17th Summer Institute in Statistical Genetics

July 9-27, 2012

UNIVERSITY *of* WASHINGTON

Seattle, Washington

www.biostat.washington.edu

AGTGAAGCTACTACTTAGAAAGTACTGCTACTGGTGAAAAT

SISG
2012

Sponsored by:

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

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

Administrative Details

General Information

The 17th Summer Institute in Statistical Genetics (SISG 2012) will be held at the University of Washington in Seattle, Washington. A map showing this location is on the **Institute website: http://sisg.biostat.washington.edu**.

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 2012 there will be two concurrent Institutes in Seattle. Details about the Summer Institute in Statistics and Modeling for Infectious Diseases in Seattle, 9-25 July, are available online: http://depts.washington.edu/sismid/. Registration for each Institute is available only online via each Institute's website. Also, the University of Edinburgh will host the European Institute in Statistical Genetics (EISG 2012), to run in the weeks on either side of the International Conference in Quantitative Genetics, 13-15 June and 25-29 June 2012. Details for EISG 2012 can be found online: http://www.eisg2012.org.uk/

Registration Deadlines and Fees

Registration fee per module: $650 (USD); Early-bird rate: $550 (USD).

Reduced academic and government registration fee per module: $500 (USD); Early-bird rate: $400 (USD).

Early-bird deadline is Monday, 11 June 2012. Participation in any module cannot be guaranteed for registrations received after Monday, 11 June 2012.

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 of this brochure *(please refer to the Institute's website for updated module descriptions)*.

Refund Policy: A $200 processing fee will be deducted from refunds requested after Monday, 25 June 2012. No refunds will be processed after Monday, 2 July 2012.

Payment can be made with Diners Club, Visa or Mastercard credit cards online via a secure server. Mailed payments to the mailing address shown below can be made with a purchase order (U.S. companies and organizations only) or by check or money order in U.S. dollars drawn on a U.S. bank. Checks should be made payable to the University of Washington. For wire-transfers, please refer to your Registration Invoice for instructions.

Scholarships

Some registration-fee and travel scholarships are available for students. Applications are due on or before **Friday, 30 March 2012**, and applicants will be notified by Friday, 13 April 2012. Applicants should send a letter explaining their reason for wishing to attend the Institute, the Scholarship Cover Sheet listing the modules they wish to attend, a one-page CV, and a letter from an advisor or supervisor via email to: sisg@uw.edu. Applications can also be via regular mail or air courier to the mailing addresses shown below. **Applicants must also Register for the modules and check the "Yes" button to the question: "Are you applying for a Summer Institute scholarship?" so that payment will not be required at time of registration and to receive a Scholarship Cover Sheet.** 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 to bring laptop computers with them. They will have free online access while they are on the University of Washington campus. Participants will receive USB-drives with copies of software and datasets when they arrive at the Institute. It is suggested, however, that participants follow the online instructions at the Institute website to download software and data before they arrive.

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 2012 c/o UW BIOSTATISTICS
4333 Brookyn Avenue NE
UW Tower 15T-410
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 Applied Mathematics and Statistics
University of California at Santa Cruz

Philip Awadalla
Genome Quebec Chair of Population and Medical Genomics
University of Montreal

Peter J. Bradbury
Computational Biologist, ARS, USDA
Cornell University

Elias Chaibub Neto
Research Fellow
Sage Bionetworks

Rebecca W. Doerge
Professor of Statistics
Purdue University

Joseph Felsenstein
Professor of Genome Sciences and Biology
University of Washington

Christopher Gaiteri
Research Fellow
Sage Bionetworks

Gregory C. Gibson
Professor, School of Biology
Georgia Institute of Technology

Stephanie Gogarten
Senior Research Scientist
University of Washington

Jérôme Goudet
Associate Professor of Population Genetics
University of Lausanne

Justin Guinney
Research Fellow
Sage Bionetworks

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

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

Cathy C. Laurie
Senior Principal Research Scientist
University of Washington

Thomas Lumley
Professor of Statistics
University of Auckland

Alison Motsinger-Reif
Assistant Professor of Statistics
North Carolina State University

William M. Muir
Professor of Genetics
Purdue University

John Novembre
Assistant Professor of Ecnology and Evolutionary Biology
University of California Los Angeles

Kenneth M. Rice
Associate Professor of Biostatistics
University of Washington

Ali Shojaie
Assistant Professor of Biostatistics
University of Washington

Solveig Sieberts
Director of Statistical Genetics
Sage Bionetworks

John D. Storey
Associate Professor, Lewis-Sigler Institute,
Department of Molecular Biology
Princeton University

Elizabeth A. Thompson
Professor and Chair of Statistics
University of Washington

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, University of Queensland
Diamantina Institute

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

Daniela Witten
Assistant Professor of Biostatistics
University of Washington

Brian S. Yandell
Professor of Statistics
University of Wisconsin

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: J. Hughes and D. 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.

Module 2: Computing for Statistical Genetics
Instructors: T. Lumley and K. Rice
This module introduces software for analysis of genetic data, in the R statistical environment. Data management in R, programming concepts for R, and standard regression analyses will be discussed. These topics will be followed by analysis more specific to genetic data, including association analysis, and handling large date files. Use of the extensive collection of genomics packages from the Bioconductor project will be introduced. Finally, the use of R as an interface to other more specialized, 'legacy' software will be demonstrated. Reference will be made to current analyses of whole-genome association study data. This module assumes no prior knowledge of R. It will provide a foundation for computation for later modules.

Module 3: Bayesian Statistics for Genetics
Instructors: P. Hoff and J. 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 4: Regression and Analysis of Variance
Instructors: R. Hubbard and L. 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 5: Molecular Genetics and Genomics
Instructors: J. Akey and G. Gibson
This module provides an overview of the basic principles of molecular genetics, but also incorporates an introduction to the latest genomic approaches. Starts with the laws of Mendelian inheritance and the roles of DNA and RNA as genetic material, discuss mutations and transmission genetics, and moves on to describe the foundations of population and quantitative genetics. This course builds the necessary concepts and introduces the methodologies that will enable students to take more detailed modules dealing with the structure and distribution of molecular variation, linkage and association studies for dissection of quantitative traits, phylogeny reconstruction, and gene or protein expression profiling. Also touches on such topics as comparative genomics, mutational genetic analysis, and regulation of gene expression. Recommended text: Gibson, G. & S. Muse. (2009). "A Primer of Genome Science." 3rd edition, Sinauer Associates.

Module 6: Population Genetic Data Analysis
Instructors: J. Goudet and B. Weir
This module overlaps substantially with Module 8. It 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 Weir, B.S. 2009. Genetics in geographically structured populations: defining, estimating, and interpreting FST. Nature Reviews Genetics 10:639–650. Weir, B.S. and Laurie, C.C. 2011. Statistical genetics in the genome era. Genetics Research 92:461–470.

Module 7: Quantitative Genetics
Instructors: W. Muir and B. Walsh
Assumes the material in Modules 1, 4 and 5. Provides a foundation for modules 11, 12 and 18. 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 cross-breeding, 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 Walsh, B. 1998. Genetics and analysis of quantitative traits. Sinauer Associates.

Module 8: Population Genetics and Association Mapping
Instructors: K. Kerr and T. Thornton
This module overlaps substantially with Module 6. It 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 concepts are illustrated with public domain software such as R and HAPLOVIEW. Background reading: Weir, B.S. (1996). "Genetic Data Analysis II." Sinauer Associates; Thornton and McPeek (2010) "ROADTRIPS: Case-Control Association Testing with Partially or Completely Unknown Population and Pedigree Structure." American Journal of Human Genetics 86:172–184.

Module 9: Gene Expression Profiling
Instructors: G. Gibson and J. Storey
This course covers all aspects of the statistical analysis of gene expression profiling; the methods are also relevant to analysis of proteomic and metabolomic data. Theory will be integrated with case studies demonstrating the principles of quality control,

normalization, analysis of variance and hypothesis testing, time series, surrogate variable analysis, and optimal discovery procedures. Discussion will include microarray and nextgen sequencing applications, downstream data-mining and network analysis approaches, and relevant statistical software will be demonstrated.

Module 10: MCMC for Genetics
Instructors: E. Anderson and J. Novembre
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., Painter, I.S. and Weir, B.S. (1999). Bayesian statistics in genetics. Trends in Genetics 15:354–358. Beaumont, M.A. and Rannala, B. (2004). The Bayesian revolution in genetics. Nature Reviews Genetics 5:251–261. Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1996). "Markov Chain Monte Carlo in Practice." Chapman and Hall.

Module 11: Introduction to QTL Mapping
Instructors: R. Doerge and Z-B. Zeng
Assumes the material in Modules 1,4,5 and 8. Material in Modules 6 or 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.

Module 12: Mixed Models in Quantitative Genetics
Instructors: W. Muir and B. 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 Walsh, B. 1998. Genetics and analysis of quantitative traits. Sinauer Associates.

Module 13: Molecular Phylogenetics
Instructors: J. Felsenstein, M. Holder and J. Thorne
Assumes the material in Modules 1 and 5. 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 14: Inference of Relationships and Relatedness
Instructors: E. Anderson and E. Thompson
This module focuses on methods for inferring relationships and relatedness between individuals in natural populations using multilocus genetic data. Emphasis is given to applications in managed or endangered populations of plant and animal species. Topics covered in the underlying theory include: gene identity by descent (ibd) versus gene identity in state (iis); calculation of probabilities of gene ibd conditional on relationships and genetic data; coefficients of inbreeding and kinship; information gain by considering joint relationships of additional (more than two) relatives; linked loci, genome scans, and the lengths of chromosomal ibd segments. Estimation problems covered include: estimation of pairwise relatedness and relationships; parentage and paternity inference and pedigree reconstruction in natural populations; inference of sibling groups in the absence of parental information; relationship inference and validation from linked loci; the estimation of population mixtures and hybrid individuals. The focus will be primarily on likelihood and Bayesian methods of estimation. This module assumes knowledge of the material in basic statistics (module 1) and basic population genetics (modules 6 or 8). Module 10 would be helpful, but it is not a strict prerequisite.

Module 15: Systems Genetics for Experimental Crosses
Instructors: E. Chaibub Neto and B. Yandell
This module will take a holistic, technical look at systems genetics for experimental crosses. The field of systems genetics, also known as "genetical genomics," views "omics" molecular phenotypes such as mRNA expression, protein and metabolite levels as quantitative traits, amenable to quantitative genetical analyses. We begin with model selection for the genetic architecture of a single trait, building on the Introduction to QTL Mapping module. We then address QTL mapping of multiple correlated traits, modeling the correlation structure of the traits. Extensions of the Churchill-Doerge permutation tests are developed to assess the legitimacy of alleged QTL hotspots, i.e., loci where many traits have LOD peaks. The remainder of the module concerns causal phenotype models driven by QTL. For pairs of molecular phenotypes mapping to the same locus, we compare models using one trait as a covariate of the other to assess the causal ordering among the phenotypes. This analysis is used to identify key drivers of subsets of traits, i.e., traits that seem to have a causal effect on most of the other co-mapping traits. The key drivers and their co-mapping traits are then organized into causal phenotype networks. Finally, we show how biological pathway information, coming from GO, KEGG, TF or PPI databases can improve causal models. This module presumes material covered in modules 1, 2, 4 and 11.

Modules

Module 16: Coalescent Theory
Instructors: P. Awadalla and M. Kuhner
This module is an introduction to the coalescent and its applications to modern population genetics and genomics. Assumes material in Modules 1 and 5. Material in Module 6 or 8 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 17: High-Dimensional Omics Data
Instructors: A. Shojaie and D. Witten
In this course, we will cover 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 half 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 half of the course, we will discuss unsupervised approaches for the analysis of omics data, such as clustering, principal components analysis, and network estimation techniques. 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.

Module 18: Human Quantitative Genetics
Instructors: P. Visscher and B. Weir
This module assumes the material in Modules 1 and 6 or 8. Material in Module 7 would be helpful. A quantitative genetic framework for association mapping. Topics include: genetic correlations for individuals and for traits; Haseman-Elston regression for linkage analysis for quantitative traits; experimental design; estimating genetic variance associated with genome-wide identity by descent; estimation of heritability. The use of relatives to estimate heritability within families. The use of GWAS data to estimate the contribution of all SNPs simultaneously. Background reading: Visscher, P.M. et al. (2007) Genome partitioning of genetic variation for height from 11,214 sibling pairs. American Journal of Human Genetics 81:11041110 (2007); Visscher, P.M. 2009. Whole genome approaches to quantitative genetics. Genetica 136:351–357; Weir, B.S. 2008. Linkage disequilibrium and association tests. Annual Reviews of Genomics and Human Genetics 9:129–142.

Module 19: Advanced R Programming for Bioinformatics
Instructors: T. Lumley and K. Rice
This module covers object-oriented programming, SQL database use, some of the Bioconductor data infrastructure, 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 2 would not be sufficient preparation. Background reading: Gentleman. R. (2008) R Programming in Bioinformatics. Taylor & Francis.

Module 20: GWAS Data Cleaning
Instructors: S. Gogarten and C. Laurie
Genome-wide Association Studies need to take care with genotypic data quality in order to maximize power and reduce false positives. The GENEVA Coordinating Center at the University of Washington has developed a comprehensive protocol that addresses issues of sample and SNP quality. The process begins with formatting SNP intensity and annotation data in the netCDF format and then using the R packages GWASTools and SNPRelate. The procedures include examination of: missing call rates; heterozygosity; gender and sex chromosome aneuploidy; relatedness estimation; principal component analysis to detect population stratification; Hardy-Weinberg testing; and basic association testing. Participants will apply these procedures to HapMap genotypic data. Basic proficiency in R is required (as in Module 2). Advanced R is not necessary, but would help participants who need to extend or modify the code in this module (see Module 19). Background Reading: Laurie et al. 2010. Quality control and quality assurance in genotypic data for genome-wide association studies. Genetic Epidemiology 34,591-602.

Module 21: Network and Pathway Analyses of Omics Data
Instructors: Gaiteri, Guinney, Motsinger-Reif and Sieberts
This module covers a range of commonly used and newly emerging data-mining approaches for genetic and genomic data analysis. Coexpression networks based on gene-gene correlations are a tool that can be used to sample many regulatory networks and provide a window into complex diseases. The network structure itself can be used to understand how biological functions are implemented and homeostasis is maintained. We will review strategies for leveraging coexpression network structure to detect the collective influence of multiple contributing factors in complex diseases. We will discuss both pattern recognition and dimensionality reduction approaches, and will discuss details of highly successful methods like Classification and Regression Trees, Random Forest, and Multifactor Dimensionality Reduction. Throughout the course we will cover general issues with data-mining approaches such as variable selection, hypothesis testing, multiple comparisons, and predictive modeling. Finally, we will cover pathway-based analyses both in the gene expression study and GWAS frameworks. Pathway-based analyses can be used to leverage biological knowledge available from literature, gene ontologies or previous experiments to identify the pathways associated with disease or outcome. This approach can be illuminating in the case where each individual gene or loci shows small-to-moderate association, which might not overcome the significance burden of multiple testing which accompanies high-dimensional -omics analyses. Software tools for implementing the analyses discussed will be emphasized.

Module 22: Plant and Animal Association Mapping
Instructors: P. Bradbury
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, issues in polyploid organisms, multiple testing issues, and genotyping strategies. Examples for real data, including a discussion of linkage disequilibrium in plant and animal populations. Hands-on experience with publicly available software packages, including TASSEL. Assumes material in Modules 1, 4 and 5. Material in Modules 6 or 8 would be useful.

Daily Schedule

Time	Daily Activity
8:00 am – 8:30 am	Coffee (and Registration 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 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 – 7:00 pm	Tutorials (Tuesdays and Thursdays - optional)

Calendar

Monday (8:30 am–6 pm)	Tuesday (8:30 am–7 pm)	Wednesday (8:30 am–noon)	Wednesday (1:30 pm–6 pm)	Thursday (8:30 am–7 pm)	Friday (8:30 am–5 pm)
July 9	July 10	July 11	July 11	July 12	July 13
Mod 1: Probability and Statistical Inference Mod 2: Computing for Statistical Genetics Mod 3: Bayesian Statistics for Genetics			Mod 4: Regression and Analysis of Variance Mod 5: Molecular Genetics and Genomics Mod 6: Population Genetic Data Analysis		
July 16	July 17	July 18	July 18	July 19	July 20
Mod 7: Quantitative Genetics Mod 8: Population Genetics & Association Mapping Mod 9: Gene Expressing Profiling Mod 10: MCMC for Genetics			Mod 11: Introduction to QTL Mapping Mod 12: Mixed Models in Quantitative Genetics Mod 13: Molecular Phylogenetics Mod 14: Inference of Relationships and Relatedness		
July 23	July 24	July 25	July 25	July 26	July 27
Mod 15: Systems Genetics for Experimental Crosses Mod 16: Coalescent Theory Mod 17: High-Dimensional Omics Data Mod 18: Human Quantitative Genetics Mod 19: Advanced R Programming for Bioinformatics			Mod 20: GWAS Data Cleaning Mod 21: Network & Pathway Analyses of Omics Data Mod 22: Plant and Animal Association Mapping		

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

SISG 2012
c/o UW BIOSTATISTICS
4333 Brooklyn Avenue NE
UW Tower 15T-410
Box 359461
Seattle, WA 98195-9461

17th Summer Institute
in Statistical Genetics
University of Washington
9-27 July 2012

17th Summer Institute in Statistical Genetics

9–27 July 2012

University of Washington
Seattle, Washington

http://sisg.biostat.washington.edu

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