Pleiotropic Enrichment Analysis with Diverse Omics Data

Wenlong Tang1*, Jigang Zhang2,3 and Dongdong Lin2,4

1Department of Electrical and Computer Engineering, University of Alabama, Tuscaloosa, AL, USA
2Center of Genomics and Bioinformatics, Tulane University, New Orleans, LA, USA
3Department of Biostatistics and Bioinformatics, Tulane University, New Orleans, LA, USA
4Department of Biomedical Engineering, Tulane University, New Orleans, LA, USA

*Corresponding author: Wenlong Tang, Department of Electrical and Computer Engineering, University of Alabama, Tuscaloosa, AL, 35487, USA, Tel: 504-988-1341; E-mail: wtang11@bama.ua.edu

Rec date: 31 Oct, 2014; Acc date: 03 Nov, 2014; Pub date: 05 Nov, 2014

Introduction

In human complex diseases, the phenomenon that a genetic variant or gene can affect multiple diseases (or phenotypes) is referred as “Pleiotropy” [1]. Pleiotropy is usually considered as the most important source for genetic correlation between traits [2]. The pleiotropy highlights the fact that some genes in human genome perform multiple biological functions. For example, epidemiological studies indicate that individuals with increased genetic susceptibility to diabetes have the decreased risk of prostate cancer [3,4]. The association analysis between these two diseases show that several genes can affect both type 2 diabetes and prostate cancer risk via pleiotropy [3,5]. The identification and characterization of pleiotropic genes can offer a unique way to dissect complex etiology of these diseases.

Although examples of pleotropic genes have been discovered in human complex diseases [6], pleiotropy remains a poorly understood and there have been very few systematic studies [7]. Recently systematic phenome-wide association studies (PheWASs) have been proposed, in which a SNP with an established association with a phenotype is screened for association with hundreds of other phenotypes [8,9]. Compared with single-trait analysis, the multiple-trait analysis in PheWASs can combine the information from genetically correlated traits to identify pleotropic genes associated with disease risk [10]. Genome-wide association studies with an abundance of genetic variants and various traits have provided an opportunity to fully examine pleiotropy in a systematic manner for the human complex diseases. For example, numerous approaches extended the regression framework for testing the association between a genetic variant and multiple correlated phenotypes, using variations of generalized estimating equations [11]. Other approaches included a dimension reduction technique, e.g. Principal components analysis, on the phenotypes before testing the association with the genetic variant in a genetic association analysis [12]. There were also approaches to combine the association results across various phenotypes by meta-analysis to identify those variants that are associated with multiple phenotypes [13].

But these methods mainly focused on the detection of single pleiotropic SNPs or genes in single omics data. It is increasingly recognized that complex diseases result from the joint effects of multiple genes which are from the same biological pathways [14,15]. Genes that belong to the same pathway might have a similar pattern of pleiotropic effects. Therefore, only testing pleiotropy of single SNPs or genes is insufficient to dissect the genetic structures of complex diseases. Additionally, identification of pleiotropic genes for complex diseases is limited by the relatively modest effects of genetic variants [16]. Thus it is especially meaningful to identify groups of pleiotropic pathways associated with the traits of interest, in which there are enrichment of pleiotropic genes.

In the past few years, the availability of high-throughput techniques can lead to the opportunity to generate different omics data (genomics, transcriptomics, epigenomics, and proteomics data) for systematic analysis of genome-wide levels of pleiotropy at multi-omics levels. Characterizing the underlying biological mechanism of a pleotropic effect is a big challenge with multi-omics data. Systems-level approaches must be developed to detect pleiotropic variants and pathways (network modules). We aim to gain insight into the extent and pattern of pleiotropy in the genetics of complex disease to characterize pleiotropic genes for clusters of diseases and disease traits.

Here we present a Pleiotropic Enrichment Analysis (PEA) for multi-omics data to identify pleiotropic functional modules and genes underlying multiple genetically related diseases as following:

First, systematically identify pleiotropic genes. We can compute crosstalk scores by combining the pleiotropic evidence of genes obtained from individual uni-omics studies through a meta-analysis based approach [17].

Second, perform pathway-enrichment analysis with Signaling Pathway Impact Analysis to identify the pathway enriched with pleiotropic genes [18], which calculates the significance of a pathway according to both the over-representation evidence and perturbation-based evidence using the topology of the network.

Third, incorporate the prior knowledge for pleiotropic network reconstruction. This strategy incorporates the information of prior knowledge (e.g. protein-protein interaction network and online databases for miRNA target prediction) into pathway enriched with pleiotropic genes to rebuild pleiotropic regulation modules [19,20].

Overall, we provide a general guideline for detection of pleiotropy using multi-omics data by PEA. The identification and characterization of pleiotropy is crucial for a comprehensive biological understanding of complex disease. As genetic information from different omics data is increasingly integrated into medical practice, characterizing pleiotropic effects may improve the accuracy of these genetic tests and the interpretation of results on diseases. The existence of pleiotropic pathways (or modules) in distinct disorders may suggest new opportunities and challenges for drug discovery. Drugs developed for one disorder could be repurposed to treat another disorder if the therapeutic target genetic factors is common to the biology of both disorders. Furthermore, we anticipate that the pleiotropic pathways/modules might give clues to underlying common mechanisms on genetically correlated diseases.
References