Editorial

Genome-wide Association Studies: Promises and Pitfalls

Eng-King Tan,[1-3] *MBBS (S'pore), FRCP (UK),*

Genetic testing is an important means to confirm the diagnosis of an inheritable disease.[1] For this to be feasible, genes that are associated with the disease need to be identified. Hunting for the genes that cause or are associated with a particular disease is a challenging task. Traditionally, mapping causative genes or "genetic susceptibility loci" involves intensive linkage analysis in many families containing clusters of the disease of interest.

Genetic association study helps to unravel gene variants that shows an association with a defined disease trait.[2-4] This association study is usually done using a case-control methodology. However, an association does not necessarily equate to a causative effect of the gene variant. Furthermore, the relation between genetic variants and disease status depends on many confounding factors, such as the pattern of physical proximity of the gene variants (linkage disequilibrium) (also described as the non-random association of alleles at two or more loci) in the population; the type of cases and controls; ethnic differences; and selection bias in recruiting the subjects. This "non-rocket science" approach has yielded many observations that by and large could not be consistently replicated.

With technological advances and the completion of the Human Genome Project, it has become easier to search for genetic associations of common diseases. In recent years, genome-wide association studies (GWAS), sometimes referred to as whole genome association (WGA) studies, have generated considerable interest among investigators in genetic epidemiology.[2,5] The fundamental approach in GWAS is nothing new, as this is essentially a genetic association study, albeit with a much more intricate, and arguably more reliable, approach.

Earlier genetic association studies frequently involved the analysis of a single or a few gene variants of selected candidate genes of interest.[2-4] However, with an estimated 30,000 genes in the human genome, and possibly million of both common and rare gene variants, it may be unrealistic to expect investigators to strike a "jackpot" and uncover susceptibility or causative genes easily. In GWAS, the analysis involves examining variants across the whole-genome (i.e. the entire DNA sequence) for genetic variations that may associate with certain traits or diseases. Modern chip technology can allow the interrogation of up to 1 million single nucleotide variants (SNPs) at one go.

To date, GWAS have identified gene variants across a spectrum of diseases, such as age-related macular degeneration, diabetes mellitus (both Type 1 and Type 2), rheumatoid arthritis, Parkinson's disease, and hypertension. New GWAS are published every month in leading scientific journals, and it is difficult to keep abreast unless one has a particular interest in them. GWAS is a non-hypothesis-driven approach, because the entire genome is assessed for the genetic associations that may occur with a specific condition.[5,6]

The potential limitations are fairly similar to genetic associations in general. Sample size is a big consideration because of the large number of SNPs and the replications contained in numerous statistical comparisons. Genetic variants that are rare, or likely to exert a small effect, are unlikely to be detected. Varying risk genes in different populations and varying risk alleles in different ethnic groups are common confounders. Different genotyping platforms may provide variable coverage of the genome, i.e. with incomplete overlap of information. GWAS-derived data in diseases with no objective phenotype marker can be misleading. Replication of the data in independent datasets still remains the litmus test of the validity of GWAS.

How does one translate the GWAS data into clinical care, leading to a better diagnosis or treatment, or both? Genetic variants or loci that are rare or have small effect size generally have little predictive value at the population level. Genetic variants that are located in non-coding regions of the DNA may not be the real causative factors, and it is difficult to make sense of the significance of multiple variants located in regions of unknown genes.

Thus, it is generally accepted that most of the current GWAS data confer no immediate benefits to patients. However, hidden among the multiple limitations and caveats, there is an enormous potential for a better understanding of the molecular biology of the disease. Correlation of GWAS data with gene and protein expression studies will allow a better understanding of the biological role of some of the genetic variations.

[1] Department of Neurology, Singapore General Hospital, Singapore
[2] National Neuroscience Institute, Singapore
[3] Duke-NUS Graduate Medical School, Singapore
Address for Correspondence: Dr Eng-King Tan, Department of Neurology, Singapore General Hospital, Outram Road, Singapore 169608.
Email: gnrtek@sgh.com.sg

Through collaborative efforts, researchers have been able to combine multiple GWAS studies in more comprehensive genome-wide pooled analysis (more properly, meta-analysis) that could lead to new gene discoveries.[7] Improvement in bioinformatics software has allowed a systems biology approach, in which a pathway analysis of GWAS data could lead to the identification of a network of biologically important SNPs of various genes that may have implications in specific diseases.

In the near future, it is conceivable that detailed genetic profiling would allow a custom-made approach in drug-development and in predicting therapeutic responses for the individual.[8] Building on the GWAS platform, some research groups are applying complementary genotyping approaches such as whole genome re-sequencing. The latter approach promises to yield confirmatory information at lower cost.

Together with advances in the evaluation of epigenetic and environmental modulation, the GWAS and genome re-sequencing approaches provide some cautious optimism that discoveries from these technologies can translate into better care for those suffering from, or at risk of developing, inheritable diseases.

Acknowledgements

The author thank the support from Singapore General Hospital, National Neuroscience Institute, Duke-NUS Graduate Medical School, Singapore Millennium Foundation and National Medical Research Council.

REFERENCES

1. Tan EK. Re-defining neurological syndromes: the genotype meets the phenotype. Ann Acad Med Singapore 2006;35:63-5.
2. Tan EK, Khajavi M, Thornby JI, Nagamitsu S, Jankovic J, Ashizawa T. Variability and validity of polymorphism association studies in Parkinson's disease. Neurology 2000;55:533-8.
3. Tan EK. The role of common genetic risk variants in Parkinson disease. Clin Genet 2007;72:387-93.
4. Tan EK. Identification of a common genetic risk variant (LRRK2 Gly2385Arg) in Parkinson's disease. Ann Acad Med Singapore 2006;35:840-2.
5. Pearson TA, Manolio TA. How to interpret a genome-wide association study. JAMA 2008;299:1335-44.
6. Simón-Sánchez J, Singleton A. Genome-wide association studies in neurological disorders. Lancet Neurol 2008;7:1067-72.
7. de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum Mol Genet 2008;17(R2):R122-8.
8. Thomas DL, Thio CL, Martin MP, Qi Y, Ge D, O'Huigin C, et al. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. Nature 2009;461:798-801.