
ABCD Human Subject Study

Adolescent Brain Cognitive Development – ABCDSTUDY.org

Release Notes: Adolescent Brain Cognitive Development Study (ABCD) Fix Release 2.0.1

Genetics

<http://dx.doi.org/10.15154/1503209>

March 2019

Change Log

March 2019 – ABCD Data Release 2.0

- Initial release

July 2019 – ABCD Fix Release 2.0.1

- Added notice of Fix Release 2.0.1 and updated text

List of Instruments

*(Updated in Fix Release 2.0.1)

Name of Instrument	Short Name
Genomics Sample*	genomics_sample_03
Experiment Description	omics_experiments
ABCD Youth Genetic Blood (RUCDR)	biocf01
ABCD Youth Genetic Saliva (RUCDR)	abcd_ygs01

General Information

The following information refers to the Adolescent Brain Cognitive Development (ABCD) Study Data Release 2.0/2.0.1 available from <https://data-archive.nimh.nih.gov/abcd>. An overview of the ABCD Study is at <https://abcdstudy.org> and detailed descriptions of elements of the study are listed in <https://abcdstudy.org/scientists/protocols>.

This document describes the contents of various instruments available for download. To understand the context of this information, see *Release Notes ABCD README FIRST* and *Release Notes ABCD Imaging Instruments*.

July 2019 Fix Release 2.0.1 Please ensure updated instruments contained in Fix Release 2.0.1 replace instruments downloaded from the original Data Release 2.0 as indicated in the above table.

SmokeScreen Data

Genomics_sample_03. Each entry in this instrument references the same plink in a zip-file.

1. genotyping platform

Affymetrix NIDA SmokeScreen Array

The sample preparation and genotyping are performed by Rutgers RUCDR. This includes:

- extraction kit: Chemagen bead based/Chemagic STAR DNA Saliva4k Kit (CMG-1755-A)
- processing: DNA fragmentation, labeling, ligation & hybridization
- equipment: Affimetric GeneTitan Instrument.

Additional information can be found in the NIMH experiment description #1194.

The smokescreen array contains 733,293 SNPs

(<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4769529/>)

2. genotyping QC process

RUCDR has performed DNA quality controls based on calling signals and variant call rates.

573,845 genetic variants passed the quality controls of raw genotype calling process.

DAIC of ABCD performed the subsequent study-based QC process, following the recommendation of Ricopili pipeline. 527,285 SNPs have call rate greater than 99 percent across ABCD samples.

3. Sample QC process

For individuals who have been successfully genotyped, further QC was performed. Individuals who have a potential problem in sample contamination or have a missing rate greater than 20 percent were removed from the data. An additional 34 individuals were found to have conflicting IDs, and were removed from the data. After QC, there are 10627 subjects in Release 2.0.1.

NOTE

“Sex” data were removed from all plink files after the discovery of many “sex” mismatches with the curated ABCD data repository. “Sex” has been set to missing (-9). Researchers who intend

to perform genetic analyses would need to analyze phenotype data from the ABCD data repository.

In our QC process, we found the plate 461 to be especially problematic. We recommend not using genotype data from plate 461, or at least including plate number as a covariate.

Data content

1. Minimally processed genotype data in plink format
[ABCD_release_2.0.1_r.*](#)

2. SNP info from the genotyping arrays
[Genotyped_SNP.updated.info](#)

3. Subject basic information, including plate number of genotyping and the processing batch.
[ABCD_release_2.0_r.batch.info](#)

Sections of this document are included in the genomics download package from NDA. The instrument that shares this data on NDA is genomics_sample03. Each entry in that instrument references the same plink in a zip-file.

To use genomic data for population indices, such as genetic ancestry, population stratification, genetic relationships, and zygosity inference (described below), the user can employ the released QCed genetic data directly. Official release of these indices will be included in a future release.

Zygosity Calculation

The zygosity is determined by probability of identity-by-descent. We recommend the following steps to calculate the zygosity:

a. Selecting independent SNPs based on i) allele frequency > 0.05, ii) missing rate < 0.1, iii) pruning to ensure independency.

We suggest using PLINK with default parameters to perform this task.

b. Based on selected SNPs, calculate probability of identity-by-descent.

We suggest using the PLINK command `--genome` to get the probability of identity-by-descent, `PI_HAT`

c. Determine zygosity based on calculated probability of identity-by-descent

Theoretically, the $P(\text{IBD})$ greater than 0.5 means full siblings or dizygotic twins, 1 means monozygotic twins. However, given the estimations are actually based on identity-by-state and there is inherent noise in the genotyping data, we recommend using 0.4 and 0.8 as the

thresholds. All the processing pipelines will be available on ABCD GitHub (<https://github.com/ABCD-STUDY>).

ABCD Youth Genetic Blood (RUCDR) - Information collected by RAs at the time of venipuncture blood collection for genetics studies.

ABCD Youth Genetic Saliva (RUCDR) - Information collected by RAs at the time of saliva collection for genetics studies