



dbGaP: Genotype and Phenotype



National Center for Biotechnology Information ■ National Library of Medicine ■ National Institutes of Health ■ Department of Health and Human Services

Understanding the relationship between the genetic makeup of an organism (genotype) and the expression of a genetic trait in response to environmental variables (phenotype) is essential for the accurate diagnosis, treatment and prevention of human disease. A growing number of large-scale and long-term studies involving many individuals are producing data sets of genomic characteristics associated with complex traits and outcomes. **The Genotype and Phenotype Database (dbGaP)** has been established at the NCBI to archive, distribute, and support the submission of data that correlate genomic characteristics with observable traits. Important sources of data in dbGaP are whole genome association (WGA) studies. WGA data are contributed by a number of sources including the Genetic Association Information Network (GAIN) and research centers at the National Institutes of Health (NIH). Other data may be from medical sequencing, molecular diagnostic assays, and surveys of association between genotype and non-clinical traits. More information on WGA studies and the NIH-funded programs that support them is available on the **WGA homepage**:

www.ncbi.nlm.nih.gov/WGA/

Data in dbGaP

NCBI organizes the data into four different data types: *Studies, Study Documents, Phenotypic Variables, and Genotype-Phenotype Analyses*. Completed studies have associated Documents and Variables and some have pre-computed analyses. A summary of the completed association studies is shown in Table 1.

Table 1: Studies with whole genome association (WGA) data in dbGaP (October 12, 2007)

WGA Study	Variables	Participants	‡ Embargo Release Date	Analyses
 Framingham SNP Health Association Resource (SHARE)	13,183	15,876	October 1, 2008	yes
 International ADHD Genetics Project	438	2,835	March 26, 2008	no*
 Search for Susceptibility Genes for Diabetic Nephropathy in Type 1 Diabetes	-	1,835	July 16, 2008	no*
 Major Depression: Stage 1 Genome-wide Association in Population-Based Samples	-	3,786	August 16, 2008	no*
 National Eye Institute (NEI) Age-Related Eye Disease Study (AREDS)	174	600	June 11, 2007	yes
 National Institute of Neurological Disorders and Stroke (NINDS) Parkinson's Disease & Control	43 & 66	1,283 & 2,723	---	no

‡ Investigators who have submitted data to dbGaP retain the exclusive publications rights for a period of approximately 9-12 months after the data are released in dbGaP. Restricted access to the data may be granted to other scientific investigators, but they may not publish their analyses of the data until after the **embargo release date** has expired.

* The analyses for these WGA studies will be public after the embargo release date expires.

NCBI assigns unique identifiers (accessions) to the data in dbGaP, and researchers can search dbGaP data as a part of the NCBI Entrez system. The dbGaP Homepage also has a convenient browser that allows direct access to the Studies, Variables and Documents and Analyses:

www.ncbi.nlm.nih.gov/sites/entrez?db=gap

Open-access data can be browsed on-line or downloaded from the dbGaP ftp site without prior permission or authorization.

<ftp://ftp.ncbi.nlm.nih.gov/dbgap>

Data and Privacy Considerations

To protect the identity of the individuals involved in the studies, NCBI only accepts data with anonymous identifiers. Studies may have open-access data and analyses available, and some of the individual-level data may be detailed enough that privacy could be compromised. Therefore, access to the individual level data is controlled and only available to approved researchers. Access to individual level data in dbGaP is granted by an NIH Data Access Committee or DAC. Researchers wishing access to controlled data must submit a Data Use Certification, or DUC, to the appropriate NIH DAC for approval. More details on access to controlled data and the application process are available from the dbGaP pages:

<http://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login>

Analyses in dbGaP

Some public genome wide association study variables such as Age-Related Macular Degeneration (AMD) Status in the National Eye Institute (NEI)-Age-related Eye Disease Study (AREDS) have associated analyses that can be displayed in the dbGaP Genome Viewer and Chromosome Browser. The Genome Viewer quickly shows regions of the genome that contain phenotype-associated alleles (Figure 1). Linking to the Chromosome Browser provides detailed information about the alleles from dbSNP and precise locations on the NCBI Map Viewer with links to Entrez Gene and Entrez Nucleotide (Figure 2).



Figure 1: The dbGaP Genome Viewer showing the location of the sequence polymorphisms associated with Age-Related Macular Degeneration (AMD).

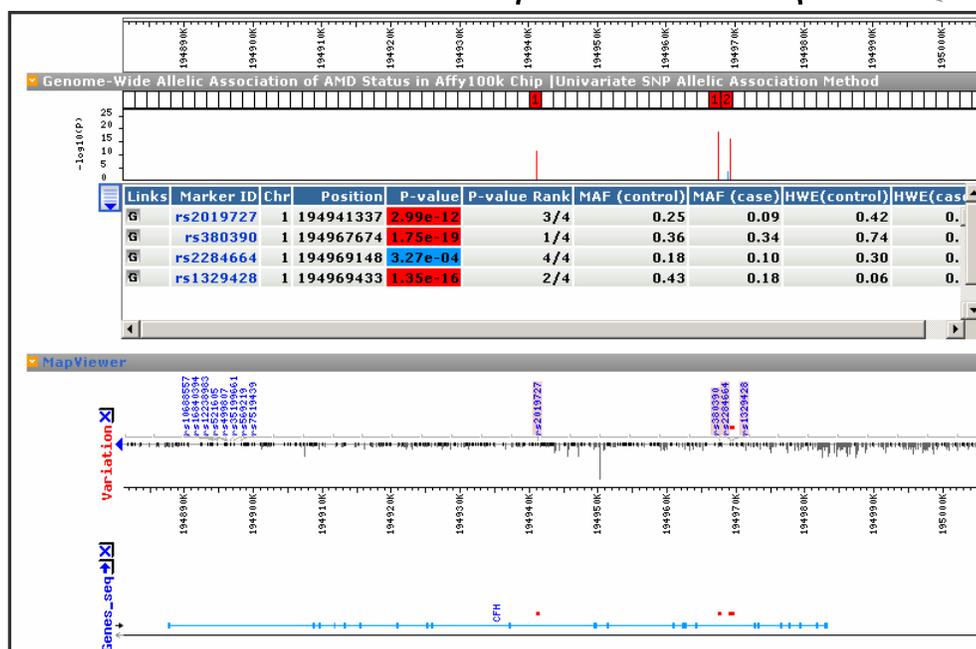


Figure 2: The dbGaP Chromosome Browser showing the locations of the sequence polymorphisms in the NCBI MapViewer and their mapping to a candidate gene.

Reference

Mailman, MD, *et al.* 2007 The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet* 39(10):1181-6.

Further Questions?

Contact the dbGaP help-desk: dbgap-help@ncbi.nlm.nih.gov