

QUANTITATIVE TRAINING REQUIREMENT FOR T32 CBTG TRAINEES

Handling of large data sets or “big data” (exp; Cancer Genomics/Bioinformatics) is becoming a routine aspect of cancer research that touches almost all aspects of basic and translational research. Appropriate quantitative training is required for CBTG trainees to pursue cutting-edge biomedical cancer research and keep pace with research tools that may include the use of publically available databases and diverse formats for big data acquired from emerging genomics/genetics technologies. Starting in Spring of 2016, as a new course work requirement, all CBTG trainees will be required to elect one course (of 14 options) that offers introductory training in bioinformatics, genomics, and biostatistics for basic and translational cancer research. As a required part of this course, trainees will be exposed to experimental design and data analysis, including methods for handling large data sets using computer-based tools and probing publically available databases. For example, one of the recommended courses is entitled “Special Topics in Computer Science/Introduction to Computing in Biology” and is taught by CBTG preceptor Dr. Chad Myers. Dr. Myers uses examples from cancer research projects to teach course objectives that include understanding current questions in genomics/proteomics and related technologies, recognizing common problem types that arise in computational biology and understanding their solution, fundamental programming principles for structuring data and using algorithms, and manipulation of large data files for conversion between popular formats and to extract specific information within “big data”. Trainees with a strong computational background would be exempt from this requirement but strongly encouraged to remain current by auditing this or another course. This course would be taken in the 2nd year of training (other course requirements are usually completed in the spring semester of the first year of training but may be spread out over 2 years for predocs or 3 years for postdocs; many trainees have already completed the Cancer Biology course prior to entering our program).

CBTG trainees must complete one of the following courses:

BIOC 5361 - Microbial Genomics and Bioinformatics

(3.0 cr; Prereq-College-level courses in [organic chemistry, biochemistry, microbiology]; fall, spring, every year)

Introduction to genomics. Emphasizes microbial genomics. Sequencing methods, sequence analysis, genomics databases, genome mapping, prokaryotic horizontal gene transfer, genomics in biotechnology, intellectual property issues.

BIOL 5272 - Applied Biostatistics

(3.0 cr; =[BIOL 3272]; Prereq-One semester of college-level [[calculus or statistics or computer programming], general biology]; A-F only, fall, every year)

Conceptual basis of statistical analysis. Statistical analysis of biological data. Data visualization, descriptive statistics, significance tests, experimental design, linear model, simple/multiple regression, general linear model. Lectures, computer lab.

BIOL 5485 - Bioinformatics: Experimental Design and Computational Analysis in Systems Biology

(3.0 cr; Prereq-4003 or &4003 or equiv; A-F only, fall, every year)

Modern computational tools used in molecular biology and genomics research. When/how to use particular tools, how to interpret results. Principles, advantages/disadvantages of various methods.

CSCI 3003 - Introduction to Computing in Biology

CSCI 5980 – Special Topics in Computer Science (grad level, held in conjunction with CSCI 3003)

(3.0 cr; spring, every year)

Emphasizes computing tasks common in biology. Programming techniques: variables, flow control, input/output, strings, pattern matching, arrays, hash tables, functions, subroutines. Concepts in computing: algorithms, complexity, documentation, regular expressions/grammars, local variables, encapsulation. Students complete lab projects in Perl language.

CSCI 5461 - Functional Genomics, Systems Biology, and Bioinformatics

(3.0 cr; Prereq-3003 or 4041 or #; spring, every year)

Computational methods for analyzing, integrating, and deriving predictions from genomic/proteomic data. Analyzing gene expression, proteomic data, and protein-protein interaction networks. Protein/gene function prediction, Integrating diverse data, visualizing genomic datasets.

CSCI 5481 - Computational Techniques for Genomics

(3.0 cr; Prereq-4041 or #; fall, every year)

Techniques to analyze biological data generated by genome sequencing, proteomics, cell-wide measurements of gene expression changes. Algorithms for single/multiple sequence alignments/assembly. Search algorithms for sequence databases, phylogenetic tree construction algorithms. Algorithms for gene/promoter and protein structure prediction. Data mining for micro array expression analysis. Reverse engineering of regulatory networks.

GCD 5005 – Computer Programming for Cell and Developmental Biology

(3.0 cr, starting fall 2014)

The use of computer programming in biology is rapidly expanding, and in this course students will build their own biologically-oriented computer programs. Typical applications of computer programming in cell and developmental biology are for computational modeling, automated image analysis, and for data analysis. The objective of this course will be to teach basic MATLAB computer programming skills, with applications designed specifically for students to learn how programming is used for modeling of biological processes, for advanced data analysis, and also for quantitative image analysis. Students will gain confidence in building their own code, as well as in applying published code to specific problems. Students who proceed to graduate school, medical school, or employment in industry will find programming a useful skill that will provide them a competitive advantage in the market.

NSC 8320, Sect 17 – Neurostatistics

Offered every other fall

PUBH 6414 - Biostatistical Methods I

(3.0 cr; =[PUBH 6450]; Prereq-[Public hlth [MPH or certificate] student or environmental hlth [MS or PhD]] or #; A-F only, fall, spring, summer, every year)

Descriptive statistics and graphical methods. Use of R-Commander. Probability definitions and rules. Binomial, normal, and t- distributions. Statistical inference: estimates and confidence intervals of means, proportions, relative risks, odds ratios; t-tests; chi-square tests; ANOVA; correlation/linear regression. Correlation/regression. Inference/causality.

PUBH 6450 – Biostatistics I

(4.0 cr; =[PUBH 6414]; Prereq-[College-level algebra, health sciences grad student] or #; A-F only, fall, spring, every year)

Descriptive statistics. Gaussian probability models, point/interval estimation for means/proportions. Hypothesis testing, including t, chi-square, and nonparametric tests. Simple regression/correlation. ANOVA. Health science applications using output from statistical packages.

PUBH 6451 - Biostatistics II

(4.0 cr; Prereq-[[[6420, 6450] or [6414, 6415]] with grade of at least B, health sciences grad student] or #; spring, every year)

Two-way ANOVA, interactions, repeated measures, general linear models. Logistic regression for cohort and case-control studies. Loglinear models, contingency tables, Poisson regression, survival data, Kaplan-Meier methods, proportional hazards models.

Geared towards trial and epidemiology cohort data analysis

PUBH 7445 - Statistics for Human Genetics and Molecular Biology

(3.0 cr; Prereq-[6450, [6451 or equiv]] or #; background in molecular biology recommended; spring, every year)

Introduction to statistical problems arising in molecular biology. Problems in physical mapping (radiation hybrid mapping, DDP), genetic mapping (pedigree analysis, lod scores, TDT), biopolymer sequence analysis (alignment, motif recognition), and micro array analysis.

PUBH 8446 - Advanced Statistical Genetics and Genomics

(3.0 cr; Prereq-[7445, statistical theory at level of STAT 5101-2; college-level molecular genetics course is recommended] or #; spring, every year)

Genetic mapping of complex traits in humans, modern population genetics with an emphasis on inference based observed molecular genetics data, association studies; statistical methods for low/high level analysis of genomic/proteomic data. Multiple comparison and gene network modeling.

STAT 5021 - Statistical Analysis

(4.0 cr; =[STAT 3011, ANSC 3011, ESPM 3012]; Prereq-=: 3011; College algebra or #; Stat course recommended; fall, spring, every year)

Intensive introduction to statistical methods for graduate students needing statistics as a research technique.