**[Supplementary Data Set 1: GWASdb entries with MeSH terms mapped for each trait and genes annotated as described in the Online](https://static-content.springer.com/esm/art%3A10.1038%2Fng.3314/MediaObjects/41588_2015_BFng3314_MOESM11_ESM.txt)**[**Methods**](https://www.nature.com/articles/ng.3314#Sec4)**.**

**Disease**: the name of the trait for the corresponding genetic association as provided by GWASdb, which is generally taken directly from the GWAS catalog or whatever source from which the association was derived.

**snp**\_id: the identifier (generally a dbSNP rs ID) of the SNP reported to be associated with disease.

**Link**: the reference for the association. Most references are a PubMed ID for the published paper.

**pvalue**: the *P* value reported for the association of snp\_id with disease.

Source: the origin of the association information. It may be the following:

GWAS:A/B: results listed in the NHGRI GWAS Catalog. The publications the associations of GWAS:B are drawn from published tables and supplementary information.

Omim: from Online Mendelian Inheritance in Man. All *P* values are zero. GWASCentral: from GWAS Central. dbGaP: associations from dbGaP.

**SNP.Trait.Cnt**: the number of associations of the same snp\_id with the same disease in the original data set. These have been reduced to a single row in this data set, and the minimum *P* value was selected.

**MSH**: Medical Subject Heading for disease. Manually mapped by Computational Biology.

**MSH.Top**: the MeSH term for the top level of the branch to which the trait is mapped. In most instances, there are many branches to which a single MSH may be mapped. When this occurs, the most common top-level term in GWASdb is selected.

**snp.ld**: a SNP in linkage disequilibrium (LD) with snp\_id that provides a plausible connection to a gene.

**Gene**: a gene that snp.ld is within 5 kb of, is an eQTL for or sits in a DNase I hypersensitivity site that is correlated with, or is within the transcription start site of.

**r2**: LD between snp\_id and snp.ld.

**eqtl**: indicates whether snp.ld is an eQTL for a gene. The eQTL data are drawn from eqtl.uchicago.edu.

**rdb**: indicates whether snp.ld-gene mapping is the result of a DHS correlation (from Maurano *et al*. (2012), provided by J. Stamatoyannopoulos).

**Cat.rdb**: RegulomeDB category of the SNP (if rdb is "yes"). Lower values indicate more lines of converging functional evidence.

**eCat**: a derivative of Cat.rdb, filling in values where rdb is "no." If eqtl is "yes" but rdb is "no," then it gets a value of 2. If snp.ld is a missense variant (amino acid change), the value is 0. The value is 9 otherwise.

**AAEffect**: amino acid effect of snp.ld.

**AAScore**: Condel score from VEP for nonsynonymous variants.

**GeneScore**: an overall assessment of the evidence that the associated variant has a causal effect on the gene in question, ranging from values of zero to eight. Higher scores imply higher weight of causal evidence. The contributions to GeneScore are summarized on a separate GeneScore Wiki page.

**Rank**: the rank for the given gene for its strength of connection to snp\_id. This takes LD and functional evidence into account. (TXT 12043 kb)