



Childhood and Adolescence
Psychopathology:
unravelling the complex etiology
by a large Interdisciplinary
Collaboration in Europe

Workshop 2, Statistical genetics – genome-wide association studies Deliverable number: 7.4

Work Package: WP7
Lead partner: Imperial College London, 999993468
Author(s): Marjo-Riitta Jarvelin; Vincenzo Salerno
Submission date: 18/07/2018
Version number: 0.1 **Status:** Draft

Grant Agreement N°: 721567
Project Acronym: CAPICE
Project Title: Childhood and Adolescence Psychopathology: unravelling the complex etiology by a large Interdisciplinary Collaboration in Europe
Call identifier: H2020-MSCA-ITN-2016
Funding scheme: MSCA-ITN-ETN - European Training Networks
Topic: MSCA-ITN-2016 - Innovative Training Networks
Start date of the project: February 1st, 2017
Duration: 48 months

Dissemination Level	
PU: Public	✓
PP: Restricted to other programme participants (including the Commission)	
RE: Restricted to a group specified by the consortium (including the Commission)	
CO: Confidential, only for members of the consortium (including the Commission)	



Project funded by the European Commission

under the Horizon 2020 Programme

Marie Skłodowska-Curie Actions - Innovative Training Networks

Disclaimer:

This document has been produced in the context of the CAPICE Project. The CAPICE project is part of the European Community's Horizon 2020 Program (Marie Skłodowska-Curie Actions - Innovative Training Networks) for research and development and is as such funded by the European Commission. All information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission has no liability with respect to this document, which is merely representing the authors' view.

Table of Contents

1.	Introduction to the statistical analysis of genome-wide association studies	4
2.	Tables.....	5
3.	Annex I.....	7
3.1.	Selected reading material and links.....	7
3.2.	Selected useful links.....	8
3.3.	Data resources	9

1. Introduction to the statistical analysis of genome-wide association studies – London, 2-6 July 2018

From the 2nd to the 6th July, the CAPICE Early Stage Researchers (ESR) had the opportunity to attend the GWAS course in London titled, Introduction to the statistical analysis of genome-wide association studies. The course was organised by Imperial College London and held at the Hammersmith Hospital Campus.

This constituted, for the CAPICE consortium, the second network-wide training event (workshop 2) in line with the CAPICE programme.

The course was designed for geneticists facing the need to analyse from small to large-scale genotype data in relation to common human traits and diseases; scientists and students in training aiming to undertake SNP-based and genome-wide association studies and their meta-analyses; researchers willing to understand better the statistical approaches and analytical procedures including quality control issues for the genetic association studies.

This very intense course covered the statistical background for association studies; primer on scripting in the most frequently used computational environments; design and analysis of such studies and interpretation of the results. Lectures were followed by practical exercise, using state-of-art software tools and example datasets, which were tailored to illustrate the ideas discussed during lectures, complemented with discussions of the results.

Two additional meetings and discussions were organised during the week involving the CAPICE consortium only. On Monday 2nd July, before the start of the course during introduction session (and first break) and ahead of the busy schedule for the ESRs, Prof Marjo-Riitta Jarvelin (WP7 leader) and Mr Vincenzo Salerno, Project Manager for Professor Jarvelin, introduced the course to the ESRs, eliciting thoughts and discussions on how the content of the course would feed into the ESRs' wide-training goals and relatively to their individual projects. Additionally, the ESRs were asked to think of how to disseminate their learning during that week to the wider public, through a dissemination task which was followed upon on the 4th day of the course. On Thursday 5th July, during the lunch break, a session was organised for the CAPICE ESRs to discuss about their personal opinion and feedback about the GWAS course, their learning outcomes and general issues. The meeting also focused on their individual short dissemination report to be complete in two weeks following the course and shared with Prof. Marjo-Riitta Jarvelin and CAPICE Coordinator. Overall, ESRs thought the course had been very useful and for the most of them really a new learning experience. Discussion about dissemination their own work further was initiated and everybody had a chance to talk about their own views. This task needs to be taken forward locally in their own universities and again during the next training session.

2. Tables

Table 1 – Course Timetable

Course timetable					
	Day 1 – Monday 02.07.2018	Day 2 –Tuesday 03.07.2018	Day 3 – Wednesday 04.07.2018	Day 4 – Thursday 05.07.2018	Day 5 – Friday 06.07.2018
Teaching 9.00-10.00 Room 3S2 CAPICE STUDENTS' WELCOME AT 8.30	Welcome & introductions LECTURE Dr. Inga Prokopenko Introduction to statistics for geneticists (part I)	LECTURE Dr. Reedik Mägi Quality Control (QC) for GWAS	LECTURE Dr. Inga Prokopenko Association analysis	LECTURE Dr. Inga Prokopenko Imputation of GWAS	LECTURE Dr. Krista Fischer, Dr. Reedik Mägi Genetic risk scores, Mendelian Randomization
Tea/Coffee Break 10.00-10.15 Room 3S1 in breakout area layout 9.50-10.20: Introduction to CAPICE ESRs					
Teaching 10.15-11.15 Room 3S2	LECTURE Dr. Marika Kaakinen Introduction to statistics for geneticists (part II)	COMPUTER WORKSHOP Dr. Reedik Mägi QC for GWAS	COMPUTER WORKSHOP Dr. Inga Prokopenko Association analysis	COMPUTER WORKSHOP Dr. Reedik Mägi Imputation of GWAS	COMPUTER WORKSHOP Dr. Krista Fischer, Dr. Reedik Mägi Genetic risk scores, Mendelian Randomization
Biobreak 11.15-11.30					
Teaching 11.30-12.30 Room 3S2	COMPUTER WORKSHOP Dr. Natalia Pervjakova Introduction to Unix and R	COMPUTER WORKSHOP QC for GWAS	COMPUTER WORKSHOP Association analysis	COMPUTER WORKSHOP Imputation of GWAS	COMPUTER WORKSHOP Genetic risk scores, Mendelian Randomization
LUNCH 12.30-13.30 Room 3S1 in breakout area layout			Group photo	CAPICE STUDENTS – Discussion about the course and dissemination task (student projects)	Course certificates & feedback forms
Teaching 13.30-14.30 Room 3S2	COMPUTER WORKSHOP Dr. Natalia Pervjakova Introduction to Unix and R	INVITED LECTURE‡ Prof. Patricia Munroe Genetics of arterial blood pressure: from common to rare variants	LECTURE Prof. Andrew P. Morris Population structure	INVITED LECTURE‡ Prof. Bertram Müller-Myhsok Systems genetics of major depression and other stress related phenotypes	INVITED LECTURE‡ Prof. Zoltán Kutalik Mendelian randomisation to reveal biomarkers, obesity subtypes and boost GWAS power All course leaders Q&A session
Biobreak 14.30-14.45					

Teaching 14.45-15.45 Room 3S2	LECTURE Dr. Marika Kaakinen Introduction to GWAS	LECTURE Dr. Krista Fischer Statistical models for genetic association analysis	COMPUTER WORKSHOP Population structure	LECTURE Prof. Andrew P. Morris Meta-analysis of GWAS	LECTURE Dr. Andrew Morris Analysis of rare variants
Tea/Coffee Break 15.45-16.00 Room 3S1 in breakout area layout 15.40-16.10					
Teaching 16.00-17.00 Room 3S2	COMPUTER WORKSHOP Dr. Marika Kaakinen Introduction to GWAS	LECTURE Dr. Krista Fischer Statistical models for genetic association analysis	COMPUTER WORKSHOP Population structure	COMPUTER WORKSHOP Meta-analysis of GWAS	COMPUTER WORKSHOP Analysis of rare variants
17.00-18.00	Canapés and refreshments			Course dinner	COURSE CLOSURE

‡ Wolfson Education Centre, Lecture Theatre 2

Table 2: List of attendees

List of attendees – CAPICE ESRs	
Beneficiary	Early Stage Researcher
VU University	Wonu Akingbuwa
VU University	Eshim Shahid
King's College	Andrea Allegrini
University of Gothenburg	Sabrina Doering
University of Bristol	Elis Haan
University of Bristol	Laura Schellhas
Erasmus University	Elizabeth Diemer
Karolinska Institutet	Ashley Thompson
University Twente	Kratika Agarwal
University of Cagliari	Hema Sekhar Reddy Rajula
Imperial College	Ville Karhunen
Janssen	Marica Leone

3. Annex I

3.1. Selected reading material and links

1. 10 Years of GWAS Discovery: Biology, Function, and Translation. [https://www.cell.com/ajhg/pdf/S0002-9297\(17\)30240-9](https://www.cell.com/ajhg/pdf/S0002-9297(17)30240-9) by PM Visscher - 2017
2. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet.* 2005;6(2):95-108.
3. Balding DJ. A tutorial on statistical methods for population association studies. *Nat Rev Genet.* 2006;7(10):781-91.
4. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet.* 2008;9(5): 356-369.
5. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661-678 (2007).
6. Pearson TA, Manolio TA. How to interpret Genome-wide Association Study. *JAMA.* 2008;299(11):1335-1344.
7. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature.* 2009 Oct 8;461(7265):747-53.
8. GWAS imputation. Read the paper about the Haplotype Reference Consortium (HRC) reference panel A reference panel of 64,976 haplotypes for genotype imputation. <https://www.nature.com/articles/ng.3643>
9. Highly discussed recent article related to GWAS: An Expanded View of Complex Traits: From Polygenic to Omnigenic. Evan A.Boyle Yang I.Li Jonathan K.Pritchard <https://www.sciencedirect.com/science/article/pii/S0092867417306293?via%3Dihub>
10. Interesting paper discussing genomic inflation factor on biorxiv.org Estimating inflation in GWAS summary statistics due to variance distortion from cryptic relatedness
11. Meta-analysis of GWAS Quality control paper: Winkler TW et al. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc.* 2014 May;9(5):1192-212

3.2. Selected useful links

1. Tutorial - good intro into UNIX <https://www.tutorialspoint.com/unix/index.htm>
2. R – software webpage: www.r-project.org download from: <http://cran.r-project.org> or mirror sites, www.rstudio.com
3. Introduction to R, <http://cran.r-project.org/doc/manuals/r-release/R-intro.html> www.r-project.org , www.rseek.org
4. Quick R: data management, statistics, graphs <https://www.statmethods.net/>

5. www.r-bloggers.com (Rbloggers are also on Facebook)
6. We will refer to PLINK software <http://zzz.bwh.harvard.edu/plink/>
7. We will use EPIACTS for the analyses <https://genome.sph.umich.edu/wiki/EPIACTS>
8. We will also use BEAGLE for imputation
<http://faculty.washington.edu/browning/beagle/beagle.html>
9. Explore UCSC web browser for associations within the FTO gene locus
http://genome-euro.ucsc.edu/cgi-bin/hgTracks?db=hg19&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr16%3A53737875%2D54148379&hgid=228412872_I2oNFw7kwLjDr2wTpWVIMEkadsYB
10. Explore GWAS catalog for the GWAS signals reported for the same FTO gene locus
<http://www.ebi.ac.uk/gwas/search?query=FTO>
11. Explore PubMed for papers reporting the GWAS signals for the same FTO gene locus
<https://www.ncbi.nlm.nih.gov/pubmed/?term=FTO+association>
12. demo(graphics) <http://www.ms.ut.ee/mart/R/Rgraafika.html> (Author: Märt Möls)
13. Additional software tools for Quality control of GWAS: EasyQC (Winkler et al.)
<http://www.uni-regensburg.de/.../genetisc.../software/index.html>
14. excellent genetic data converter from team of J.Marchini
http://www.well.ox.ac.uk/~gav/qctool_v2/#overview
http://www.well.ox.ac.uk/~gav/qctool_v2/#file_formats
15. Power and sample size calculators:
<http://sampsizes.sourceforge.net/iface/index.html>
16. A new web-based version of power calculator for Genetic Association Studies
http://csg.sph.umich.edu/abecasis/gas_power_calculator/
17. We will use <http://biostats.usc.edu/Quanto.html> for power calculations (limitation – not available for Mac)

3.3. Data resources

For application to the UK Biobank data <http://www.ukbiobank.ac.uk/register-apply/> to test your research hypotheses in 500,000 British individuals (it costs ~£3,000 per project)