What are Genomic Summary Results & How Can "GSR" Inform Research and Clinical Care?

What are Genomic Summary Results?

Genomic Summary Results, or "GSR", which are sometimes called "aggregate genomic data" or "summary data", are the output of analyses of genomic data across the many individuals included within a specific study's dataset. For most studies in dbGaP, for example, this means that GSR represent a summary of the information generated from hundreds, or thousands, of research participants. This type of information is, thus, different from "individual-level data" about specific research participants, which usually are only available for appropriate research use through controlled access databases. There are two broad classes of GSR information: 1) allele frequency information; and 2) association analysis statistics.

Allele Frequency Information

An **allele** frequency is the proportion of a specific allele, or variation in the DNA code, relative to other possible alleles at the same position in the code in a given population, or in some cases, an entire species. It is calculated by counting the number of times an allele occurs divided by the total number of observations. Population allele frequencies are estimated by genotyping or sequencing many DNA samples from the population or group of interest. The accuracy of the estimate increases as the number of genotypes or sequences generated increases. Allele frequency can vary across ancestral groups (or populations), in response to environmental exposures or development of the group over time; therefore, displays of allele frequency estimates are usually calculated and displayed by population. Allele frequency information is used in the fields of Genomics, Population Genetics, and Clinical Genetics to help interpret the potential for links between the presence of specific alleles and observed "outcomes", such as physical traits or disease risks. For example, this type of information is a very powerful tool for clinical genetics, because if the allele observed in a patient with a rare disorder is very common among the general (healthy) population, it is unlikely to be the causal allele for their disease.

dbSNP and ExAC are public websites that show allele frequency data in humans.

Association Analysis Statistics

In genomics, association analysis statistics are the information generated when investigators evaluate the correlation of genotype to **phenotype**. Phenotypes studied may be diseases (e.g., diabetes), traits (e.g., height), or molecular traits (e.g., mRNA or protein expression levels). Examples of these kinds of statistics are: p-values, beta values in regression, the odds ratio, and effect size. Analyses are usually focused on a single phenotypic trait (or "main variable") and performed on a single population (e.g., Europeans). However, analyses often contain co-variates (e.g., diabetes and body weight), and there are numerous methods and variations on how association analyses can be performed.

A dbGaP website that shows genetic association analysis statistics in humans can be seen here.