Imperial College London

EUROPEAN SOCIETY OF HUMAN GENETICS

Introduction to the statistical analysis of genome-wide association studies

DATE: 2 to 6 July 2018

VENUE: Section of Genomics of Common Disease
Imperial College London
Hammersmith Hospital Campus
Du Cane Road, London, W12 0NN, UK

AUDIENCE:

Geneticists facing the need to analyse large-scale human genotyping data in relation to their effect on common human traits and diseases

Scientists aiming to undertake genome-wide association studies and their meta-analyses

Researchers willing to understand better the statistical approaches and analytical procedures for the genetic association studies

COURSE LEADERS:

Inga Prokopenko, PhD, Senior Lecturer in Human Genomics, Imperial College London, London, UK

Andrew P. Morris, PhD, Professor of Statistical Genetics, Wellcome Trust Senior Research Fellow in Basic Biomedical Science, University of Liverpool, Liverpool, UK

Reedik Mägi, PhD, Senior Research Fellow, Head of Bioinformatics workgroup, Estonian Genome Center, University of Tartu, Tartu, Estonia

Krista Fischer, PhD, Senior Research Fellow, Estonian Genome Center, University of Tartu, Tartu, Estonia

Marika Kaakinen, PhD, PostDoctoral Marie Curie Research Fellow, Imperial College London, London, UK

INFO: https://www.imperial.ac.uk/department-of-medicine/study/short-courses/genomic-studies/

CONTACT: **GCD Office,** e-mail: gcdshort.courses@imperial.ac.uk
tel: +44 (0)207 594 1603
Dr. Inga Prokopenko, e-mail: i.prokopenko@imperial.ac.uk

Course programme:

Day one:

Introduction to statistics for geneticists - Dr. Inga Prokopenko and Dr. Marika Kaakinen
Basics of probability theory, binomial and normal distribution, polygenic inheritance and complex traits, allele frequencies in population.

Introduction to Linux and R – Ms. Natalia Pervjakova
Interface, command line and basic commands, functions, text editors, saving commands in scripts and running scripts, installing software tools for statistical analysis of genetic data, versions, data storage. Linux as environment for PLINK software tool. Basics of R usage to run graphical tools for genome-wide data and analysis results.

Introduction to genome-wise association studies (GWAS) – Dr. Marika Kaakinen
Principles of linkage disequilibrium (LD) and SNP tagging for genome-wide genotyping array design, analysis and imputation; haplotypes, study design, sample size and statistical power, use UCSC browser and EBI GWAS catalog,

Day two:

Quality Control (QC) for GWAS –Dr. Reedik Mägi
Sample and variant QC: for individual (sample) and variant missingness, gender checks, duplicates and cryptic relatedness, population outliers, heterozygosity and inbreeding, minor allele frequency, and Hardy-Weinberg equilibrium.

Invited Speaker:
TBC

Statistical models for genetic association analysis – Dr. *Krista Fischer*
Hardy-Weinberg equilibrium, linear and logistic regression, additive genetic model, test significance, type I error and multiple testing.

Day three:

Association analysis - Dr. Inga Prokopenko
Analyses of data using PLINK software, including genetic models used for statistical analysis, covariates and adjustments, basic types of single-variant analyses, graphical representation of the output results

Population structure – Prof. Andrew P. Morris
Identification of population outliers in GWAS and methods for detecting and accounting for structure within populations. Use of PLINK for principal components analysis and association analysis adjusting for structure.

Day four:

Imputation of GWAS - Dr. Inga Prokopenko
GWAS reference panels, including HapMap and 1000 Genomes Projects, reference haplotypes, imputation with IMPUTE software, phasing and imputation steps, chromosome chunks, combining chinks for imputed data analysis, quality of imputation, imputed genotypes probability.

Invited Speaker:
TBC

Meta-analysis of GWAS - Prof. Andrew P. Morris
Combining association summary statistics across GWAS using fixed-and random-effects meta-analysis. GWAMA software to perform meta-analysis.

Day five:

Genetic risk scores, Mendelian Randomization – Dr. Krista Fischer and Dr. Reedik Mägi
Calculation of genetic risk scores, statistical approaches for causal inference, including instrumental variable analysis, within Mendelian Randomization

Invited Speaker:
TBC

Q&A session – all course leaders

Analysis of rare variants – Prof. Andrew P. Morris
Rationale for rare variant analysis. Methods for assaying rare variation. Methods for the analysis of rare variants. GRANVIL software for testing association with rare variants.