

19TH SUMMER INSTITUTE IN STATISTICAL GENETICS



July 7-25, 2014

UNIVERSITY *of* WASHINGTON

Seattle, Washington

sisg.biostat.washington.edu

Sponsored by:



DEPARTMENT OF BIOSTATISTICS
SCHOOL OF PUBLIC HEALTH
UNIVERSITY *of* WASHINGTON

AGTGAAGCTACTTTAGAAAGTACTGCTACTGGTGAAT
SISG
2014

Supported by the National Institutes of Health and the National Science Foundation



Administrative Details

General Information

The 19th Summer Institute in Statistical Genetics (SISG 2014) will be held at the University of Washington in Seattle, Washington during July 7-25, 2014. The Institute consists of a series of two-and-a-half day workshops designed to introduce geneticists to modern methods of statistical analysis and to introduce statisticians to the statistical challenges posed by modern genetic data. Prerequisites are minimal, and the modular nature of the Institute enables participants to design a program best suited to their backgrounds and interests. Most participants take two or three modules. Individuals attending the Institute will receive certificates of course completion in recognition of their participation.

In 2014 there will be two concurrent Summer Institutes in Seattle. Details about the Summer Institute in Statistics and Modeling in Infectious Diseases (SISMID 2014), 7-23 July 2014, are available online at: depts.washington.edu/sismid/. The SISG 2014 modules are timed such that participants may also register for SISMID 2014 modules. We are also hosting the Summer Institute in Statistics for Clinical Research (SISCR 2014), 23-27 June 2014. Details for SISCR 2014 are available online at: <http://www.biostat.washington.edu/suminst/siib/general>.

Registration Deadlines and Fees

Registration fee per module: \$675 (USD); Early-bird rate: \$575 (USD).

Reduced academic, government, and not-for-profit org. fee per module: \$525 (USD); Early-bird rate: \$425 (USD).

Early-bird deadline is 11:59 PM (Pacific USA) on Monday, 9 June 2014. Participation in any module is subject to space availability, and participants are encouraged to register early. Seating for each module is limited.

Registration fees cover tuition, course materials, coffee breaks, networking, and tutorials. **Meals, travel and lodging are not covered.** No textbooks are required or supplied, but books recommended as background reading are listed in the module descriptions (*please refer to the Institute's website for updated module descriptions*).

Refund Policy: A \$25 processing fee will be deducted from refunds requested before 12:00 AM, Tuesday, 24 June 2014. A \$200 processing fee will be deducted from refunds requested after 12:01 AM, Tuesday, 24 June 2014 and before 12:00 AM, Tuesday, 1 July 2014. No refunds will be processed after 12:01 AM, Tuesday, 1 July 2014.

Payments can be made with Visa, Mastercard, or Diner's Club credit cards online via a secure server. For payments by Wire Transfer or Purchase Order, please refer to the online registration site for instructions.

Scholarships

Some registration-fee and travel scholarships are available for students. Students can apply for up to a maximum of three (3) modules for consideration of a scholarship. Applications may be submitted between 11:59 pm (US Pacific Time) on Monday 3 February 2014, and 11:59 pm (US Pacific time) on Friday, 25 April 2014. Scholarships will be granted on a rolling basis and most decisions will be sent to applicants via email within 10 days of receipt of your completed application. Scholarship funds may be exhausted before April 25, 2014. Participation in any module is subject to space availability and seating for each module is limited. Students can register for up to a maximum of six (6) modules (one module per session); however, any scholarship award will only cover up to a maximum of three (3) modules, and extra modules will be paid for by the student. Due to University policies, we are able to grant tuition awards based on an applicant's qualifications, but no travel awards, to students residing outside the United States.

Computing

Most modules will incorporate computing and participants are encouraged (and for some modules, required) to bring laptop computers with them. They will have free online access while they are on the University of Washington campus. Participants are also highly encouraged to follow the online instructions at the Institute website to download public domain software and data before they arrive. The Institute does not loan out laptops.

Lodging Accommodations

A list of local hotels offering special rates to participants is on the Institute website along with information for requesting dormitory accommodation.

Conference Mailing Address

SISG 2014 c/o UW BIOSTATISTICS
4333 Brooklyn Avenue NE
UW Tower 15T-420
Box 359461
Seattle, WA 98195-9461



Instructors

Joshua M. Akey

Associate Professor of Genome Sciences
University of Washington

Eric C. Anderson

Research Molecular Geneticist
National Marine Fisheries Service
Assistant Adjunct Professor of
Physical and Biological Sciences
University of California at Santa Cruz

Philip Awadalla

Genome Quebec Chair of Population
and Medical Genomics
University of Montreal

John S. Buckleton

Principal Scientist
ESR, New Zealand

Rebecca W. Doerge

Anderson Distinguished Professor
and Chair of Statistics
Purdue University

Karen Edwards

Professor of Epidemiology
University of California, Irvine

Joseph Felsenstein

Professor of Genome Sciences & Biology
University of Washington

Stephanie Fullerton

Associate Professor of Bioethics
and Humanities
University of Washington

Nanibaa' Garrison

Assistant Professor of Pediatrics
Vanderbilt University

Michel Georges

Professor of Animal Genomics
University of Liège

Gregory C. Gibson

Professor, School of Biology
Georgia Institute of Technology

Michael Goddard

Professorial Fellow in Animal Genetics
University of Melbourne, Australia
Victorian Department of Primary
Industries

Jérôme Goudet

Associate Professor of Population
Genetics
University of Lausanne

Peter D. Hoff

Professor of Statistics
University of Washington

Mark T. Holder

Assistant Professor of Ecology
and Evolutionary Biology
University of Kansas

Rebecca Hubbard

Assistant Investigator
Group Health Research Institute

James P. Hughes

Professor of Biostatistics
University of Washington

Caroline Hutter

Program Director, Division of Genomic
Medicine, National Human Genome
Research Institute

Lurdes Y. T. Inoue

Associate Professor of Biostatistics
University of Washington

Kathleen F. Kerr

Associate Professor of Biostatistics
University of Washington

Mary K. Kuhner

Research Associate Professor of
Genome Sciences
University of Washington

Thomas Lumley

Professor of Biostatistics
University of Auckland

Alison Motsinger-Reif

Associate Professor of Statistics
North Carolina State University

William M. Muir

Professor of Genetics
Purdue University

Dahlia Nielsen

Research Associate Professor of Genetics
North Carolina State University

Kenneth M. Rice

Associate Professor of Biostatistics
University of Washington

Ali Shojaie

Assistant Professor of Biostatistics
University of Washington

Matthew Stephens

Professor of Statistics and Human
Genetics
University of Chicago

John D. Storey

Professor of Genetics and Statistics
North Carolina State University

Jeffrey L. Thorne

Professor of Genetics and Statistics
North Carolina State University

Timothy Thornton

Assistant Professor of Biostatistics
University of Washington

Peter M. Visscher

Professor of Quantitative Genetics
University of Queensland

Jonathan C. Wakefield

Professor of Statistics and Biostatistics
University of Washington

J. Bruce Walsh

Professor of Ecology & Evolutionary
Biology
University of Arizona

Bruce S. Weir

Professor and Chair of Biostatistics
University of Washington
Director, Summer Institute in Statistical
Genetics

John S. Witte

Professor of Epidemiology
and Biostatistics
University of California, San Francisco

Daniela Witten

Assistant Professor of Biostatistics
University of Washington

Naomi Wray

Associate Professor of Statistical
Genetics
University of Queensland

N. David Yanez

Associate Professor of Biostatistics
University of Washington

Zhao-Bang Zeng

Reynolds Distinguished Professor of
Statistics and Genetics
North Carolina State University

Modules

Module 1: Probability and Statistical Inference

Instructors: Jim Hughes and David Yanez

This module covers the laws of probability and the binomial, multinomial, and normal distributions. It covers descriptive statistics and methods of inference, including maximum likelihood, confidence intervals and simple Bayes methods. Classical hypothesis testing topics, including type I and II errors, two-sample tests, chi-square tests and contingency table analysis, and exact and permutation tests. Resampling methods, such as the bootstrap and jackknife, are covered as well. This module serves as a foundation for almost all of the later modules. Co-taught with the Summer Institute in Statistics and Modeling in Infectious Diseases.

Module 2: Forensic Genetics

Instructors: John Buckleton and Bruce Weir

Although the use of genetic profiles is now routine, there are still issues for which care must be taken when attaching statistics to matching profiles. These include the effects of relatives and population structure, the use of lineage markers and the interpretation of mixtures. An introductory account will be given of the use of likelihood ratios, familial searching, the “birthday problem” and Y-STR profiles. A discussion of the use of peak heights and allelic dropout will be given.

Module 3: Molecular Genetics and Genomics

Instructors: Josh Akey and Greg Gibson

The molecular genetics and genomics 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. We start with the key concepts of quantitative and Mendelian genetics and then illustrate how these have been reconciled with molecular biology. Two half days are then spent on the basics of genome-wide association mapping as well as exome and whole genome sequencing; and on gene expression profiling and integrative genomics leading to systems biology. On the final afternoon the instructors provide their perspectives on the future of personalized medicine, and on evolutionary genetics.

Module 4: Bayesian Statistics for Genetics

Instructors: Peter Hoff and Jon Wakefield

The use of Bayesian methods in genetics has a long history. In this introductory module we will begin by discussing introductory probability. We will then describe Bayesian approaches to binomial proportions, multinomial proportions, two-sample comparisons (binomial, Poisson, normal), the linear model, and Monte Carlo methods of summarization. Advanced topics will be touched on, including 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. Background Reading: P.D. Hoff (2009). *A First Course in Bayesian Statistical Methods*. Springer-Verlag.

Module 5: Regression and Analysis of Variance

Instructors: Rebecca Hubbard and Lurdes Inoue

This module is designed as a foundation for the quantitative genetics and QTL modules as well as for the association mapping modules. It assumes the material in Module 1 and it will cover the basic commands in R. It covers linear regression and analysis of variance. This module includes both lectures and interactive data analysis using R. Specific topics discussed are: simple linear regression; multiple linear regression; residual analysis; transformations; one-way ANOVA; two-way ANOVA; analysis of covariance; multiple comparisons.

Module 6: Introduction to R

Instructors: Ken Rice and Tim Thornton

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. Examples and exercises will use data drawn from biological and medical applications, including infectious diseases and genetics. **Hands-on use of R is a major component of this module; users require a laptop and will use it in all sessions.** Co-taught with the Summer Institute in Statistics and Modeling in Infectious Diseases.

Module 7: Gene Expression Profiling

Instructors: Greg Gibson and John Storey

The gene expression module will cover the theory and application of transcriptomics, including both microarray and RNA-Seq methodologies. The focus of the module is on the statistical basis of hypothesis testing, covering the central role of normalization strategies with the opportunity for students to work examples using the SNM module in R, as well as the fundamentals of ANOVA and the False Discovery Rate procedure. In addition, we will discuss options for downstream processing by clustering and module detection, finishing with expression QTL analysis and integrative genomics to infer pathway structure.

Module 8: Population Genetic Data Analysis

Instructors: Jérôme Goudet and Bruce Weir

This module serves as a foundation for many of the later modules. Estimates and sample variances of allele frequencies, Hardy-Weinberg and linkage disequilibrium, characterization of population structure with F-statistics. Relationship estimation. Statistical genetic aspects of forensic science and association mapping. Concepts illustrated with R exercises. Background reading: Holsinger, K. and B.S. Weir. 2009. Genetics in geographically structured populations: defining, estimating, and interpreting FST. *Nature Reviews Genetics* 10:639–650. Weir, B.S. and C.C. Laurie. 2011. Statistical genetics in the genome era. *Genetics Research* 92:461–470.

Modules

Module 9: Elements of R for Genetics and Bioinformatics

Instructors: Thomas Lumley and Ken Rice

This module introduces programming skills required for analysis of genetic data, in the R statistical environment. The module assumes prior knowledge of R, to the level described in Module 6. We will briefly review how R scripts are built up from interactive commands, and then discuss how to turn these into R programs based on user-defined functions. We will cover R's debugging system, its tools for enhancing efficiency of code, and its methods for handling errors and warnings. In many genomic applications, R packages already exist to perform specialized statistical and bioinformatic analyses, and we will introduce packages for handling large datasets, and the Bioconductor repository of genomic packages. **Users require a laptop and will use it in all sessions.**

Module 10: Population Genetics and Association Mapping

Instructors: Katie Kerr and Tim Thornton

This module assumes the material in Module 1 and it serves as the foundation for many later modules. Topics covered include: basic probability and Mendelian genetics; Hardy-Weinberg equilibrium; inbreeding coefficients; population structure; recombination and genetic linkage; linkage disequilibrium; measures of relatedness; haplotype frequency estimation with unphased genotypes, genetic association testing; association testing in the presence of population structure and/or relatedness. Many methods are illustrated with implementation in R, and the module is most useful for students with basic familiarity with R. Other public domain software will be introduced, such as HAPLOVIEW, LOCUSZOOM, and PLINK. There is substantial overlap with Module 8. Background reading: Weir, B.S. 1996. Genetic Data Analysis II. Sinauer Associates; Thornton and McPeck. 2010 ROADTRIPS: Case-Control Association Testing with Partially or Completely Unknown Population and Pedigree Structure. American Journal of Human Genetics 86:172–184.

Module 11: Quantitative Genetics

Instructors: Bill Muir and Bruce Walsh

Assumes the material in Modules 1, 5 and 8. Provides a foundation for modules 14, 15 and 23. 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, and all microarray data, a working knowledge of quantitative genetics is critical in diverse fields from plant and animal breeding, human genetics, genomics, to ecology and evolutionary biology. The course will cover the basics of quantitative genetics including: Fishers variance decomposition, covariance between relatives, heritability, inbreeding and crossbreeding, and response to selection. Also an introduction to advanced topics such as: Mixed Models, BLUP, QTL mapping; correlated characters; and the multivariate response to selection. Background reading: Lynch, M. and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer Associates.

Module 12: Molecular Phylogenetics

Instructors: Joe Felsenstein, Mark Holder and Jeff Thorne

Assumes the material in Modules 1 and 8. Overview of methods for analysis of interspecific DNA and protein sequence data. Coverage will include parsimony, maximum likelihood, distance-based, and Bayesian methods for phylogenetic estimation. Probabilistic models for sequence change will be emphasized. Related topics that will be presented are the comparative method, divergence time estimation, phylogenetic hypothesis testing, and detection of positive selection. Statistical methodology will be a focus and some related computational algorithms will be outlined. Brief introductions will be made to software packages such as PAUP*, Beast, PHYLIP, and MrBayes. Background readings: Felsenstein, J. (2004) "Inferring Phylogenies." Sinauer Associates, or Yang, Z. (2006) "Computational Molecular Evolution (Oxford Series in Ecology and Evolution)." Oxford University Press.

Module 13: Advanced R Programming for Bioinformatics

Instructors: Thomas Lumley and Ken Rice

This module covers techniques needed for software development in R. The module includes advanced graphics, object-oriented programming, SQL database use, some of the Bioconductor infrastructure, writing packages, and calling C code from R. The module is aimed at people who have either substantial R experience or programming experience in other languages; Module 9 without some experience would not be sufficient preparation. Background reading: Gentleman, R. (2008) R Programming in Bioinformatics. Taylor & Francis. **Users require a laptop and will use it in all sessions.**

Module 14: QTL Mapping

Instructors: Rebecca Doerge and Zhao-Bang Zeng

Assumes the material in Modules 1, 5 and 11. Material in Module 8 would be helpful. This module will systematically introduce statistical methods for mapping quantitative trait loci (QTL) in experimental cross populations. Topics include experimental designs, linkage map construction, single-marker analyses, interval mapping, composite interval mapping and multiple interval mapping. Significance thresholds for genome scan and model selection will also be discussed. Uses public domain software Windows QTL-Cartographer for computer lab exercises. Emphasis is on procedures for QTL mapping data analysis and appropriate interpretation of mapping results rather than on formulas. Users require a laptop and will use it in all sessions.

Module 15: Mixed Models in Quantitative Genetics

Instructors: Bill Muir and Bruce Walsh

The analysis of linear models containing both fixed and random effects. 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, direct and associative effects models of general group and kin selection, genotype by environment interaction models. Background reading: Lynch, M. and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer Associates.

Modules

Module 16: Genetic Epidemiology

Instructors: Karen Edwards and Carolyn Hutter

Assumes the material in Models 1 and 8. This module will provide an overview of genetic epidemiology, with a focus on design, analysis and interpretation in studies of complex disease. The focus is on methods for discovering how genetic factors influence health and disease. Topics covered will include twin studies, family studies, segregation analysis, linkage analysis, population-based association studies, and gene-environment interactions. The course will include practical examples using on-line resources and statistical analysis programs. Background reading: M.A. Austin (editor). 2013. Genetic Epidemiology: methods and applications.

Module 17: MCMC for Genetics

Instructors: Eric Anderson and Matthew Stephens

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 as well as some population genetics. 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 practicals 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. Software used: R, STRUCTURE, PHASE. Background reading: Shoemaker, J.S., I.S. Painter and B.S. Weir. (1999). Bayesian statistics in genetics. Trends in Genetics 15:354–358. Beaumont, M.A. and B. Rannala. (2004). The Bayesian revolution in genetics. Nature Reviews Genetics 5:251–261. Gilks, W.R., S. Richardson and D.J. Spiegelhalter. (1996). “Markov Chain Monte Carlo in Practice.” Chapman and Hall.

Module 18: High-Dimensional Omics Data

Instructors: Ali Shojaie and Daniela Witten

In this course, we will present a number of statistical machine learning methods for the analysis of high-dimensional biological data, often referred to as ‘omics.’ Examples include genomic, transcriptomic, metabolomic, proteomic, and other large-scale data sets, typically characterized by a huge number of molecular measurements (such as genes) and a relatively small number of samples (such as patients). In the first part of the course, we will cover supervised learning methods that are useful in the analysis of omics data. These include penalized approaches for performing regression, classification, and survival analysis in the high-dimensional setting. In the second part of the course, we will discuss unsupervised approaches for the analysis of omics data,

such as clustering and principal components analysis. Throughout the course, we will highlight the effects of high dimensionality and focus on common pitfalls in the analysis of omics data, and how to avoid them. The techniques discussed will be demonstrated in R. This course assumes a previous course in regression and statistical hypothesis testing, and some familiarity with R or other command line programming languages. **Users require a laptop and will use it in all sessions.**

Module 19: Statistical & Quantitative Genetics of Disease

Instructors: John Witte and Naomi Wray

This module focuses on the analysis of genetic data for disease and its interpretation. We will consider in detail the quantitative genetics of binary disease with emphasis on the equivalences and relationships between different models. We will contrast and synthesize the traditional viewpoints of quantitative geneticists and epidemiologists. Topics will include: risk models on different scales the observed (or disease) scale, the liability threshold scale, the log risk scale; estimation of heritability from familial risk ratios; estimation of the contribution of individual and multiple risk loci to disease; estimation of variance attributable to genome-wide SNPs individually and together; and risk profile scoring. Some of the underlying methods can be traced to classic papers, however, whole genome methods applied to SNP data have been recently developed. There are no prerequisite modules, although modules 6 and 8 may be helpful. Basic R programming, matrix algebra and statistical methods are assumed.

Module 20: Plant and Animal Association Mapping

Instructors: Michel Georges and Dahlia Nielsen

This module is an introduction to association mapping, focusing on plant and animal populations. Topics include theory of linkage disequilibrium and mapping, population and family-based association techniques for discrete and continuous traits, methods for detecting and accounting for population structure, genomic selection, issues in polyploid species, and multiple testing issues. We will also discuss developing genome resources, with a focus on next-gen sequencing strategies. Assumes a basic background in genetics and genomics, such as the material covered in Module 3.

SISMID Module 13: Introduction to Metagenomic Data Analysis

Instructors: Alexander V. Alekseyenko and Paul J. McMurdie

Module description: This course is concerned with analysis of microbial community data generated by next-generation sequencing technologies. These high-throughput methods allow for deep surveying of microorganisms inhabiting their biological hosts. We will cover the steps for preprocessing Roche 454 sequencing data, necessary to produce abundance tables. We will then examine methodology for associating microbial abundance data with experimental factors and outcomes. Programming will be done in R, and overview of available programs for preprocessing the data will be provided. Pre-requisites: Module 1: Probability and Statistical Inference. Co-listed with Summer Institute in Statistics and Modeling of Infectious Diseases.

Modules

SISMID Module 14: Evolutionary Dynamics and Molecular Epidemiology of Viruses

Instructors: Philippe Lemey and Marc A. Suchard

Module description: This module covers the use of phylogenetic and bioinformatic tools to analyze pathogen genetic variation and to gain insight in the processes that shape their diversity. The module focuses on phylogenies and how these relate to population genetic processes in infectious diseases. In particular, the module will cover Bayesian Evolutionary Analysis by Sampling Trees (BEAST, <http://beast.bio.ed.ac.uk/>). This software will be used in class exercises that are mainly focused on estimating epidemic time scales, reconstruction changes in viral population sizes through time and inference of spatial diffusion of viruses. Evolutionary processes including recombination and selection will also be considered. Assumes material in Module 1. Co-listed with Summer Institute in Statistics and Modeling of Infectious Diseases.

Module 21: Coalescent Theory

Instructors: Philip Awadalla and Mary Kuhner

This module is an introduction to the coalescent and its applications to modern population genetics and genomics. Assumes material in Modules 1 and 3. Material in Module 8 or 10 and 13 would be helpful. Derivation and properties of basic coalescent model and extension to include factors such as recombination, geographic structure and natural selection. Use of the coalescent in analyzing data for disease gene mapping, recombination rate estimation, and detection of recent adaptive evolution. Use of coalescent methodologies in large-scale surveys of genetic variation. Applications to standard or next-generation sequencing data for inferences from natural populations and disease cohorts. Use of public domain software.

Module 22: Network and Pathway Analysis of Omics Data

Instructor: Alison Motsinger-Rief and Ali Shojaie

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 will be 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 a previous course in regression, previous exposure to the material covered in Module 18, and familiarity with R or other command line programming languages. Users require a laptop and will use it in all sessions.

Module 23: Advanced Quantitative Genetics - NEW!

Instructors: Mike Goddard and Peter Visscher

This module assumes the material in Modules 1 and 6. Material in Modules 9, 11 and 15 would be helpful. This module focuses on the genetics and analysis of quantitative traits in human populations, with emphasis on estimation and prediction analysis using genetic markers. It is a good match with Module 19 that deals with similar topics but with a focus on disease (binary) outcomes. 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. For computer exercises, we will be using R, the Merlin suite of software, PLINK GCTA (www.complextraitgenomics.com/software/gcta/). Background reading: Vinkhuyzen et al. 2013. Estimation and partition of heritability in human populations using whole-genome analysis methods. *Annu Rev Genet.* 23;47:75-95. doi: 10.1146/annurev-genet-111212-133258. Yang, J. et al. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* 42: 565-569. Wray NR et al. 2013. Pitfalls of predicting complex traits from SNPs. *Nature Review Genetics* 14:507-15. doi: 10.1038/nrg3457.

Module 24: Ethics for Statistical Geneticists - NEW!

Instructors: Malia Fullerton and Nanibaa' Garrison

Statistical approaches to the study of genetic and genomic data entail a range of study design and analytical decisions with potential ethical implications. In this course, we will discuss issues that may arise in the conduct of human genomic research, including the inclusion of diverse populations and generalizability of study findings, data integrity and reproducibility, privacy and potential for re-identification, data sharing, result dissemination, return of individual results, clinical translation, and forensic applications. Relevant ethical principles and professional guidelines will be introduced and discussed. Background reading: Ioannidis, JP. 2005. Why most published research findings are false. *PLoS Med.* 2005 Aug;2(8):e124.

Daily Schedule

Time	Daily Activity
8:00 am – 8:30 am	Coffee (and Registration Check-In on Mondays)
8:30 am – 10:00 am	Class Session
10:00 am – 10:30 am	Break
10:30 am – 12:00 pm	Class Session
12:00 pm – 1:30 pm	Lunch (and Registration Check-In on Wednesdays)
1:30 pm – 3:00 pm	Class Session
3:00 pm – 3:30 pm	Break
3:30 pm – 5:00 pm	Class Session
5:00 pm – 6:00 pm	Participants' and Instructors' Reception (Mondays and Wednesdays)
5:00 pm – 6:00 pm	Tutorials (Tuesdays and Thursdays - optional)

Calendar

Monday (8:30 am–6 pm) July 7	Tuesday (8:30 am–7 pm) July 8	Wednesday (8:30 am–noon) July 9	Wednesday (1:30 pm–6 pm) July 9	Thursday (8:30 am–7 pm) July 10	Friday (8:30 am–5 pm) July 11
<i>Week 1, Session 1</i> Mod 1: Probability and Statistical Inference * Mod 2: Forensic Genetics Mod 3: Molecular Genetics and Genomics Mod 4: Bayesian Statistics for Genetics			<i>Week 1, Session 2</i> Mod 5: Regression and Analysis of Variance Mod 6: Introduction to R * Mod 7: Gene Expressing Profiling Mod 8: Population Genetic Data Analysis		
July 14	July 15	July 16	July 16	July 17	July 18
<i>Week 2, Session 1</i> Mod 9: Elements of R Mod 10: Population Genetics & Association Mapping Mod 11: Quantitative Genetics Mod 12: Molecular Phylogenetics			<i>Week 2, Session 2</i> Mod 13: Advanced R Programming in Bioinformatics Mod 14: QTL Mapping Mod 15: Mixed Models in Quantitative Genetics Mod 16: Genetic Epidemiology		
July 21	July 22	July 23	July 23	July 24	July 25
<i>Week 3, Session 1</i> Mod 17: MCMC for Genetics Mod 18: High-Dimensional Omics Data Mod 19: Statistical & Quantitative Genetics of Disease Mod 20: Plant and Animal Association Mapping SISMID 13: Introduction to Metagenomics Analysis † SISMID 14: Evolutionary Dynamics and Molecular Epidemiology of Viruses †			<i>Week 3, Session 2</i> Mod 21: Coalescent Theory Mod 22: Pathway & Network Analysis for Omic Data Mod 23: Advanced Quantative Genetics Mod 24: Ethics for Statistical Geneticists		

Only one (1) module may be taken for each 2.5-day block.

* Co-taught with Summer Institute in Statistics and Modeling in Infectious Diseases (SISMID 2014).

† SISMID Modules 13 and 14 are co-listed with SISG 2014, and their module descriptions are listed in this brochure for your convenience. *Registration for these two SISMID 2014 modules is only available via SISMID's website.* Please be sure to avoid schedule conflicts between the two summer institutes.

SISMID 2014 will be held concurrently with SISG 2014, 7-23 July 2014, at the University of Washington. Information about SISMID 2014 can be found on its conference website at: depts.washington.edu/sismid. If you have specific questions about SISMID 2014, please send an email to: sismid@uw.edu.

SISG 2014
c/o UW BIOSTATISTICS
4333 Brooklyn Avenue NE
UW Tower 15T-420
Box 359461
Seattle, WA 98195-9461

19th Summer Institute
in Statistical Genetics
University of Washington
7-25 July 2014

19th Summer Institute in Statistical Genetics

7–25 July 2014

University of Washington
Seattle, Washington
sisg.biostat.washington.edu



Sponsored by:
Department of Biostatistics
School of Public Health
University of Washington