**BABC – Broad Antisocial Behavior Consortium**

Analysis plan for GWAS on antisocial phenotypes

Last updated: February 7th 2014

This document provides the instructions for the data preparation and GWA analyses on antisocial phenotypesin each of the participating cohorts. Standardisation of the procedures is very important, as it will increase the precision of the meta-analyses across all samples of the consortium.

**Study aim**

The Broad Antisocial Behavior Consortium has been created to combine the results of multiple genome-wide association studies of broad antisocial behavior (measured by symptom counts of antisocial personality disorder, ratings of aggression, conduct problems, delinquency, psychopathic personality etc) in meta-analyses in order to increase the probability of detection of genetic variants associated with individual differences in liability to antisocial behaviors. At this first discovery phase our aim is to include at least ~20,000 individuals.

**Deadline**

Please upload the GWAS results of your samplebefore: 28 February 2014.

If you know you cannot make that deadline please let us know as soon as possible.

**Contact details:**

Please read carefully and if you have any questions, do not hesitate to email Jorim Tielbeek, [j.tielbeek@debascule.com](mailto:j.tielbeek@debascule.com) or Ada Johansson, [ada.johansson@qimrberghofer.edu.au](mailto:ada.johansson@qimrberghofer.edu.au).

**The working group**:

Prof dr. Danielle Posthuma, [danielle.posthuma@vu.nl](mailto:danielle.posthuma@vu.nl)

Jorim Tielbeek Msc, [j.tielbeek@debascule.com](mailto:j.tielbeek@debascule.com)

A/Prof Sarah Medland, [sarah.medland@qimrberghofer.edu.au](mailto:sarah.medland@qimrberghofer.edu.au)

Dr. Ada Johansson, [ada.johansson@qimrberghofer.edu.au](mailto:ada.johansson@qimrberghofer.edu.au)

Prof. dr. Stephen Faraone, [sfaraone@childpsychresearch.org](mailto:sfaraone@childpsychresearch.org)

**Content of this document:**

1. Participating cohorts
2. Instructions for phenotype and covariate coding
3. Instructions for genotype handling
4. Analysis outline
5. Instructions for uploading data
6. Meta-analyses
7. Authorship
8. **Participating cohorts (as of February 7th 2014)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Study sample** | **Country** | **Data analyst** | **Principal investigator(s)** |
| 1 | **QIMR Berghofer** | Australia | Jorim Tielbeek | Nick Martin |
| 2 | **COGA** | USA | Jessica Salvatore | Danielle Dick |
| 3 | **ALSPAC** | UK | TBA | Marcus Munafo |
| 4 | **Michigan State University Twin  Registry** | USA | Xiaowen Kong | Alexandra Burt |
| 5 | **Colorado Twin Registry** | USA | Jaime Derringer | Robin Corley |
| 6 | **Helsinki Birth Cohort Study** | Finland | Jari Lahti | Katri Räikkönen |
| 7 | **The Southern Illinois Twins and Siblings Study (SITSS)** | USA | TBA | Lisabeth DiLalla |
| 8 | **BinnensteBuiten Study** | Netherlands | Mark Patrick Roeling | Danielle Posthuma |
| 9 | **TEDs** | UK | Jorim Tielbeek | Essi Viding/Robert Plomin |

1. **Instructions for phenotypes and covariate coding**

*Inclusion*

We propose to limit the analyses to subjects from European ancestry. Please let us know if you have a large group of individuals of non-European ancestry.

*Antisocial phenotypes*

Please perform the analyses on continuous data. If you have multiple measures at different ages with the same measurement instrument, please contact Jorim Tielbeek ([j.tielbeek@debascule.com](mailto:j.tielbeek@debascule.com)) to discuss which instrument(s)/age group(s) are most relevant for the consortium.

*Covariates - Please use these variables as a covariate in your GWAS analysis:*

* **Age** at the time of the phenotypic assessment (in years since birth).
* **Sex** coded as 1=male, 0=female.
* **Population structure**. Please use the first four principal components to correct for population structure in your sample. If necessary add study-specific covariates such as study site or batch effects.

1. **Instructions for genotype handling**

Genotyped SNPs (Affymetrix, Illumina, Perlegen)

Please exclude monomorphic SNPs and provide a list of these SNPs. No other filtering on call rate/HWE/MAF/imputation quality is required (QC metrics to be reported, and filtering will be done at meta-analysis stage).

*Imputation\**

Analyse all SNPs including SNPs on the X chromosomes. For analysis of the X chromosome, males will be treated as homozygous females. Please use dosage data.

Reference set used for imputation –1000 genomes (GIANT reference panel or the version 3 references). Please let us know the program and settings you have used for your imputation (If you have used the settings from a 1000 genome cookbook please let us know which one).

*Software for imputation:*

IMPUTE(2), MACH, BEAGLE or other

If you would like to use another reference set for imputation, please contact Jorim Tielbeek ([j.tielbeek@debascule.com](mailto:j.tielbeek@debascule.com)) to discuss the options.

1. **Analysis outline**

In order to maximize power, continuous traits are preferred. If continuous data are not available, such as in some clinical cohorts with matched controls, please perform the analysis on the dichotomous trait.

Please perform the analyses for:

1. Only males
2. Only females
3. Combined

If you have two selected measurement instruments (e.g. SDQ and APSD), relevant for the consortium please run the three analyses for the two instruments separately (so six analyses in total).

Exclusions:

- One of siblings / twins or use a method that accounts for genetic relatedness (e.g. using MERLIN)

- Non-Caucasian

*Continuous/quantitative:*

For continuous measures please use linear Regression onto estimated dose (as included in PLINK, MACHQTL, ProbABEL, SNPTEST, MERLIN), while adjusting for population structure and covariates.

Model Linear Regression: ASB score ~ sex + age + PCs + SNP

- SNP = genotype (estimated dose from 0 to 2)

- Sex: coded 1 for male, 0 for female

- Age: at measurement (in years)

- PC’s: principle components\* (ancestry)

*Dichotomous/diagnostic:*

In case you only have case control status, please use logistic regression with dichotomous outcome. Please create a dichotomized score using the specific cutoff per measurement instrument, if no standard (such as DSM-IV/V) criteria exist, please contact Jorim Tielbeek ([j.tielbeek@debascule.com](mailto:j.tielbeek@debascule.com)) to discuss.

Model Logistic Regression: ASB (dichotomous measure) ~sex + age + PCs + SNP

- SNP = genotype (estimated dose from 0 to 2)

- Sex: coded 1 for male, 0 for female

- Age: at measurement (in years)

- PC’s: principle components\* (ancestry)

1. **Instructions on the format of the input files for the Meta-analysis**

Data format - Please strictly adhere to the data format as specified below in this Table!

|  |  |  |  |
| --- | --- | --- | --- |
| Column: | Variable Name: | Description: | Type/Format |
| 1 | SNP | CHR\_BP name if 1000G was used for imputation | STRING |
| 2 | CHR | Chromosome | NUMERIC, no decimals |
| 3 | POS | Position | NUMERIC, no decimals |
| 4 | EFF\_ALL | Effect or Coded allele, for which the linear regression effect is reported (A/T/G/C). | STRING |
| 5 | NONEFF\_ALL | Non-effect or Non-coded allele (A/T/G/C). | STRING |
| 6 | Beta/OR | Beta (linear regression)  Odds Ratio (logistic regression) | NUMERIC, 4 decimals |
| 7 | SE | Standard error of effect of additive test. | NUMERIC, 4 decimals |
| 8 | P | P-value of additive test | NUMERIC, scientific notation, e.g. 1.02E-06 |
| 9 | AF\_COD | Allele Frequency of the Effect/Coded allele | NUMERIC, 4 decimals |
| 10 | HWE\* | HWE p-value \*optional: only if provided by your output | NUMERIC, scientific notation, e.g. 1.02E-03 |
| 11 | IMP | Whether a SNP was observed (=0) or imputed (=1) | NUMERIC, no decimals |
| 12 | INFO | Imputation quality for imputed SNPs, set to 1 if the SNP was directly genotyped. Report R2hat or proper\_info depending on software used. | NUMERIC, no decimals |
| 13 | INFO\_TYPE | Software used for imputation (IMPUTE(2), MACH, BEAGLE or other) | STRING |
| 14 | N\_EFF | Effective sample size (number of individuals with genotype (imputed or direct) and phenotype data. Note that this can differ per SNP) | NUMERIC, no decimals |

* Please save your file with results as plain space-delimited text file.
* Missing values should be coded as -999
* Please provide the variable names in the first row of the file. Please use the exact same phrasing as in the Table.
* Please keep the order of the variables as requested above.

Header line: SNP CHR POS EFF\_ALL NONEFF\_ALL BETA SE P AF\_COD HWE IMP INFO INFO\_TYPE N\_EFF

1. **Instructions for uploading data**

The results from the association analyses can be uploaded using secure file transfer protocol (sftp). To upload the data you need an sftp program. Both filezilla and winscp are freely downloadable ([filezilla-project.org/](http://filezilla-project.org/) ; [winscp.net/eng/download.php](http://winscp.net/eng/download.php)) and work well on our site.

***WinSCP:*** *If you have administrator rights, choose the 'Installation package', otherwise the 'Portable executables'. Install the package and start up winscp: choose for a new session, type lisa.surfsara.nl under 'Host name', type your login name under 'Login' and choose SCP (or SSH2) for 'protocol'. Click Login, and type your password when asked for. A graphical interface will appear, showing left the files in your Windows system, on the right the files on Lisa. You can drag and drop files from your Windows system to lisa. See also* [*the winscp documentation*](http://winscp.net/eng/docs/start)*.*

The host name is: **lisa.surfsara.nl**

The login name is: babc

The password will be sent to you in a separate email – please let us know when you’ve finished the analysis so we can send you the password.

For information about lisa/sara:

<https://www.surfsara.nl/systems/lisa/new-users>

Please put the result files in the subdirectory:

**Results\_GWA\_BABC\_Antisocial**

The following naming scheme for the files with your association results is preferred:

**STUDY.PHENOTYPE.SEX.DATE.txt**

Where,

STUDY is a short identifier for the cohort studied

PHENOTYPE is the measure used, such as ‘SDQ’

SEX is the type of analyses: ‘Male’, ‘Female’ or ‘Combined’.

DATE is the date on which the file was prepared, in the format DDMMYYYY

After uploading please send an email to Jorim Tielbeek (j.tielbeek@debascule.com)

1. **Meta-analysis**

The meta-analysis will be performed using METAL.

1. **Authorship**

All participating groups can send us the names of the co-authors who were involved in this study (if more than 3 colleagues are involved please contact Jorim Tielbeek ([j.tielbeek@debascule.com)](mailto:j.tielbeek@debascule.com)))

Apart from the main analysts and overseeing analysis group members, authors will be grouped based on their contributions:

1. Analysts (main researchers who carried out the analyses)
2. Other researchers (DNA collection, phenotype collection, cleaning data, QC etcetera)
3. PI (head of the group/project)

We will sort on alphabetical order within each group. The structure will be:

[Members of working group(s)]- [1] – [2] – [3] – [Members of working group(s)]

‘Member of working group(s) will be the researchers who largely contributed to the development of the analyses plan, who carried out the meta-analyses, who contributed largely to the manuscript and who were in charge of this meta-analyses project.