March 18th, 2015
Alquicira Hernández José

# Ideas on clinical validity of GWAS

The clinical utility of genome-wide association studies depends on the study design and the power of the association. If common variants have a small effect size but common diseases show a strong inheritance in families (high heritability), then almost by definition the disease must be influenced by multiple genetic factors. If a SNP has a modest effect on disease risk, it can only account for a small portion of the total variance due to genetic factors. This is a considerable problem when trying to use a SNP as a predictor in the clinic for some specific disease. Although the susceptibility loci identified via GWAS analyses have been useful in providing a new insight regarding the biology of the disease, they have not resulted in new genetic tests.

Some of the following criteria could be useful to assess clinical relevance of SNPs:

- *Considering quantitative traits*
  A quantitative trait could provide more confidence to make predictions in the clinic, since it comprises a broader output of phenotype states that are associated with an allele. If a correlation between these outputs and an allele or genotype exists, it is more informative than associations between just a single phenotype state and a locus or loci.

- *Use of covariates in studies*
  Depending on the type of phenotype/disease to consider, the incorporation of covariates in the study design for some associations could be suitable when considering if a study is valuable for the clinic, since they can prevent spurious associations from being detected and correct the phenotype outcome due to genetic factors.

- *Population substructure*
  One of the most important covariates for GWAS is genetic ancestry. If a quality

control was made for population stratification (PCA, admixture analysis, deviation from Hardy-Weinberg equilibrium, etc.), the results of the analysis will be more reliable. Considering the genetic distance between populations from different continents, care must be taken when examining studies with samples from different populations.

➔ *Multiple testing corrections*

Given the number of tests in genome-wide studies, the use of a stringent correction for multiple testing must be taken into account when deciding which SNPs could be considered to be useful. If the analysis of some GWAS included a Bonferroni correction, they may be more trusted than other studies with less strict corrections.

➔ *Replications studies*

If a study was replicated using different samples and the sizes are big enough (compared to the original samples), this may be taken into account when determining if the inference about the genetic associations is dependable.

➔ SNP falling into biological significant region

Most of the SNPs associated with phenotypes fall into intergenic regions. However, some of them must be localized in regulatory regions that could be crucial in understanding the relation between the SNP and the phenotype.

If it is possible to know the biological relevance of these intergenic SNPs along with the intragenic SNPs (which are more trusted to be involved is phenotypes) either at genomic or metabolic level, the association may be more significant for clinical purposes, since a biological evidence supports the association.

Considering the data included in the GWAS catalog from the NHGRI, some available values for each study could be used to try to elucidate the confidence of each association. The previous suggestions can be applied for some studies of this catalog as a first quality filter criteria, since some information regarding the experimental design and inference are included.