**Minimum-Required Information for Association Data**

Association analyses vary on trait, variant type, frequency, and analytic method. To facilitate the data sharing, we propose a unified guideline for Minimum Information Required for Association Data (MIRAD). It includes four essential parts of data elements.

**1. Locus Identifier**

The identifier includes (not limited) rs#, gene ID and SV# for SNP, gene, and structural variant respectively. They can be mapped to current genome build and also evolved with future reference genome assemblies and NCBI annotations.

**2. Variation summary**

It contains information about alleles, allele frequencies, sample size, and genotype counts per sample group within each locus. To limit the ability of unauthorized parties to infer individual participants, data like counts and frequency are only accessible to authorized researchers.

**3. Statistical significance and Effect size**

p-value and/or FDR, which come either from univariate test on common variants (printed on chips) or from group-based multivariate test on a set of rare variants provided by sequencing projects. The effect size includes odds ratio, regression coefficient, relative risk, etc., on particular allele (s). These data not only help users to find causal variant and haplotype, but also can be used to estimate locus contribution to the heredity of the trait or disease.

**4. Phenotype Definition and Analysis Metadata**

The main trait or disease analyzed should be defined to a controlled vocabulary such as UMLS, HPO, etc. Descriptions of the analysis and method, include phenotypic covariates and parameters, are needed for reproducing the result set once the individual data are fully available.

**Reasoning**

Sharing of these data elements allow other researchers to evaluate supporting evidence and independently verify discoveries with different samples and data models. In the event that individual level genotype is inaccessible, people are able to directly use them for meta-analysis to increase statistical power or for the development of hypotheses.

**Our practice**

Using MIRAD, dbGaP has developed several templates for data submission and browser exhibition. You are welcome to join the discussion, make suggestions and comment on the MIRAD proposal. The dbGaP team is committed to bringing new discoveries to public and research communities and are happy to work with researchers to promote data sharing within the scientific community.

If you have any question, or need any help from the service, please contact the dbGaP at: [dbgap-help@ncbi.nlm.nih.gov](mailto:dbgap-help@ncbi.nlm.nih.gov)