

dbGaP Study Release Notes



Release Notes for NIMH Autism Genome Project, phs000267.v1.p1 "Autism Genome Project (AGP) Consortium - GWAS and CNV - Stage I"

For any questions or comments, please contact: dbgap-help@ncbi.nlm.nih.gov.

September 20, 2011 Version 1 Data set release date

[2011-09-20](#)

Version 1 Data set release for NIMH Autism Genome Project now available

This release includes phenotype tables, genotype calls, and raw genotype intensity data with the following breakdown. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): Autism and Related Disorders (ARD)

	phenotype	genotype
subjects	4448	4431

Additional notes for phenotypes

The submitter has also requested that 2 raw phenotype dataset files along with their respective data dictionaries be included in the final dump files. These files have not been processed by dbGaP and are not referenced by dbGaP accessions. The file names are:

- SubjectRawPedigreeFile.AGP.phs000267.v1.p1.20110613.xls (data dictionary)
- SubjectRawPedigreeFile_AGP_phs000267_v1_p1_20110613.txt (dataset)
- SubjectRawPhenotypeFile.AGP.phs000267.v1.p1.20110613.xls (data dictionary)
- SubjectRawPhenotypeFile_AGP_phs000267_v1_p1_20110613.txt (dataset)

All subjects in the Raw Pedigree and Phenotype files were included in the Subject Consent file, phs000267.v1.pht002302.v1.p1.Autism_Genome_Project_Subject.MULTI.txt.

Additional notes for genotypes

- See download component phg000143.v1.p1.NIMH_AutismGenomeProject.tar-file-indices.MULTI.tar.gz for lists of files within each released genotype .tar file
- See download components phg000143.v1.p1.NIMH_AutismGenomeProject.sample-info.Illumina.MULTI.tar.gz and phg000143.v1.p1.NIMH_AutismGenomeProject.marker-info.Illumina.MULTI.tar.gz for a manifest of the genotyped study samples, description of SNP array used for the genotyping, and for pre and post QC subject-SNPs filters.
- Genotype data can be found in several formats:
 - Genotype calls in individual format ("indfmt") containing a file per sample. There are two tar files – one containing 4067 QC-ied subjects for which matrix data exist and another containing 367 files for non-QC-ied subjects.
 - Genotype calls in the original submission ("original-submission") follow the same rule like the indfmt files above. Those are FinalReports CSV files.
 - Genotype calls in a matrix format ("genotype-calls-matrixfmt") comprise of PLINK format binary and text genotypes only for the QC-ied 4067.
- Quality control information is in phg000143.v1.p1.NIMH_AutismGenomeProject.genotype-qc.Illumina.c1.ARD.tar.gz.

dbGaP Study Release Notes



FTP site

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- <ftp://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs000267/phs000267.v1.p1>