GRADUATE STUDENT HANDBOOK

Master of Science in Computational Biology & Quantitative Genetics
Contents

1 Introduction 1

2 Master of Science in Computational Biology and Quantitative Genetics 3
  2.1 Core Competencies .................................................. 3
  2.2 Admissions Procedures and Requirements .......................... 4
    2.2.1 Harvard T.H. Chan School of Public Health Requirements 4
    2.2.2 Program Requirements ........................................... 4
  2.3 Intradepartmental Biostatistics Degree Program Switch Protocol 4
  2.4 Advising and Degree Program Approval ............................ 5
    2.4.1 Academic Advisor ............................................... 5
    2.4.2 Departmental Approval of Program ............................. 5
    2.4.3 Epidemiology Requirement ..................................... 5
    2.4.4 Research Ethics Requirement .................................. 5
  2.5 Satisfactory Progress Requirements ................................ 6
  2.6 Degree Requirements ............................................... 6
    2.6.1 Course Requirements for the SM2 Degree ...................... 6
    2.6.2 Collaborative Research Thesis .................................. 7

3 ADMINISTRATIVE TIMELINE 9

4 PROGRAM FORMS 10
1 Introduction

This handbook describes the Master’s Degree Program in Computational Biology and Quantitative Genetics offered by the Departments of Biostatistics and Epidemiology at Harvard University. The SM program provides students with the rigorous quantitative training and essential skills needed to successfully meet the challenges presented by large-scale public health data – “Big Data” – in biomedical research. The program is designed to prepare students for a career as a bioinformatics analyst or bioinformatics engineer in universities and hospitals, research organizations, and the pharmaceutical and biotechnology industries. It can also provide the foundation for further doctoral studies.

The sections of this handbook include information and regulations concerning entrance requirements, program descriptions, degree requirements, and other program policies. Policies and official requirements of the School of Public Health are set forth in the Harvard T. H. Chan School of Public Health Student Handbook (https://www.hsph.harvard.edu/student-handbook/). Each graduate student is responsible for general knowledge of, and adherence to, the policies and requirements of the degree program in which the student is enrolled. Additional program information is available at the website https://www.hsph.harvard.edu/sm-computational-biology/program/. Vitally important for our community is that all members demonstrate respect for each other and our discipline. For all members of the community, respect is demonstrated by attending all scheduled classes or meetings, and arriving on time, fully prepared, and ready to participate.

This handbook was prepared by the Program Director and approved by the Executive Committee of the Program in Computational Biology and Quantitative Genetics. The Program Director is responsible for reviewing the student’s program of study, and has the authority to consider exceptions to the rules and regulations established by the Executive Committee. Recommendations of the Program Director are forwarded to the Executive Committee for final approval. Both the Program Director and the Executive Committee welcome suggestions and comments.
Program Leadership and Administration

**Executive Committee**

Dr. John Quackenbush, Program Co-Director  
Dr. Peter Kraft, Program Co-Director  
Dr. Lori Chibnik  
Dr. Immaculata De Vivo  
Dr. Curtis Huttenhower  
Dr. Junwei Lu  
Dr. Rulla Tamimi

**Administration**

Jelena Follweiler  
Manager of Academic Services, Biostatistics  
Phone: 617-432-1087  
Email: jtillots@hsph.harvard.edu

Eric Creighton  
Graduate Program Coordinator, Biostatistics  
Phone: 617-432-5061  
Email: ecreighton@hsph.harvard.edu

Eric DiGiovanni  
Assistant Director of Graduate Studies, Epidemiology  
Phone: 617-432-1328  
Email: edigiova@hsph.harvard.edu

**Career Development**

Dr. Erin Lake  
Director of Student Development, Biostatistics  
Email: eklake@hsph.harvard.edu

**Websites**

Master of Science in Computational Biology and Quantitative Genetics  
Department of Biostatistics  
Department of Epidemiology
2 Master of Science in Computational Biology and Quantitative Genetics

This 80-credit program, offered jointly by the Departments of Biostatistics and Epidemiology, is designed to provide students with:

- The biological background needed to understand and interpret data
- A bioinformatics background providing familiarity with essential tools and data resources
- Computational skills used to analyze and manage “Big Data”
- Statistical skills required to appropriately analyze large quantitative datasets
- Epidemiological skills necessary for the design, conduct, and analysis of experiments

The SM in Computational Biology and Quantitative Genetics is intended as a terminal professional degree which will enable you to launch your career in bioinformatics. It can also provide the foundation for further doctoral studies in biostatistics, epidemiology, computational biology, and other related fields.

Students will receive training in quantitative methods, including linear and logistic regression, survival analysis, longitudinal data analysis, statistical computing, clinical trials, statistical consultation and collaboration, and epidemiology. Students will also gain a strong foundation in modern molecular biology and genetics, computer programming, the use and application of tools for analysis of genomic data, methods for integrative analysis, and meta-analysis of genes and gene function.

2.1 Core Competencies

The curriculum for the Master’s in Computational Biology and Quantitative Genetics will provide students with the skills essential to contribute to research projects involving the large, complex genomic datasets that are increasingly common in all areas of biomedical, biological, and public health research. These skills include core competencies in five areas:

1. Biological Background
   - Working knowledge of molecular genetics, the structure and organization of the human genome, gene expression regulation, epigenetic regulation, gene functional descriptions, and modern technologies including genotyping, genome-seq, exome-seq, RNA-seq, ChIP-seq, etc, and their applications, and as well understanding of metagenomics.

2. Bioinformatics Background
   - Familiarity and ability to use the major genomics data resources, basic knowledge of sequence analysis, familiarity with gene functional annotation and pathway analysis, ability to write data management and analysis scripts, working knowledge of data mining and statistical analysis techniques as well as machine learning approaches, and understanding of modern network modeling techniques.

3. Computational Skills
   - Working knowledge of UNIX, a scripting language such as perl or python, an advanced programming language such as c, c++, or java, and R/Bioconductor, and familiarity with database programming and modern web technologies.
4. Biostatistics Skills
   • Fundamental understanding of basic statistical inference and applied regression, survival, longitudinal, and Bayesian statistical analysis

5. Epidemiology Skills
   • Ability to critique the existing evidence for a particular research topic, review and summarize information from many studies, to develop robust research questions and design experiments including power calculations, to understand ethical issues in a given study, to assess and correct for measurement errors, and to access and integrate multiple data resources.

2.2 Admissions Procedures and Requirements

2.2.1 Harvard T.H. Chan School of Public Health Requirements
Application for admission to the SM program is available online on the Admissions Office website (https://www.hsph.harvard.edu/admissions/admissions/how-to-apply/application-requirements/). For information on general requirements for admission, contact the Admissions Office by phone (617/432-1031) or through their website (https://www.hsph.harvard.edu/admissions/).

2.2.2 Program Requirements
All candidates for admission to the SM in Computational Biology and Quantitative Genetics program should have successfully completed the following:

   • An undergraduate degree in mathematical sciences or allied fields (e.g., biology, psychology, economics),
   • Calculus through partial differentiation and multivariable integration,
   • One semester of linear algebra or matrix methods,
   • Either a two-semester sequence in probability and statistics or a two-semester sequence in applied statistics,
   • At least one semester of training in biology, with some familiarity with molecular biology and genetics.

In addition, applicants are encouraged to have completed other courses in quantitative areas and in areas of application in the biological sciences. Practical knowledge of computer scripting and programming as well as experience with a statistical computing package such as R is highly desirable. Additional research or work experience is beneficial, but not required. Applicants should show excellence in written and spoken English.

Evidence that these requirements have been fulfilled should form part of the application.

2.3 IntraDepartmental Biostatistics Degree Program Switch Protocol
The Department of Biostatistics offers a number of master’s degree programs. These programs each have their own goals and requirements and make independent decisions about admissions. Students should carefully choose the program to which they apply and we expect students admitted to a program to meet the requirements of that program.

However, we recognize that the interests of some students may change during their time in graduate school. Therefore, the Department has established the following procedure for students applying for a change in program. Please note that transfers between programs are not automatic and may not be approved, and if a student has received a scholarship or other funds from a degree program, that funding will not transfer to the new degree.
1. Students must complete at least one full semester of coursework before applying for a program transfer.

2. Students must enroll in and successfully pass any required coursework for their current program before beginning the transfer process. Please see your current degree program handbook for a list of required coursework. Students may also need to complete coursework in their proposed program so as to not fall behind in requirements to finish their program on time if approved for transfer.

3. Students must complete and submit a formal application and include an updated statement of purpose describing the reason for seeking a program transfer, current CV and one letter of recommendation.

4. The Directors of department master’s programs, including the Directors of the student’s current program and prospective program, will schedule an interview with the students to assess their application. At the meeting, students will provide a copy of their CV and statement of purpose to each Director, and will briefly explain their decision to apply for a transfer. The Directors will consider the applications and render a decision whether to approve or deny the transfer. These meetings will be scheduled in early January for students applying for transfer to start in the Spring semester, and in late May for students applying to transfer after the Spring semester.

5. The Directors will discuss and notify the student of their decision with three days. All decisions are final.

6. If approved, completed paperwork must be submitted to the Registrar’s Office, and then to the Manager of Academic Services.

Note that all program transfers are at the discretion of the program Directors and are not guaranteed. If approved, students will be assigned a new academic advisor affiliated with the chosen degree program. It is the responsibility of the students to ensure that they complete all requirements for their degree program.

2.4 Advising and Degree Program Approval

2.4.1 Academic Advisor

All entering students are assigned an academic advisor to help plan course loads and explain program requirements. At the earliest possible date, the student and the academic advisor will develop a program of study. Students should bring their Master’s Degree Program form (Section 2.4.2) to all meetings with their advisors to keep on track with their requirements. Should a student wish to change his/her academic advisor, he or she is encouraged to discuss this with the Program Director. In addition, SPH provides services for all students with clinically documented learning and/or physical disabilities.

2.4.2 Departmental Approval of Program

The Master’s Degree program plan must be submitted to the student’s academic advisor and the Program Director for approval, using the Master’s Degree Program form provided at least one semester prior to their expected graduation date.

2.4.3 Epidemiology Requirement

The School of Public Health requires that Master’s students must successfully pass one epidemiology course. The program requires that EPI 201 be taken to satisfy this requirement.

2.4.4 Research Ethics Requirement

Students must satisfy a research ethics requirement by completing a course in responsible conduct of research or by completing an online training course during the first year in the program (see the timeline in Section 3 for details). Students who feel they have already completed an equivalent training program
must submit adequate documentation to, and receive approval from, the Manager of Academic Services in Biostatistics during the first semester in residence.

2.5 Satisfactory Progress Requirements

For students in the SM2 program, a total of 80 credits are required with a minimum of 55 ordinal credits from the core courses, tracks, and electives listed in Section 2.6.1. In addition, SPH students must remain in good academic standing, must complete program requirements within the designated time to degree, and must maintain a cumulative average of 2.70 or above. All ordinal grades for courses used to satisfy program requirements specified in Section 2.6.1 must be at the level of B- or higher. Courses taken on a pass/fail basis cannot be used to satisfy ordinally graded program requirements.

A detailed presentation of SPH’s regulations for Master’s students is found at https://www.hsph.harvard.edu/student-handbook/. All Master’s students and their advisors should make sure that SPH and CBQG program requirements are met according to schedule.

2.6 Degree Requirements

A total of 80 credits are required for the SM2 in Computational Biology and Quantitative Genetics. A minimum of 55 ordinal credits of coursework must be taken from the core courses, tracks, and electives listed below. Students with prior equivalent background to any of the required courses or strong reasons to take a different course can request permission from the Program Director for a substitution of one or more of the required courses. Although academic advisors and departmental staff will work with students to monitor progress, it is ultimately each student’s responsibility to ensure that all requirements are met.

2.6.1 Course Requirements for the SM2 Degree

Fifty-five credits of ordinally graded courses must be taken from the following courses. This includes a 12.5 credit ordinally graded core curriculum consisting of:

- BST 210 Applied Regression Analysis (5 credits, Fall)
- BST 280 Introductory Genomics & Bioinformatics for Health Research (2.5 credits, Fall 2)
- EPI 201 Introduction to Epidemiology Methods I (2.5 credits, Fall 1)
- EPI 249 Molecular Biology for Epidemiologists (2.5 credits, Fall 1)

An additional ten credits comprised of courses in either one of the two following tracks:

**Statistical Genetics Track**

- BST 227 Introduction to Statistical Genetics (2.5 credits, Fall 2)
- EPI 293 Analysis of Genetic Association Studies (2.5 credits, Wintersession)
- EPI 507 Genetic Epidemiology (2.5 credits, Fall 2)
- EPI 511 Advanced Population and Medical Genetics (5 credits, Spring)
  
  or

- BST 247 Advanced Statistical Genetics (2.5 credits, Spring 2)

**Computational Biology Track**

- BST 281 Genomic Data Manipulation (5 credits, Spring)
- BST 282 Introduction to Computational Biology and Bioinformatics (5 credits, Spring)

A minimum of 22.5 additional credits will come from the alternative track or the following list of elective courses:

- BST 212 Survey Research Methods in Community Health (2.5 credits, Spring)
- BST 214 Principles of Clinical Trials (2.5 credits, Spring 1)
BST 217  Statistical & Quantitative Methods for Pharmaceutical Regulatory Services (2.5 credits, Spring 2)
BST 222  Basics of Statistical Inference (5 credits, Fall)
BST 223  Applied Survival Analysis (5 credits, Spring)
BST 226  Applied Longitudinal Analysis (5 credits, Spring)
BST 228  Applied Bayesian Analysis (5 credits, Fall)
BST 230  Probability Theory and Applications I (5 credits, Fall)
BST 231  Statistical Inference I (5 credits, Spring)
BST 232  Methods I (5 credits, Fall)
BST 234  Introduction to Data Structures and Algorithms (5 credits, Spring)
BST 254 Sec 3 Measurement Error and Misclassification (2.5 credits, Fall)
BST 260  Introduction to Data Science (5 credits, Fall)
BST 261  Data Science II (2.5 credits, Spring)
BST 263  Statistical Learning (5 credits, Spring)
BST 267  Introduction to Social and Biological Networks (2.5 credits, Fall)
BST 270  Reproducible Data Science (2.5 credits, Fall)
BST 273  Introduction to Programming (2.5 credits, Fall)
BST 283  Cancer Genome Analysis (5 credits, Spring)
BST 290  Advanced Computational Biology and Bioinformatics (5 credits, Fall)
EPI 202  Elements of Epidemiologic Research: Methods 2 (2.5 credits, Fall)
EPI 203  Study Design in Epidemiologic Research (2.5 credits, Spring)
EPI 204  Analysis of Case-Control and Cohort Studies (2.5 credits, Spring)
EPI 221  Pharmacoepidemiology (2.5 credits, Fall)
EPI 271  Propensity Score Analysis (1.25 credits, Winter)
EPI 288  Data Mining and Prediction (2.5 credits, Spring)
EPI 289  Models for Causal Inference (2.5 credits, Spring)
ID  271  Advanced Regression for Environmental Epidemiology (2.5 credits, Spring)
RDS 280  Decision Analysis for Health and Medical Practices (2.5 credits, Fall)
RDS 282  Economic Evaluation of Health Policy & Program Management (2.5 credits, Spring)
RDS 285  Decision Analysis Methods in Public Health and Medicine (2.5 credits, Spring)
BIOPHYS 170  Quantitative Genomics (5 credits, Fall)
BIOPHYS 376  Functional & Computational Genomics Studies of Transcription Factors & Cis Regulatory Elements (5 credits, Fall or Spring)
BMI  701  Foundations in Biomedical Informatics I (5 credits, Fall)
BMI  702  Foundations of Biomedical Informatics II (5 credits, Spring)
BMI  703  Precision Medicine I: Genomic Medicine (2.5 credits, Fall)
BMI  705  Precision Medicine II: Integrating Clinical and Genomic Data (2.5 credits, Fall)
BMI  713  Computational Statistics for Biomedical Sciences (2.5 credits, Fall)
BMI  715  Computing Skills for Biomedical Informatics (2.5 credits, Fall)

1 May be taken concurrently with BST 210 in the fall if in the BIO or CBQG degree programs. This course is a prerequisite for the doctoral-level core Biostatistics theory courses.
2 These courses are strongly recommended.

2.6.2 Collaborative Research Thesis

The Collaborative Research Thesis provides students with valuable real-world experience doing research in Boston’s premier biomedical institutions. Students are mentored by a member of the program faculty or other affiliated quantitative scientist with a faculty-level appointment working in clinical and epidemiologic research projects at SPH, Harvard University, or a Harvard-affiliated hospitals. In many cases, students may instead choose to undertake their thesis work in an industry setting; our faculty have many ties to the biotech world and can help arrange appropriate research projects. A link to a list of possible thesis advisors
A student must complete a 10-20 credit ordinarily graded CBQG Collaborative Research Thesis (CBQG 325), normally undertaken during the fourth semester, after the required core course work has been completed. This will typically involve data analysis for a research project under the direction of one or more mentors. Students should notify the Manager of Academic Services in Biostatistics about their thesis advisor choice by the end of their first academic year. Students may begin their thesis research during the summer following their second semester. Once a thesis advisor is chosen, and a research project has been decided, the student should submit a thesis proposal by email for approval to the Program Director (see timeline in Section 3 for details). Once the proposal has been approved by the Program Director, send the approved proposal to the Manager of Academic Services in Biostatistics. Examples of previous successful proposals may be obtained from the Manager before submission.

In this Collaborative Research Thesis, a student will perform activities related to the design, conduct, and analysis of research studies with a focus on data analysis and scientific presentation. The student will carry out an extensive data analysis with the goal of addressing a relevant question requiring the use of methods in computational biology or quantitative genetics. Usually these projects will involve interacting with a group of people with varied disciplinary backgrounds. The student will then write a Master’s thesis of approximately 20-25 double-spaced pages excluding tables, figures, and references that describes the medical or public health problem of interest, describes the analytical methods used and their appropriateness, summarizes the data analyses, and provides a scientific interpretation of the data, in a standard scientific writing style. The student will also orally present this work in a seminar of approximately 30 minutes in length. The Master’s thesis and oral presentation will primarily be the work of the student, with only advisory input from the mentor(s). The Master’s thesis and oral presentation will be evaluated by a thesis committee consisting of a minimum of three members. The members will include the student’s thesis advisor(s), the Program Director, and other Biostatistics faculty members or surrogates as needed (potentially including the student’s academic advisor). The student will submit a Thesis Committee Nomination form before scheduling the thesis defense. The thesis defense should be scheduled by contacting the Manager of Academic Services in Biostatistics by late March or early April (see timeline in Section 3 for more details). The Master’s thesis must be submitted to the thesis committee at least two weeks prior to the scheduled oral presentation.
3 ADMINISTRATIVE TIMELINE

Detailed requirements and deadlines for degree completion are given on the Harvard T.H. Chan School of Public Health webpage. All forms linked below are also located on the last page of this Graduate Student Handbook.

- **Summer Before Entering Program**
  
  - Take online course in R. One option is https://www.datacamp.com/courses/free-introduction-to-r

- **Year One**
  - **First Semester**
    - Complete Research Ethics requirement by taking the free online CITI program or attending HPM 548 (See section 2.3.4). You will receive a reminder about where to find the online course before beginning the Fall term.
    
    - If planning to waive courses, ask for those waivers of fall core courses (BST 210 or BST 280) or the track courses by emailing the Program Director and the Manager of Academic Studies in Biostatistics with details about the course(s) taken or experience that you have that may qualify you for a waiver. To waive EPI 201 which is a school-wide core requirement, you must submit this Waiver of Core Courses Form (found at bottom of page) to the instructor teaching the course for approval, and then to the Registrar’s Office. You may **not** waive out of EPI 249.
  
  - **Second Semester**
    
    - Complete or waive spring track courses.
    
    - Attend scheduled meeting about thesis advisors and research projects (March or April).
    
    - Search for thesis advisor (see Section 2.6.2) and potential project.
    
    - Notify the Manager of Academic Services in Biostatistics of your thesis advisor choice by May 15, and remind her to set up sections of CBQG 325 for your advisor if those sections do not already exist in the system.

  - **Summer**
    
    - May start thesis research over the summer. Signing up for thesis credit (CBQG 325) is not necessary or advisable during the summer.

- **Year Two**
  
  - **Third Semester**
    
    - Continue to complete any necessary coursework.

    - Start or continue thesis research, and sign up for research credit (CBQG 325) at this time, counting the summer work as a part of this enrollment. The grade for CBQG 325 in the fall may be “incomplete” until you finish your thesis and defend it in the spring term.

    - Work on your thesis proposal with your thesis advisor, and submit the proposal by email to the Program Director by the beginning of Fall 2 term. Once the proposal has been approved by the Program Director, send the approved proposal to the Manager of Academic Services in Biostatistics (or you may cc the Manager on the original email). Examples of previous successful proposals may be obtained from the Manager prior to submission.
Fourth Semester

- Continue to complete any necessary coursework.
- Start or continue thesis research, and sign up for research credit (CBQG 325).
- Turn in your final program form by February 15.
- Choose your thesis committee members, and complete this form. Submit it to the Manager of Academic Services in Biostatistics by the end of Spring 1 term (no later than March 15).
- Work with Manager of Academic Services in Biostatistics to schedule your thesis defense date/time. The Program Director should have dates/times blocked off in late April or early May for all thesis defenses.

4 PROGRAM FORMS

- CBQG SM2 Degree Program Form
  https://content.sph.harvard.edu/biostats/publications/cbqg_handbook/SM2_Degree_Form_CBQG.pdf

- Thesis Committee Nomination Form