The goal is to screen 510k+ genome variants for association with 14 social behavior phenotypes under three settings: male only, female only, combined. A total of 42 test is performed.

The initial sample size is 1000 children which are 500 twin pairs.

Age and gender are covariates for combined test, for male /female only settings, gender is dropped.  
Analysis is done using R gee package. Each twin pair is assigned unique family ID and sorted accordingly to ensure adjacency in the sample queue, which is required by R gee.

General cleanup:

Exclude samples of non-European origin(ethnicity not equal to 6, 136 out of 1000).

Exclude variants with low minor allele frequency(MAF<0.05).

Exclude variants not on autosomal and sex chromosomes(CHR>25).

Exclude variants on Y chromosome(CHR=24) under female setting.

Per analysis cleanup:

Exclude samples with any missing value in covariants or testing phenotype.

Other fix:

Tow sample(51021100 and 51021101 ) have inconsistent family id(being 520211, but 510211 expected) in the original phenotype sheet “2014.04.24 Child Twin CP AGG RB Data.xlsx”, these are fixed by changing family id to 510211 before any analysis.

Output:

A total of 42 tables are generated for the tests of 14 phenotypes under 3 settings. These tables are named and formated to meet the requirements listed in “Analysis Plan Broad Antisocial Behavior Consortium 7Feb2014.docx”.

Coveriants under different settings: