Tentative 2020 SISG Module Schedule, July 13-31

This schedule is subject to change. A final schedule will be posted March 1.

Overview:

<table>
<thead>
<tr>
<th>Monday, July 13</th>
<th>Tuesday, July 14</th>
<th>Wednesday, July 15</th>
<th>Thursday, July 16</th>
<th>Friday, July 17</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-Probability &amp; Statistical Inference</td>
<td>5-Regression Methods: Concepts &amp; Applications</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-Introduction to Genetics and Genomics</td>
<td>6-Integrative Genomics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-Introduction to R</td>
<td>7-Population Genetics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Endangered Populations &amp; Non-model Organisms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Monday, July 20</th>
<th>Tuesday, July 21</th>
<th>Wednesday, July 22</th>
<th>Thursday, July 23</th>
<th>Friday, July 24</th>
</tr>
</thead>
<tbody>
<tr>
<td>8-Bayesian Statistics for Genetics</td>
<td>11-Gene</td>
<td>12-Mixed Models in Quantitative Genetics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-Quantitative Genetics</td>
<td>Genes</td>
<td>13-Statistical Genetics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-Pathway &amp; Network Analysis for Omics Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Monday, July 27</th>
<th>Tuesday, July 28</th>
<th>Wednesday, July 29</th>
<th>Thursday, July 30</th>
<th>Friday, July 31</th>
</tr>
</thead>
<tbody>
<tr>
<td>14-Association Mapping: GWAS &amp; Sequencing Data</td>
<td>17-Computational Pipeline for WGS Data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-MCMC for Genetics</td>
<td>18-Advanced Quantitative Genetics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16-Forensic Genetics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Descriptions for each module are outlined below.
Module 1: Probability and Statistical Inference

Instructor(s): Hughes, James; Moodie, Zoe

Module dates/times: Monday, July 13; 8:30 a.m. -5 p.m.; Tuesday, July 14, 8:30 a.m.-5 p.m., and Wednesday, July 15, 8:30 a.m.-Noon

This module serves as an introduction to statistical inference using tools from mathematical statistics and probability. It introduces core elements of statistical modeling, beginning with a review of basic probability and some common distributions (such as the binomial, multinomial, and normal distributions). Maximum likelihood estimation is motivated and described. The central limit theorem and frequentist confidence intervals are introduced, along with simple Bayes methods.

We then cover classical hypothesis testing scenarios such as one-sample tests, two-sample tests, chi-square tests for categorical data analysis, and permutation tests. The course concludes with an overview of resampling methods, such as the bootstrap and jackknife, and a discussion of multiple testing corrections such as false discovery rate control. This module serves as a foundation for almost all of the later modules.

Training in calculus is not a prerequisite for this module, but a willingness to attempt math problems and some comfort with basic algebra will be necessary. Suggested pairing: Modules 4 and 7.

Access 2019 course materials.

Jim Hughes is Professor of Biostatistics at the University of Washington. He is interested in the application of statistical methods to problems in AIDS and other sexually transmitted diseases. He is particularly interested in cluster randomized trial designs and statistical methods for dealing with misclassified data. He is heavily involved in graduate and undergraduate teaching and graduate student advising, and he has won teaching awards. He recently published “On the design and analysis of stepped wedge trials.” Contemporary Clinical Trials. 45(Pt A):55-60, 2015.

Zoe Moodie - TBD
**Module 2: Introduction to Genetics and Genomics**

Instructor(s): Gibson, Greg; Queitsch, Christine

**Module dates/times:** Monday, July 13; 8:30 a.m. -5 p.m.; Tuesday, July 14, 8:30 a.m.-5 p.m., and Wednesday, July 15, 8:30 a.m.-Noon

This module covers the theory and practice of modern genetics. It is designed to provide biologists with the foundations upon which statistical genetics is built, and/or an introduction to the concepts of classical and contemporary genetics for statisticians and informaticians.

The module starts with the key concepts of quantitative and Mendelian genetics and then illustrates how these have been reconciled with molecular biology. Three half-days are then spent on the basics of genome-wide association mapping as well as exome and whole genome sequencing; on evolutionary and population genetics particularly as they pertain to human biology; and on gene expression profiling and integrative genomics leading to systems biology, also touching on personalized medicine. Suggested pairing: Modules 5 and 9.

Access [2019 course materials](#).

**Greg Gibson** is Professor and Director of the Center for Integrative Genomics at Georgia Tech. He conducts research on genomic approaches to human genetics; variability of gene expression; systems biology of disease; theory of canalization and biological robustness. He recently published “Constraints on eQTL fine mapping in the presence of multisite local regulation of gene expression.” *G3 Genes, Genomes, Genetics* 7:2532-2544, 2017.

**Christine Queitsch** is Associate Professor of Genome Sciences at the University of Washington. Her research focuses on two related fields: the genetic architecture of complex traits and the role of gene regulation and protein folding in generating heritable phenotypic variation. She advances complex trait genetics by ascertaining uncharacterized sequence variation and by resolving the relative importance of additive variation and epistasis in complex traits. Her most recent publication is “Variability in a short tandem repeat mediates complex epistatic interactions in Arabidopsis thaliana.” *Genetics* 205:455, 2017.
Module 3: Introduction to R

Instructor(s): Rice, Ken; Thornton, Timothy

Module dates/times: Monday, July 13; 8:30 a.m. - 5 p.m.; Tuesday, July 14, 8:30 a.m. - 5 p.m., and Wednesday, July 15, 8:30 a.m. - Noon

This module introduces the R statistical environment, assuming no prior knowledge. It provides a foundation for the use of R for computation in later modules.

In addition to discussing basic data management tasks in R, such as reading in data and producing summaries through R scripts, we will also introduce R's graphics functions, its powerful package system, and simple methods of looping.

Hands-on use of R is a major component of this module; users require a laptop and will use it in all sessions. Examples and exercises will use data drawn from biological and medical applications, including infectious diseases and genetics. Participants require a laptop and will use it in all sessions. Suggested pairing: All later modules.

Access 2019 course materials.

Ken Rice is Professor of Biostatistics at the University of Washington. His research focuses primarily on developing and applying statistical methods for complex disease epidemiology, notably cardiovascular disease. He leads the Analysis Committee for the CHARGE consortium, a large group of investigators studying genetic determinants of heart and aging outcomes. He recently published “Large-scale genome-wide analysis identifies genetic variants associated with cardiac structure and function.” J. Clinical Investigation 127:1798-1812.

Tim Thornton is Associate Professor of Biostatistics at the University of Washington. His research interest is in the area of statistical genetics, with an emphasis on statistical methodology for genetic association studies of complex traits in samples with relatedness, ancestry admixture, and/or population structure. He recently published “Admixture mapping in the Hispanic Community Health Study/Study of Latinos reveals regions of genetic associations with blood pressure traits.” PLoS One 12:e0188400, 2017.
Module 4: Endangered Populations & Non-model Organisms

Instructor(s): Nielsen, Dahlia; Singh, Nadia

**Module dates/times:** Monday, July 13; 8:30 a.m. -5 p.m.; Tuesday, July 14, 8:30 a.m.-5 p.m., and Wednesday, July 15, 8:30 a.m.-Noon

This module provides training in conceptual foundations and practical aspects of genetics and genomics tools relevant for natural populations, threatened populations, and non-model organisms. Many of the motivating examples are relevant to issues in conservation of species. Students will learn foundational models in population and evolutionary genetics, as well as core methods used in genetics and genomics. Topics covered include genetic variation, dynamics of small populations, effective population size, population fragmentation and gene flow, phylogeography, inbreeding and inbreeding depression, and natural selection. It also covers methods for developing genomic resources for non-model organisms, with an emphasis on next generation sequence data generation and analysis (RAD-seq, whole genome sequencing, transcriptome sequencing). As it introduces many of the principles and applications of populations genetics, it serves as an excellent precursor to Population Genetics (SISG Module 7). Topics covered are easily applied to model organisms.

This module has been developed using the framework of Scientific Teaching (1). Specifically, we implement backwards design, student-centered learning, and assessment. The course highlights diversity on every axis, including the students, instructors, mode of teaching, and course content. The instructors also intentionally cultivate inclusivity in the classroom using techniques from Tanner (2) which include wait time, think-pair-share, whip around, and multiple hands multiple voices. Student-centered learning activities for this specific course include bookending with discussion (3) and group brainstorming/crowdsourcing (4). In addition, computer-based learning (4) will be implemented using R (in R Markdown) to illustrate fundamental concepts such as sampling, Hardy-Weinberg Equilibrium, and Random Genetic Drift.

**Nadia Singh** is an Associate Professor of Biology and a member of the Institute for Ecology and Evolution at the University of Oregon. Her research focused on the causes and consequences of variation in fundamental genetic processes for evolution. She is particularly interested in phenotypic plasticity in recombination and the molecular mechanisms underlying this phenomenon. She recently published “Wolbachia Infection Associated with Increased Recombination in Drosophila” G3: Genes, Genomes, Genetics, 9(1) 229-237, 2018. She is a recent graduate of the University of Oregon Mobile Summer Institute on Scientific Teaching, and recently completed Science Communication Training associated with the Alan Alda Center for Communicating Science. At the University of Oregon she is the chair of the Diversity, Equity and Inclusion committee for the Biology Department and the founder and chair of the Natural Sciences Diversity Leadership Committee. She is also currently serving on the Diversity, Equity and Inclusion committee for the Genetics Society of America.
**Dahlia Nielsen** is an Associate Professor of Biological Sciences and a resident member of the Bioinformatics Research Center at North Carolina State University. Her research focuses on methods development and applications in identifying genes underlying complex traits, including gene expression responses and molecular signaling between hosts and pathogens. She has been engaged in various projects to develop genomic resources for non-model species. She recently published “**Networks Underpinning Symbiosis Revealed Through Cross-Species eQTL Mapping.**” Genetics 206(4): 2175-2184, 2017. She has attended workshops in teaching STEM, including a multi-day workshop in Active Learning taught by Rebecca Brent and Richard Felder. Drs. Brent and Felder are pioneers in learner-centered approaches to teaching.

Module 5: Regression Methods: Concepts & Applications

Instructor(s): Hubbard, Rebecca; Shi, Xu

Module dates/times: Wednesday, July 15, 1:30-5 p.m.; Thursday, July 16, 8:30 a.m.-5 p.m., and Friday, July 17, 8:30 a.m.-5 p.m.

This module will introduce linear regression as a tool for studying relationships between continuous outcomes and continuous, binary, and categorical predictors. Using linear regression as the foundation, we will explore other regression methods, including logistic regression for the analysis of binary outcomes. Specific topics discussed are: linear regression; regression diagnostics; ANOVA; multiple comparisons; logistic regression; generalized linear models. Participants will have the opportunity for hands-on experience, using R. This module is designed as a foundation for the quantitative genetics and association mapping modules. It assumes the material in Module 1 and will cover the basic commands in R. Suggested pairing: Modules 8 and 11.

Access 2019 course materials.

Rebecca Hubbard is Associate Professor of Biostatistics at the University of Pennsylvania. Her research focuses on development and application of statistical methodology for studies that use observational data from clinical medical practice. Her work emphasizes development of statistical tools for biomedical inference and has been applied to studies of cancer screening, aging and dementia, pharmacoepidemiology, women’s health and behavioral health. Her most recent publication is “An electronic health record-based algorithm to ascertain the date of second breast cancer events,” Medical Care 55:E91-E87.

Shi, Xu – Information coming soon.
Module 6: Integrative Genomics

Instructor(s): Gibson, Greg; Powell, Joseph

Module dates/times: Wednesday, July 15, 1:30-5 p.m.; Thursday, July 16, 8:30 a.m.-5 p.m., and Friday, July 17, 8:30 a.m.-5 p.m.

This module emphasizes how the theory and application of transcriptomics can be extended to include other types of omic analysis, and then integrated using statistical and machine learning tools. It starts with the statistical basis of hypothesis testing, covering the central role of normalization strategies and the specifics of differential expression analysis. Students will be given the opportunity to work examples using open source R code that is in standard use for RNASeq data. The module then discusses options for downstream processing by clustering and module detection/comparison; extensions to methylation profiling, proteomics, and metabolomics; eQTL analysis including fine mapping of regulatory variation; and finally integrative methodologies addressing the relationship between genomic, meta-genomic, and phenotypic variation. This module deals primarily with upstream data processing methods that lead to the delineation of networks and pathways. Suggested pairing: modules 2 and 9.

Access [2019 course materials](#).

**Greg Gibson** is Professor and Director of the Center for Integrative Genomics at Georgia Tech. He conducts research on genomic approaches to human genetics; variability of gene expression; systems biology of disease; theory of canalization and biological robustness. He recently published "Constraints on eQTL fine mapping in the presence of multisite local regulation of gene expression." G3-Genes,Genomes, Genetics7:2532-2544, 2017.

**Joseph Powell** is Associate Professor and Scientific Director of the Garvan Weizmann Centre for Cellular Genomics at the Garvan Institute, Sydney. His lab develops and applies computational and statistical genomics approaches to investigate the genetic control of genome regulation and its role in contributing to the susceptibility to human disease. Specifically, his research involves the use of large-scale transcriptomic and DNA sequence data from both bulk tissues and single cells, focusing on understanding the genetic mechanisms by which heritable variants contribute to disease susceptibility at a cellular level, and ultimately achieve therapeutic and diagnostic outcomes. He has recently published "Genetic correlations reveal the shared genetic architecture of transcription in human peripheral blood." Nature Communications 8:Article 483.
Module 7: Population Genetics

Instructor(s): Hernandez, Ryan; O’Connor, Timothy

Module dates/times: Wednesday, July 15, 1:30-5 p.m.; Thursday, July 16, 8:30 a.m.-5 p.m., and Friday, July 17, 8:30 a.m.-5 p.m.

This module considers the analyses now possible for whole-genome sequence data collected on large numbers of individuals. Specific topics include characterization of de novo mutations and the comparison of growth rates among populations. Sequence data allow detailed examination of the signatures of natural selection and methods to compare selective constraints across populations and to seek evidence for recent, population-specific adaptation will be covered. The analysis of identity-by-descent segment sharing, and random projection for IBD detection (RaPID) to infer demographic history will be covered, as will methods to reconstruct the genetic architecture of major human diseases. Suggested pairing: Modules 8 and 12.

Access 2019 course materials.

Ryan Hernandez is Associate Professor at the McGill University and Genome Quebec Innovation Center. His research focuses on computational genomics: characterizing patterns of genetic variation within and between populations using large-scale genome resequencing data; developing novel population genetic simulation techniques; and exploiting population genetic models of demographic history and natural selection to interrogate the genetic basis of disease. He recently published "Prominent features of the amino acid mutation landscape in cancer." PLoS One 12:Article e0183273.

Tim O’Connor is Assistant Professor at the University of Maryland School of Medicine. His research explores the effects of evolution and population structure on the genomic architecture of disease and other phenotypes. He has a track record of developing new algorithms and statistics to interdisciplinary biological problems as well as the use of large multifaceted data sets, particularly the output of next-generation sequencing. He is especially interested in the recent evolution of New World populations such as Hispanic Americans, African Americans, and the Old Order Amish. He recently published "Accurate and equitable medical genomic analysis requires an understanding of demography and its influence on sample size and ratio." Genome Biology 18:Article 42.

Hernandez and O’Connor recently jointly published "Using genotype array data to compare multi- and single-sample variant calls and improve variant call sets from deep coverage whole-genome sequencing data." Bioinformatics 33:1147-1153.
Module 8: Bayesian Statistics for Genetics

Instructors: Rice, Ken; Wakefield, Jonathan

Module dates/times: Monday, July 20; 8:30 a.m. -5 p.m.; Tuesday, July 21, 8:30 a.m.-5 p.m., and Wednesday, July 22, 8:30 a.m.-Noon

The use of Bayesian methods in genetics has a long history. This introductory module begins by discussing introductory probability. It then describes Bayesian approaches to binomial proportions, multinomial proportions, two-sample comparisons (binomial, Poisson, normal), the linear model, and Monte Carlo methods of summarization. Advanced topics include hierarchical models, generalized linear models, and missing data. Illustrative applications will include: Hardy-Weinberg testing and estimation, detection of allele-specific expression, QTL mapping, testing in genome-wide association studies, mixture models, multiple testing in high throughput genomics. Suggested pairing: All later modules.

Access 2019 course materials.

Ken Rice is Professor of Biostatistics at the University of Washington. His research focuses primarily on developing and applying statistical methods for complex disease epidemiology, notably cardiovascular disease. He leads the Analysis Committee for the CHARGE consortium, a large group of investigators studying genetic determinants of heart and aging outcomes. He recently published "Large-scale genome-wide analysis identifies genetic variants associated with cardiac structure and function." J. Clinical Investigation 127:1798-1812.

Jon Wakefield is Professor of Statistics and Biostatistics at the University of Washington. His research interests include spatial epidemiology, space-time models for infectious disease data, small area estimation, hierarchical models for survey data, estimating national and subnational disease burden, ecological inference for non-infectious and infectious disease data, genome-wide association studies, analysis of next generation RNAseq data and the links between Bayes and frequentist procedures. He recently published "Impacts of Neanderthal-introgressed sequences on the landscape of human gene expression." Cell 168:916, 2017.
Module 9: Quantitative Genetics

Instructors: Walsh, Bruce

Module dates/times: Monday, July 20; 8:30 a.m.-5 p.m.; Tuesday, July 21, 8:30 a.m.-5 p.m., and Wednesday, July 22, 8:30 a.m.-Noon

This module assumes the material in Modules 1 and 5 and it provides a foundation for many later modules.

Quantitative Genetics is the analysis of complex characters where both genetic and environment factors contribute to trait variation. Since this includes most traits of interest, such as disease susceptibility, crop yield, growth and reproduction in animals, human and animal behavior, and all gene expression data (transcriptome and proteome), a working knowledge of quantitative genetics is critical in diverse fields from plant and animal breeding, human genetics, genomics, behavior, to ecology and evolutionary biology.

The course will cover the basics of quantitative genetics including: genetic basis for complex traits, population genetic assumptions including detection of admixture, Fisher's variance decomposition, covariance between relatives, calculation of the numerator relationship matrix based on IBD alleles and an arbitrary pedigree, the genomic relationship matrix based on AIS alleles, heritability in the broad and narrow sense, inbreeding and cross-breeding, and response to selection. Suggested pairing: All later modules.

Access 2019 course materials.

Bruce Walsh is Professor of Ecology and Evolutionary Biology, Public Health, and Plant Sciences at the University of Arizona. His interests are in applications of quantitative and statistical genetics to a diverse array from problems, from breeding to human genetics and evolution. He is the co-author of two of the leading texts in the field of quantitative genetics: "Genetics and Analysis of Quantitative Characters" (980 pp. Sinauer Associations) and "Evolution and Selection Quantitative Traits" (1500 pp. Oxford University Press).
Module 10: Pathway & Network Analysis for Omics Data

Instructor(s): Motsinger-Rief, Alison; Shojaie, Ali

Module dates/times: Monday, July 20; 8:30 a.m.-5 p.m.; Tuesday, July 21, 8:30 a.m.-5 p.m., and Wednesday, July 22, 8:30 a.m.-Noon

Networks represent the interactions among components of biological systems. In the context of high-dimensional omics data, relevant networks include gene regulatory networks, protein-protein interaction networks, and metabolic networks. These networks provide a window into biological systems as well as complex diseases, and can be used to understand how biological functions are implemented and how homeostasis is maintained. On the other hand, pathway-based analyses can be used to leverage biological knowledge available from literature, gene ontologies or previous experiments in order to identify the pathways associated with disease or an outcome of interest.

In this module, various statistical learning methods for reconstruction and analysis of networks from omics data are discussed, as well as methods of pathway enrichment analysis. Particular attention is paid to omics datasets with a large number of variables, e.g. genes, and a small number of samples, e.g. patients. The techniques discussed will be demonstrated in R. This course assumes familiarity with R or other command-line programming languages. Suggested pairing: Modules 5, 10, 15.

Access 2019 course materials.


Alison Motsinger-Reif is Associate Professor of Statistics at North Carolina State University. The primary goal of her research is the development of computational methods to detect genetic risk factors of common, complex traits in human populations. She focuses on the development of methods to detect complex predictive models in high-throughput genomic data. She recently published “Metabolic network failures in Alzheimer’s disease: A biochemical road map.” Alzheimers and Dementia 13:965-984, 2017.
Module 11: Genetic Epidemiology

Instructor(s): Fohner, Alison; Lindstroem, Sara

Module dates/times: Wednesday, July 22, 1:30-5 p.m.; Thursday, July 23, 8:30 a.m.-5 p.m., and Friday, July 24, 8:30 a.m.-5 p.m.

This module provides an overview of genetic epidemiology, with a focus on design, analysis and interpretation in studies of complex disease. The module is meant as an introduction to the field with a focus on surveying the various methods for discovering how genetic factors influence health and disease.

It discusses classic genetic epidemiology methods and study designs, including twin studies, family studies, segregation analysis, linkage analysis and population-based association studies, as well as more contemporary topics including gene-environment interactions, rare variant analysis and precision medicine applications. These topics will be reinforced through in-class exercises along with critical reading and discussion of recent publications. Suggested pairing: Modules 9 and 13.

Access 2019 course materials.

Alison Fohner is Assistant Professor of Epidemiology at the University of Washington. She studies how genetic information can improve public health, focusing on how genetic variation affects the response of an individual to particular medications. She has training in pharmacogenetic research methods and data analysis, as well as the legal, ethical, and social implications of conducting and applying that research in health care systems. She is completing a fellowship in informatics and delivery science at Kaiser Permanente Northern California, where she is applying machine learning to identify variation in response to medical treatment within extensive electronic medical record data. Her recent publications include "Carnitine palmitoyltransferase 1A P479L and infant death: policy implications of emerging data." Genetics in Medicine 19:851-857.

Sara Lindstroem is Assistant Professor of Epidemiology at the University of Washington. Her research focuses on understanding the genetic contribution to common complex traits, with a primary emphasis on cancer and linked traits. By leveraging long-running large population-based studies, she investigates how our genetics and environment affect our risk of developing disease. Her current research projects include studying the shared genetic origin between common cancers and the genetics underlying childhood obesity, breast tissue composition and venous thromboembolism. She is also seeking approaches to incorporate information about the functional characteristics of the genome in her studies. She is involved in several large-scale international collaborations that study the genetics underlying breast and prostate cancer. Among her recent publications is "Assessing the causal relationship between obesity and venous thromboembolism through a Mendelian Randomization study." Human Genetics 136:897-902, 2017.
Module 12: Mixed Models in Quantitative Genetics

Instructors: Walsh, Bruce

Module dates/times: Wednesday, July 22, 1:30-5 p.m.; Thursday, July 23, 8:30 a.m.-5 p.m., and Friday, July 24, 8:30 a.m.-5 p.m.

This module assumes the material in Modules 1 and 5 and provides a foundation for Module 14. "Mixed models" refer to the analysis of linear models with arbitrary (co)variance structures among and within random effects and may be due to such factors as relationships or shared environments, cytoplasm, maternal effects and history. Mixed models are utilized in complex data analysis where the usual assumption(s) of independence and/or homogeneous variances fail. Mixed models allow effects of nature to be separated from those of nurture and are emerging as the default method of analysis for human data. These issues are pervasive in human studies due to the lack of ability to randomize subjects to households, choice, and prior history.

In plant breeding, growth and yield data are correlated due to shared locations, but diminish by distance resulting in spatial correlations. In animal breeding, performance data are correlated because individuals maybe related and may share common material environment as well as common pens or cages. Further, when individuals share a common space, they may experience indirect genetics effects (IGEs), which is an inherited effect in one individual experienced as an environmental effect in an associated individual. The evolution of cooperation and competition is based on IGEs, the estimation of which require mixed model analysis. Detection of cytoplasmic and epigenetic effects rely heavily on mixed model methods because of shared material or parental histories.

Topics to be discussed include a basic matrix algebra review, the general linear model, derivation of the mixed model, BLUP and REML estimation, estimation and design issues, Bayesian formulations.

Applications to be discussed include estimation of breeding values and genetic variances in general pedigrees, association mapping, genomic selection, spatial correlations and corrections, maternal genetic effects, detecting selection from genomic data, admixture detection and correction, direct and indirect genetic effects, models of general group and kin selection, genotype by environment interaction models. Suggested pairing: Modules 9 and 15. Access 2019 course materials.

Bruce Walsh is Professor, Ecology and Evolutionary Biology, University of Arizona. His interests are broadly in using mathematical models to explore the interface of genetics and evolution, with particular focus on two areas: the evolution of genome structure and the analysis of complex genetic characters (aka quantitative genetics). He is well-known as co-author of “Genetics and Analysis of Quantitative Characters.” 980 pp. Sinauer Associations.
Module 13: Statistical Genetics

Instructor(s): Goudet, Jérôme; Weir, Bruce

Module dates/times: Wednesday, July 22, 1:30-5 p.m.; Thursday, July 23, 8:30 a.m.-5 p.m., and Friday, July 24, 8:30 a.m.-5 p.m.

This module serves as a foundation for many of the later modules. It includes:

- A unified treatment for the analysis of discrete genetic data, starting with estimates and sample variances of allele frequencies to illustrate genetic vs statistical sampling and Bayesian approaches.
- A detailed look at Hardy-Weinberg and linkage disequilibrium, including the use of exact tests with mid-p-values and a new look at X-chromosome Hardy-Weinberg testing.
- A new characterization of population structure with F-statistics, based on allelic matching within and between populations with individual inbreeding and relationship estimation as a special case.
- Analyses illustrated with applications to forensic science and association mapping, with particular reference to rare variants.

Concepts illustrated with R exercises. Suggested pairing: Modules 6, 8 and all later modules.

Access 2019 course materials.

Jérôme Goudet is Associate Professor of Ecology and Evolution at the University of Lausanne, Switzerland. His research concerns an understanding of the interplay of population structure, trait architecture, and selection. For this, he uses different approaches, from theory and the development of statistical tools to field observations. He recently published "apex: phylogenetics with multiple genes." Molecular Ecology Resources 17:19-26, 2017. He developed the R package hierfstat.

Bruce Weir is Professor of Biostatistics and Director of the Institute for Public Health Genetics at the University of Washington. He develops statistical methodology for genetic data with an emphasis on allelic dependencies, population structure, disease associations and relationships, and the use of genetic data for human identification. His most recent publication is "Detection and quantification of inbreeding depression for complex traits from SNP data." Proc. Natl. Acad. Sci. USA 114:8602-8607, 2017.

Goudet and Weir recently jointly published "How to estimate kinship." Molecular Ecology 27:4121-4135. They are currently working on the 3rd Edition of "Genetic Data Analysis."
Module 14: Association Mapping: GWAS and Sequencing Data

Instructor(s): Thornton, Timothy; Wu, Michael

Module dates/times: Monday, July 27; 8:30 a.m. -5 p.m.; Tuesday, July 28, 8:30 a.m.-5 p.m., and Wednesday, July 29, 8:30 a.m.-Noon

This module will provide students with the basic tools to carry out genetic association analysis within the context of genome wide association studies (GWAS) and next-generation sequencing studies with considerable emphasis on hands-on learning.

Topics covered include: case-control (disease) association testing; quantitative trait analysis; quality control processes in GWAS; multi-locus testing using gene and pathway information; population structure and ancestry inference; association testing in the presence of population structure and/or relatedness; gene-environment and gene-gene interactions; basic rare variant association analysis in sequencing studies; advanced rare variant methods; sequence kernel association tests (SKAT); meta analysis; design considerations; and other emerging topics.

An important component of this module is in-class software exercises which will provide students with hands-on experience analyzing real data using state-of-the-art analysis tools for GWAS and next generation sequencing data.

Assumes basic familiarity with R. Other public domain software that will be used includes PLINK.

Suggested pairing: Modules 12 and 16.

Access 2019 course materials.

Timothy Thornton is Associate Professor of Biostatistics at the University of Washington. His research interest is in the area of statistical genetics, with an emphasis on statistical methodology for genetic association studies of complex traits in samples with relatedness, ancestry admixture, and/or population structure. He recently published “Admixture mapping in the Hispanic Community Health Study/Study of Latinos reveals regions of genetic associations with blood pressure traits.” PLoS One 12:e0188400, 2017.

Michael Wu is an Associate Member in the Biostatistics and Biomathematics Program at the Fred Hutchinson Cancer Research Center. The major thrust of his research lies in the development and application of statistical methods for translational science and particularly for analysis of high-dimensional genomic data within the broader context of clinical trials as well as population-based genetic, genomic, epigenetic, and microbiome studies. He recently published “A fast small-sample kernel independence test with application to microbiome association studies.” Biometrics 73:1453-1463, 2017.

Thornton and Wu recently jointly published “Powerful genetic association analysis for common or rare variants with high-dimensional structured traits.” Genetics 206:1779-1790, 2017.
Module 15: MCMC for Genetics

Instructor(s): Anderson, Eric; Stephens, Matthew

Module dates/times: Monday, July 27; 8:30 a.m. - 5 p.m.; Tuesday, July 28, 8:30 a.m.-5 p.m., and Wednesday, July 29, 8:30 a.m.-Noon

This module examines the use of Bayesian Statistics and Markov chain Monte Carlo methods in modern analyses of genetic data. It assumes a solid foundation in basic statistics and the concept of likelihood. Some population genetics and a basic familiarity with the R statistical package, or other computing language, will be helpful.

The first day includes an introduction to Bayesian statistics, Monte Carlo, and MCMC. Mathematical concepts covered include expectation, laws of large numbers, and ergodic and time-reversible Markov chains. Algorithms include the Metropolis-Hastings algorithm and Gibbs sampling. Some mathematical detail is given; however, there is considerable emphasis on concepts and practical issues arising in applications.

Mathematical ideas are illustrated with simple examples and reinforced with a computer practical using the R statistical language. With that background, two applications of MCMC are investigated in detail: inference of population structure (using the program STRUCTURE) and haplotype inference (using the program PHASE). Computer exercises using both programs are included.

Further topics include the use of MCMC in model evaluation and model checking, strategies for assessing MCMC convergence and diagnosing MCMC mixing problems, importance sampling, and Metropolis-coupled MCMC.

Suggested pairing: Modules 7 and 12.

Access 2019 course materials.

Eric Anderson – Information coming soon.

Matthew Stephens is Professor of Statistics and Human Genetics at the University of Chicago. He was a developer of STRUCTURE, a widely used computer program for determining population structure and estimating individual admixture. He also was a developer of the influential Li and Stephens model as an efficient model for linkage disequilibrium. His recent publications include “Bayesian large-scale multiple regression with summary statistics from genome-wide association studies.” Annals of Applied Statistics 11:1561-1592.
Module 16: Forensic Genetics

Instructors: Aalbers, Sanne; Weir, Bruce

Module dates/times: Monday, July 27; 8:30 a.m.-5 p.m.; Tuesday, July 28, 8:30 a.m.-5 p.m., and Wednesday, July 29, 8:30 a.m.-Noon

This model covers the basic statistical and genetic methods leading to likelihood ratios (LRs) for the presentation of genetic evidence. It provides the background necessary for using analysis results from packages such as CODIS Popstats.

This module also:

- Describes forensic STR markers: mutation process, genotyping technology, and electropherogram artifacts particularly new considerations for back, forward, double back stutter and exotics.
- Reviews principles of population genetics, and measurement of relatedness.
- Covers general principles of evidence evaluation using LRs, computing LRs for identification using presence/absence of autosomal STR genotypes and for mitochondrial and Y-chromosome markers.
- Addresses the complications of mixture interpretation when the queried contributor is a relative of true contributor.
- Describes the consequences of database searches.
- Discusses briefly probabilistic interpretation of STR profiles.
- Provides information about new molecular techniques for human identification.

The module is suitable for graduate students in population genetics, forensic science practitioners, and lawyers facing DNA evidence.

Access 2019 course materials.

Sanne Aalbers is a Research Scientist in the Genetic Analysis Center at the University of Washington. She received degrees in Applied Mathematics from Delft University of Technology and in Forensic Science from the University of Amsterdam. She has worked on financial crime analytics, for Deloitte Forensic. At the University of Washington she derived ROC curves for familial DNA database searching, and developed match probabilities for Y-STR profiles. Currently she is developing new stutter models for NGS forensic data.

Bruce Weir is Professor of Biostatistics and Director of the Institute of Public Health Genetics at the University of Washington. He is a member of the Biology/DNA Scientific Area Committee of the NIST/NIJ OSAC organization. He develops statistical analysis methods for the interpretation of forensic genetic profiles. He is co-author of “Interpreting DNA Evidence,” Sinuwaer, 1998. His recent forensic publications include “Population-specific FST values: A worldwide survey.” Forensic Science International: Genetics 23:91-100, 2016.
Module 17: Computational Pipeline for WGS Data

Instructor(s): Gogarten, Stephanie; Rice, Ken; Thornton, Timothy

Module dates/times: Wednesday, July 29, 1:30-5 p.m.; Thursday, July 30, 8:30 a.m.-5 p.m., and Friday, July 31, 8:30 a.m.-5 p.m.

This module is designed to follow on from Module 14. It will be a hands-on introduction to whole genome sequence analysis pipelines, informed by the instructors’ experience with the TOPMed project (www.nhlbiwgs.org), and in particular its focus on pooled-data analysis used to study the role of rare variants on disease outcomes.

It will begin with an overview of data formats (BAM, VCF, GDS), and then cover population structure and relatedness effects on association mapping, phenotype harmonization, association testing (single-variant, burden and SKAT), variant annotation, WGS variant analysis pipelines focusing on tools used in the TOPMed Analysis pipeline and the role of cloud computing.

Access 2019 course materials.

Suggested pairing: Modules 12 and 13.


Ken Rice is Professor of Biostatistics at the University of Washington. His research focuses primarily on developing and applying statistical methods for complex disease epidemiology, notably cardiovascular disease. He leads the Analysis Committee for the CHARGE consortium, a large group of investigators studying genetic determinants of heart and aging outcomes. He recently published “Large-scale genome-wide analysis identifies genetic variants associated with cardiac structure and function.” J. Clinical Investigation 127:1798-1812.

Tim Thornton is Associate Professor of Biostatistics at the University of Washington. His research interest is in the area of statistical genetics, with an emphasis on statistical methodology for genetic association studies of complex traits in samples with relatedness, ancestry admixture, and/or population structure. He recently published “Admixture mapping in the Hispanic Community Health Study/Study of Latinos reveals regions of genetic associations with blood pressure traits.” PLoS One 12:e0188400, 2017.
Module 18: Advanced Quantitative Genetics

Instructor(s): Visscher, Peter; Yengo, Loic

Module dates/times: Wednesday, July 29, 1:30-5 p.m.; Thursday, July 30, 8:30 a.m.-5 p.m., and Friday, July 31, 8:30 a.m.-5 p.m.

This module focuses on the genetics and analysis of quantitative traits in human populations, with emphasis on estimation and prediction analysis using genetic markers. Topics include: the resemblance between relatives; estimation of genetic variance associated with genome-wide identity by descent; GWAS for quantitative traits; the use of GWAS data to estimate and partition genetic variation; principles and pitfalls of prediction analyses using genetic markers.

A series of computer exercises will provide hands-on experience of implementing a variety of approaches using R, the Merlin suite of software, PLINK and GCTA.

Suggested pairing: Modules 8 and 11.

Access to 2019 course materials.

Peter Visscher is Professor and Chair of Quantitative Genetics at the University of Queensland. His research focuses on understanding individual differences between people in traits that are important for health outcomes and aging. A better understanding of the genes that underlie variation in risk to diseases may lead to better treatments. The traits he studies include gene expression, gene methylation, height and body-mass-index, psychiatric disease and neurogenetic conditions. He recently published “Concepts, estimation and interpretation of SNP-based heritability.” Nature Genetics 49:1304, 2017.

Loic Yengo – Information coming soon.