Biophysics and Structural Biology
Molecular, Cellular, and Developmental Biology 411 – Protein Structure and Function (3)

Systemic Modeling
CSCS 501 – An Introduction to Complex Systems (3)

Electives in Bioinformatics
Bioinformatics 547 – Probabilistic Modeling in Bioinformatics (3)
Bioinformatics 575 – Programming Laboratory in Bioinformatics (3)

Important notes: Doctoral students in the Program in Bioinformatics may take advantage of several of the interdisciplinary training opportunities offered at the University of Michigan. Two programs that are highly relevant to doctoral students in Bioinformatics with interests in population, statistical, and quantitative genetics are described briefly below.

(1) The Center for Statistical Genetics is an interdisciplinary program that seeks to encourage research and training at the interface between human genetics and the mathematical sciences. One of the goals of the Center is to provide training to pre- and post-doctoral students and fellows who are interested in applying and/or developing statistical and mathematical models and methods to address problems in genetic studies of human health and disease. The Center currently provides training support, including tuition, stipend, and a small travel allowance, through the Genome Sciences Training Program (see http://csg.sph.umich.edu/index.php). Students from relevant University of Michigan graduate programs, including Bioinformatics, are eligible to apply for support at any stage in their graduate career. Pre-doctoral trainees typically complete the following core courses: Human Genetics 541 & 542, Statistics 425 & 426 or Biostatistics 601 & 602, and Biostatistics 666.

(2) The Master of Arts in Statistics is regarded as a dual degree program by the Statistics Department while a student is working toward a PhD degree in another field, for example, in Bioinformatics. The program requires a minimum of 24 credit hours of course work and must include the following: Biostatistics 601 & 602, Statistics 500 & 501, electives in Statistics (6 hours), and cognate electives (6 hours), for example, Human Genetics 542 & 544. See http://www.stat.lsa.umich.edu/graduate/regular.html for additional information.

Course Descriptions and Pre-requisites:

Bioinformatics 527 – Introduction to Bioinformatics and Computational Biology (4)
This course introduces students to the fundamental theories and practices of Bioinformatics and Computational Biology via a series of integrated lectures and labs. These lectures and labs will focus on the basic knowledge required in this field, methods of high-throughput data generation, accessing public genome-related information and data, and tools for data mining and analysis. The course is divided into four areas: Basics of Bioinformatics, Computational Phylogeny (includes sequence analysis), Systems Biology and Modeling.

Bioinformatics 547 – Probabilistic Modeling in Bioinformatics (3)
The course will be about probabilistic models of proteins and nucleic acids, and their uses in molecular biology. The core of the course will be the analysis of sequences and their biological applications such as the searching of large databases for optimal comparisons or homologies of sequences (DNA nucleotide sequences or amino acid sequences for proteins), location of genes on a string of DNA, estimation of phylogenetic trees, and structural motif recognition for proteins and RNA. Guest lecturers will address the class on topics from the pharmaceutical industry, DNA physics and chromatin structure.

Bioinformatics 575 – Programming Laboratory in Bioinformatics (3)
This course will introduce current software tools and information resources in bioinformatics. Topics will include rapid development programming languages (Python, Perl), relational databases (SQL), Java, exploratory data analysis in R and web services programming. The course will introduce computational biology applications (BLAST, BioConductor, Cytoscape and GenePattern). There will be a class project creating an information resource in genomics and molecular biology that will use these tools to retrieve bioinformatics data from multiple internet resources, build an integrated database and deliver web forms and graphical access. Grades will be based on programming exercises and participation in class discussions and cooperative development projects. Prerequisites: Some familiarity with computer programming is assumed.

Bioinformatics 602 – Journal Club (1)
The human genome sequence and large-scale gene expression data provide new resources for biology which must be accessed and interpreted by computational means. Computational readings and discussion of the current research literature acquaint the student with the methods and research questions being applied to this important data. Each student will choose and present at least one paper each semester.

Biostatistics 601 – Probability and Distribution Theory (4)
Fundamental probability and distribution theory needed for statistical inference. Probability, discrete and continuous distributions, expectation, generating functions, limit theorems, transformations, sampling theory. Prerequisites: Three terms of calculus.

Biostatistics 602 – Biostatistical Inference (4)
Fundamental theory that is the basis of inferential statistical procedures. Point and interval estimation, sufficient statistics, hypothesis testing, maximum likelihood estimates, confidence intervals, criteria for estimators, methods of constructing test and estimation procedures. Prereq: Biostatistics 601

Biostatistics 666 – Statistical Models and Numerical Methods in Human Genetics (3)
Introduction to current statistical methods used in human genetics. Topics will include sampling designs in human genetics, gene frequency estimation, the coalescent method for simulation of DNA sequences, linkage analysis, tests of association, detection of errors in genetic data, and the multi-factorial model. The course will include a simple overview of genetic data and terminology and will proceed with a review of numerical techniques frequently employed in human genetics. Prereq: Biostatistics 602 or Statistics 426

Biostatistics 695 – Analysis of Categorical Data (3)
Regression models for the analysis of categorical data: logistic, probit and complementary log-log models for binomial random variables; log-linear models for cross-classifications of counts; regression models for Poisson rates; and multinomial response models for both nominal and ordinal responses. Model specification and interpretation are emphasized, and model criticism, model selection, and statistical inference are cast within the framework of likelihood-based inference.

Biostatistics 866 – Advanced Topics in Genetic Modeling (3)
Advanced topics in quantitative genetics with emphasis on models for gene mapping, pedigree analysis, reconstruction of evolutionary trees, and molecular genetics experiments, computational mathematics, and statistical techniques such as Chen-Stein Poisson approximations, hidden Markov chains, and the EM algorithm introduced as needed. Prereq: Biostatistics 666

Center for the Study of Complex Systems 501 – An Introduction to Complex Systems (3)
This course covers a broad range of fundamental topics relevant to the study of complex systems. The course work involves weekly readings focus on "classics" in the complex systems literature, in order to give students a broad, general understanding for the variety of work that falls under the rubric of complex systems. Topics to be covered will include evolutionary systems, self-organized criticality, measures of complexity, approaches to modeling complex adaptive systems, and emergence. Authors to be covered include Holland, Axelrod, Kaufmann, Bak, and Gell-Mann.

Electrical Engineering and Computer Science 484 – Databases Management Systems (4)
Concepts and methods for the design, creation, query and management of large enterprise databases. Functions and characteristics of the leading database management systems. Query languages such as SQL, forms, embedded SQL, and application development tools. Database design, integrity, normalization, access methods, query optimization, transaction management, and concurrency control and recovery.

Epidemiology 516 – Genetics in Epidemiology (3)
This course relates genomics to the core public health discipline of epidemiology emphasizing the use of genomics to help describe disease frequency and distribution and to gain insights into biological etiologies. Topics include genetic material in disease, in families and in populations; the investigation of multifactorial traits; model-based linkage analysis; model-free linkage analysis; segregation analysis; allele association and linkage disequilibrium; and gene-gene interactions and gene-environment interactions. Issues related to implementing studies are considered. Prereq: Biostatistics 503 or equivalent

Human Genetics 541 – Molecular Genetics (3)
A combination of classic and current papers in molecular genetics will be selected to accompnay the lecture material (1-2 papers per lecture). The foundations of modern genetics will launch the course, including both the fundamentals and current research methods for analysis of gene structure and gene expression. The gene expression component will include positive and negative regulation of transcription and mRNA splicing and turnover. The basics of DNA recombination, repair, and transposition will be covered in relationship to cancer, evolution, and mutagenesis. Strategies for developmental regulation will be presented. Parallels between prokaryotes and eukaryotes will be drawn, and comparisons will be made between the temporal and spatial control of gene expression in vertebrates and invertebrates. Genetic engineering topics will include gene targeting and transgenesis, with applications to understanding tissue specific control of gene expression. The course will include discussion of the Genome Project, identification of disease genes and an introduction to the medical application of molecular genetics including gene therapy.

Human Genetics 542 – Human Genetics (3)

Human Genetics 544 – Population and Statistical Genetics (3)

Molecular, Cellular, and Developmental Biology 411 – Protein Structure and Function (3)
This course is a practical hands-on approach to extract information about protein and DNA sequences. Students will learn how to do elementary sequence analysis – databases, searching, alignment, and 3D-structure prediction using analysis tools that are available on the web. Students will learn how to get the most out of their sequence. We will also discuss how protein structure is related to function, and investigate how proteins fold.

Pathology/Microbiology 553 – Cancer Biology (3)
This course will cover a broad range of subjects relating to cancer biology. Emphasis is on the relationship between basic science and clinical aspects of cancer.

Statistics 425 – Introduction to Probability (3)
Basic concepts of probability; expectation, variance, covariance; distribution functions; and bivariate, marginal, and conditional distributions. Prereq: Calculus III

Statistics 426 – Introduction to Theoretical Statistics (3)
An introduction to theoretical statistics for students with a background in probability. Probability models for experimental and observational data, normal sampling theory, likelihood-based and Bayesian approaches to point estimation, confidence intervals, tests of hypotheses, and an introduction to regression and the analysis of variance. Prereq: Statistics 425

Statistics 500 – Applied Statistics I (3)
Linear models; definitions, fitting, identifiability, collinearity, Gauss-Markov theorem, variable selection, transformation, diagnostics, outliers and influential observations. ANOVA and ANCOVA. Common designs. Applications and real data analysis are stressed, with students

using the computer to perform statistical analyses. Prereq: Matrix Algebra and Statistics 350 or 426

Statistics 501 – Applied Statistics II (3)
Generalized linear models including logistics regression, Poisson regression, contingency tables. Random effects and repeated measures. Modern regression techniques. Regression and classification trees. Neural networks. Prereq: Statistics 500

Statistics 503 - Applied Multivariate Analysis
Topics in applied multivariate analysis including Hotelling's T-squared, multivariate ANOVA, discriminant functions, factor analysis, principal components, canonical correlations, and cluster analysis. Selected topics from: maximum likelihood and Bayesian methods, robust estimation and survey sampling. Applications and data analysis using the computer is stressed. Prereq: Stat 500

Statistics 606 – Statistical Computing (3)
This course aims to give an overview of techniques in numerical analysis that are useful in the advanced practice of statistics. The course is roughly divided into three parts: evaluation of special functions, numerical linear algebra (linear solvers, matrix factorizations, eigenvalue problems), optimization (unconstrained methods, simplex method, active set methods, penalty function methods, combinatorial optimization), and simulation (importance and rejection sampling, Markov chain methods, exact methods). The course will cover some theoretical issues, but will primarily focus on the design and implementation of algorithms. Prereq: Calculus, linear algebra, and some knowledge of probability and statistics.