

Workshop

Applied Statistical Genetics for Population-Based Association Studies

INSTRUCTORS:

Prof. Andrea Foulkes

Associate Professor of Biostatistics
University of Massachusetts, USA

Prof. Michael Greenacre

Professor of Statistics
Universitat Pompeu Fabra, Barcelona

October 13-14, 2011 • 9:00-18:00

Fundación BBVA

Palacio del Marqués de Salamanca

Paseo de Recoletos, 10

28001 MADRID

Applied Statistical Genetics for Population-Based Association Studies

PRESENTATION

Recent technological advancements, coupled with extensive genetic sequencing efforts, have led to an explosion in the availability of molecular and cellular level data for the study of complex diseases. The on-going success of large-scale genome-wide association studies (GWAS), as well as improved candidate-gene studies, has generated considerable interest among biomedical researchers in learning appropriate, well-vetted analytic approaches. At the same time, novel statistical methods have been developed to address the unique analytic challenges that arise in these settings. This workshop presents the theory and practical application of these new methodologies, while introducing fundamental concepts and analysis tools for population-based investigations of genotype-trait associations.

We begin by providing a general overview of genetic association studies and related genetic data concepts. Topics covered include types of investigations, linkage disequilibrium and Hardy-Weinberg equilibrium, as well as the role of population substructure on associated measures and tests. Further, three analytical challenges common to these settings are described, namely multiplicity, ambiguity in phase and high-dimensionality. Selected analytic approaches are presented that specifically address these challenges, including multiple testing adjustments, haplotype reconstruction methods and tree-based approaches for discovery of gene-gene and gene-environment interactions.

This course is intended for medical and public health investigators with elementary knowledge of statistical concepts at the level of a first course in statistics or biostatistics. Supplemental materials will be provided for the more advanced student.

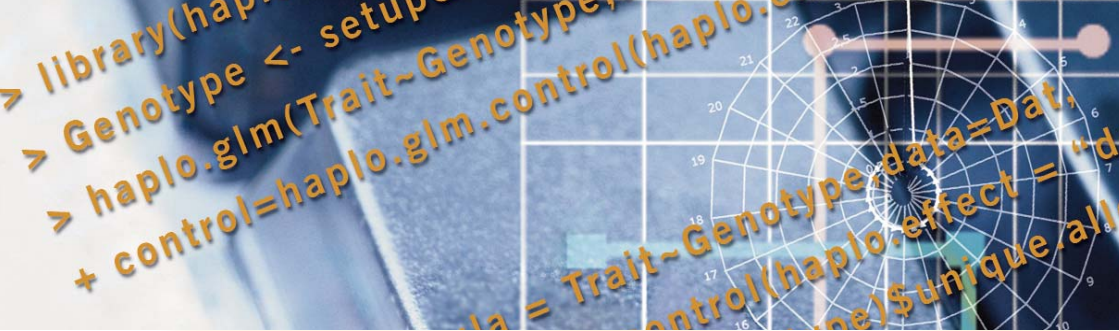
PROGRAM

October 13

- 9:00-10:30** Overview of population-based investigations • Data components and terminology • Public data examples • Analytic challenges
- 10:30-11:00** Coffee break
- 11:00-12:30** Linkage disequilibrium (LD) • Hardy-Weinberg Equilibrium (HWE) • Quality control and pre-processing • Examples and software
- 12:30-13:00** Coffee break
- 13:00-14:30** Measures of error • Single-step and step-down adjustments • Resampling-based methods • Alternative paradigms • Examples and software
- 14:30-15:30** Lunch
- 16:00-18:00** Practical session in R

October 14

- 9:00-10:30** Haplotype estimation • Estimating and testing haplotype-trait associations • Examples and software
- 10:30-11:00** Coffee break
- 11:00-12:30** Classification and Regression Trees (CART) • Random Forests (RF) • Examples and software
- 12:30-13:00** Coffee break
- 13:00-14:30** Logic Regression (LR) • Conditional Inference Trees (CIT) • Bayesian Variable Selection (BVS) • Examples and software
- 14:30-15:30** Lunch
- 16:00-18:00** Practical session in R



INSTRUCTORS

Prof. Andrea Foulkes

Andrea Foulkes is Associate Professor of Biostatistics and Head of the Biostatistics Program at the University of Massachusetts Amherst, where she is recognized for excellence in teaching. She serves as the lead statistician on multiple collaborative projects in HIV/AIDS and cardiovascular disease and is the principal investigator of an individual research award from the National Heart, Lung and Blood Institute, a division of the National Institutes of Health. Her research includes developing and applying methods to characterize the relationships among high-dimensional molecular and cellular level data and clinical measures of disease progression. She recently authored a graduate level text book on *Applied Statistical Genetics* for Springer's UserR Series.

Prof. Michael Greenacre

Michael Greenacre is Professor of Statistics at the Universitat Pompeu Fabra in Barcelona and scientific collaborator with the BBVA Foundation. His research interests are in the analysis of large data sets in the social and environmental sciences. He has authored and co-edited seven books and numerous journal articles on correspondence analysis and data visualization, the latest book being *Biplots in Practice*, published in 2010 by the BBVA Foundation and available online for free download at www.multivariatestatistics.org.

The BBVA Foundation expresses BBVA's active engagement with the societies where it does business, while upholding the two core elements of the Group's identity: namely, a stake on knowledge and innovation and the primacy of ethical principles and conduct. Its mission is to generate and disseminate knowledge and to foster the wellbeing of today's and tomorrow's society by supporting frontier scientific research and artistic creation at national and international level. The Foundation focuses on five strategic areas: Economy and Society, the Environment, Biomedicine and Health, Basic Sciences and Technology and Arts and Humanities. Across these areas, it designs, develops and funds projects and research chairs; facilitates advanced specialist training through courses, seminars and workshops; supports the organization of congresses, forums, encounters and scientific meetings; grants awards to individuals and institutions whose work has contributed significantly to the advancement of knowledge; and relays the results of its scientific and cultural projects to society through publications, music CDs and DVDs, concerts, lecture cycles, debates, exhibitions and other content in audiovisual and electronic format.

PREVIOUS BBVA FOUNDATION WORKSHOPS ON STATISTICS:

Structural Equation Modelling with LISREL,
by Karl Jöreskog and Dag Sörbom,
November 2004

Correspondence Analysis in Practice,
by Michael Greenacre and Antonietta Curci,
March 2005

Multivariate Analysis for Environmental Biologists (first edition),
by Michael Greenacre and Raul Primicerio,
November 2005

Multivariate Analysis for Environmental Biologists (second edition),
by Michael Greenacre and Raul Primicerio,
May 2006

Multivariate Statistical Modelling of Ecological Data (third edition),
by Michael Greenacre and Raul Primicerio,
April 2007

Statistical Learning (first edition),
by Trevor Hastie and Michael Greenacre,
April 2008

Multivariate Statistical Modelling of Ecological Data (fourth edition),
by Michael Greenacre and Raul Primicerio,
September 2008

Statistical Learning (second edition),
by Trevor Hastie and Michael Greenacre,
July 2009

Biplots in Practice,
by Michael Greenacre and Jan Graffelman,
October 2010

Course fees: companies, 125 euros / university staff and students, 75 euros
(includes course material, refreshments and lunches)

Registration: October 3 to 7, 2011

To register, please contact: registration@bbva.es

Simultaneous interpretation will be provided

