Available online at www.scholarsresearchlibrary.com

Extended Abstract

Journal of Computational Methods in Molecular Design, 2021 **ISSN 2231-3176**
https://www.scholarsresearchlibrary.com/journals/journal-of-computational-methods-in-molecular-design/

Efficient Genome-Wide Association Studies and Post-GWAS Integrative Analyses for Human Cancer and Neurodegenerative Diseases

Zahra Mortezaei, Laboratory of Systems Biology and Bioinformatics
E-mail: Zmortezaie@gmail.com

It is evident that in etiologies of human complex diseases, genetic factors play some important roles. Genome-wide association study (GWAS) is a standard technique to identify heritable genetic basis of complex diseases. In relation with GWAS, there exist some challenges in selecting input samples completely randomly, to biologically describe GWAS results, to translate them into clinical benefits and to compare germline variants achieved from GWAS with somatic mutations in creating, development and treatment of human complex diseases. Likelihood-based statistical methods are robust in estimating linkage disequilibrium when factors like non-randomness and population structures exist. Then the results of GWAS can be used for post-GWAS analyses to predict multiple biological components like genes, non-coding RNAs and transcription factor binding sites in association with complex diseases. An integrative analysis seeks to pool information from multiple GWAS results, somatic mutations and genetic drug targets of human complex disorders and the results of such analysis can provide new insight into the genetic and treatments of complex diseases. This presentation is prepared from the viewpoint that the robust statistical method can be applied to arrive at valuable results from GWAS and that primarily genetic information derived from GWAS is subject to further post-GWAS analysis to provide more biologically informative results in relation with genetics of human complex diseases that can be applied to real time clinical applications. Then the results of such analyses can be used to discuss and compare human cancers and neurodegenerative diseases from a genetic perspective. We concluded that in spite of the differences between human cancers and neurodegenerative diseases, the roles of germline and somatic mutations in creating, developments and treatments of those two kinds of human complex diseases are similar. For many decades after the discovery of the structure of DNA and the genetic code, the field of human genetics was largely focused on understanding the structure and function of protein-coding genes and how rare mutations in these genes cause disease. Indeed, the central dogma of molecular biology posits that genes are first transcribed into messenger RNA (mRNA), after which the mRNA is translated into protein.1 Because of the straightforward nature of the genetic code, it was easy to predict how alterations of the underlying DNA sequence would change the amino acid composition of the resulting protein.2 In addition, it was clear from Mendelian genetics that diseases that run in families in predictable patterns are caused by mutations in a single gene. Thus, beginning with the mapping of the genetic cause of the neurodegenerative disorder Huntington's Disease in 1983,3 the causative mutations underlying many Mendelian diseases were elucidated by positional cloning, and an important hurdle was overcome in our understanding of the genetic bases of human disease. Today, the genetic lesions responsible for many Mendelian diseases are known, and frequently researchers have determined how the mutation in question affects protein function, resulting in pathophysiology. However, many of the most common and burdensome diseases, such as cardiovascular disease, cancer, Alzheimer's disease, Parkinson's disease, and type 2 diabetes, are typically not (or never) caused by single mutations. Such "complex traits" are instead influenced by a combination of multiple genetic and environmental risk factors, and thus do not follow Mendelian inheritance patterns. The departure from a "one-gene, one-mutation, one-outcome" model poses a formidable challenge to elucidating the biology of these diseases. Complex traits, by definition, are influenced by many genes which may interact in additive or non-additive ways. Yet, while it may not always be necessary to understand the cause of a disease in order to successfully treat it, such a mechanistic understanding certainly increases the likelihood that a successful therapeutic intervention will be achieved. The GWAS era has been successful in the sense that thousands of loci have been statistically associated with risk for diseases and traits, and a notable number of these loci are well-replicated, suggesting that they are true associations. However, several factors have made it difficult to bridge the gap between the statistical associations linking locus and trait and a functional understanding of the biology underlying disease risk. First, the association of a locus with disease does not specify which variant (or variants) at that locus is actually causing the association nor which gene (or genes) is affected by the causal variant (the "target gene"). The former problem is due to the fact that there are often many co-inherited variants in strong linkage disequilibrium (LD) with the most significant (or "sentinel") disease-associated variant, comprising a haplotype; within the haplotype, genetic variants in strong LD often have statistically indistinguishable associations with disease risk. As a consequence, empirical validation might be needed to determine which of the linked variants are functional. The latter complication results from the fact that > 90% of disease-associated variants (daVs) are located in non-protein-coding regions of the genome, and many are far away from the nearest known gene. What might these non-coding variants be doing? One clue arises from the observation that daVs, as well as variants in strong LD with them, are enriched in predicted transcriptional regulatory regions, called "cis-regulatory elements" (CREs). This suggests that many loci implicated by GWASs to affect disease risk might do so by altering the genetic regulation of one or more target genes. However, the complex nature of eukaryotic transcriptional regulation can make it difficult to assign putative CREs to their correct target genes, necessitating the use of genomic datasets and experimental approaches to help answer this question. Indeed, while several thousand GWASs have been performed, and many thousands of loci have been confirmed as bona fide disease risk factors, the number of studies that have investigated the mechanisms underlying particular associations is orders of magnitude fewer, and the number of studies that have functionally characterized candidate causal variants at a given locus in an objective manner is even fewer still.

Bottom Note: *This work is partly presented at EuroSciCon conference on Protein, Proteomics and Computational Biology*
December 06-07, 2018 Amsterdam, Netherlands