Journal of Phylogenetics & Evolutionary Biology

Xiong, J Phylogen Evolution Biol 2015, 3:2
DOI: 10.4172/2329-9002.1000e113

Editorial Open Access

Causal Genomic and Epigenomic Network Analysis emerges as a New Generation of Genetic Studies of Complex Diseases

Momiao Xiong*

Human Genetics Center, Department of Biostatistics, The University of Texas School of Public Health, Houston, TX 77030, USA

In the past decade, rapid advances in genomic technologies have dramatically changed the genetic studies of complex diseases. Genome-wide association studies (GWAS) have been widely used in dissecting genetic structure of complex diseases. As of December 18th, 2014, A Catalog of Published Genome-Wide Association Studies (GWAS) had reported significant association of 15,177 SNPs with more than 700 traits in 2,087 publications [1]. However, numerous studies reported that the genetic loci identified by GWAS collectively explain only < 10% of genetic variation across the population in most complex diseases. About 90% of the heritability of common diseases are unexplained by a large number of identified GWA loci. Each variant usually has weak effect and make small and mild contributions to the disease. More than 1,000 loci for many complex diseases have been identified [2]. Although extremely large number of samples are collected and whole genome sequencing studies will be conducted very soon, which will lead to reducing he fraction of missing heritability, a large proportion of heritability will be still missing under the paradigm of single trait genetic analysis. The methods for heritability estimation and single trait genetic study paradigm are questionable.

A biological system consists of multiple phenotypes. The multiple phenotypes are correlated. It has been reported that more than 4.6% of the SNPs and 16.9% of the genes in previous genome-wide association studies (GWAS) were significantly associated with more than one trait [3]. These results demonstrate that genetic pleiotropic effects likely play a crucial role in the molecular basis of correlated phenotype [4]. The heritability of individual phenotype cannot reveal complicated genotype-phenotype structure and is highly unlikely to fully capture the structure of heritability of multiple phenotypes. Furthermore, the estimation of heritability by a single trait approach might be inaccurate. The concept of heritability should be extended from a single trait to multiple traits.

Consider k traits. The breeding and phenotype values for k traits are denoted by a k dimensional vector

$A= [A_{1,...,}A_k]$ and $P= [P_1,... P,_k]^T$, respectively. A breeding equation is given by

$$A=HP \tag{1}$$

Where H is a heritability matrix and denoted by

$$H=\begin{bmatrix} h_1 & \cdots & h_{1k} \\ \vdots & \ddots & \vdots \\ h_{k1} & \cdots & h_k \end{bmatrix}$$

Suppose that the phenotype can) be decomposed as a summation of additive effect, dominant effect and environment effect:k

$$P=A+D+E, \text{ where} \tag{2}$$

A, D and E represent the genetic additive, dominant and environmental effect, respectively. Denote the covariance matrix between the breeding value and phenotype values by

$$\mathrm{cov}(A,P)=\begin{bmatrix} \mathrm{cov}(A_1,P_1) & \cdots & \mathrm{cov}(A_1,P_k) \\ \vdots & \ddots & \vdots \\ \mathrm{cov}(A_k,P_1) & \vdots & \mathrm{cov}(A_k,P_k) \end{bmatrix}$$

and variance-covariance matrix of the phenotype P by

$$\mathrm{var}(P)=\begin{bmatrix} \mathrm{var}(P_1) & \cdots & \mathrm{cov}(P_1,P_k) \\ \vdots & \ddots & \vdots \\ \mathrm{cov}(P_k,P_1) & \cdots & \mathrm{var}(P_k) \end{bmatrix}$$

It is known that

$$\mathrm{cov}(A_i,P_j)=\mathrm{cov}(A_i,A_j)+\mathrm{cov}(A_i,D_j)+\mathrm{cov}(A_i,E_j),$$

which implies that

$$\mathrm{cov}(A,P)=\begin{bmatrix} \mathrm{cov}(A_1,A_1)+\mathrm{cov}(A_1,D_1)+\mathrm{cov}(A_1,E_1) & \cdots & \mathrm{cov}(A_1,A_k)+\mathrm{cov}(A_1,D_k)+\mathrm{cov}(A_1,E_k) \\ \vdots & \ddots & \vdots \\ \mathrm{cov}(A_k,A_1)+\mathrm{cov}(A_k,D_1)+\mathrm{cov}(A_k,E_1) & \vdots & \mathrm{cov}(A_k,A_k)+\mathrm{cov}(A_k,D_k)+\mathrm{cov}(A_k,E_k) \end{bmatrix}$$

It follows from equation (1) that the heritability matrix is estimated by

$$H= \mathrm{COV}(A, P)[\mathrm{var}(P)]^{-1} \tag{3}$$

Equation (3) shows that the heritability of the i^{th} trait h_{ii} is a function of the genetic covariance between the i^{th} trait and other traits. In other words, the heritability of each trait is influenced by its correlation with other multiple traits. This clearly demonstrates that the trait by trait genetic study will overlook the influence of other traits. The missing heritability may be due to trait by trait genetic analysis. The joint genetic analysis of multiple traits may increase the heritability.

There has been increasingly consensus that individual genetic and epigenetic variants, individual genes, individual linear pathway and individual trait analysis cannot capture the intrinsic genetic and epigenetic complexity of multiple phenotypes.

To completely capture the heritability, the right research direction is to jointly investigate genetic, expression, miRNA, epigenetic, metabolic variants, physiological traits, medical imaging measurements and environments in multiple traits which are often interactively organized networks. Integrative analysis of genetic, epigenetic, imaging and environmental variation in multiple phenotypes will fully uncover the heritability and facilitate the understanding the mechanism of the complex diseases. The popular methods for integrative analysis are mainly based on correlation and association analysis. These methods cannot efficiently detect, distinguish and characterize the true biological, mediated and spurious pleiotropic effects. Therefore, these approaches may not provide clear biologically or clinical relevant information that allows the mechanisms of genetic effects to be discovered and understood. To overcome these limitations, developing a new framework and novel statistical methods for inferring causal networks

***Corresponding author:** Momiao Xiong, Human Genetics Center, Department of Biostatistics, The University of Texas School of Public Health, Houston, TX 77030, USA E-mail: Momiao.Xiong@uth.tmc.edu

Received May 20, 2015; **Accepted** May 25, 2015; **Published** March 30, 2015

Citation: Xiong M (2015) Causal Genomic and Epigenomic Network Analysis emerges as a New Generation of Genetic Studies of Complex Diseases. J Phylogen Evolution Biol 3: e113. doi:10.4172/2329-9002.1000e113

of genotype-phenotypes with NGS data and detecting, distinguishing and characterizing the true biological pleiotropic, mediated pleiotropic and spurious pleiotropic effects of genetic variants are urgently needed.

An essential issue for using causal graphs to study genetics of multiple phenotypes is how to accurately and efficiently estimate the structure of causal graph from observational data. Structure learning of casual graphs has been shown to be NP-hard. Early methods for structure learning mainly focused on approximation algorithms, but such methods are unable to ensure the generation of the true causal graph. To obtain the causal graph from observation data as close to the biological causal graph as possible, "score and search"-based methods for exact learning causal graphs of genotype-phenotype to find the best-scoring structures for a given dataset are being developed. The accurate and robust estimation of the genotype-phenotype causal networks by the "score and search" methods will shift the paradigm of genetic studies of correlated multiple phenotypes from association analysis to causal inference, and dramatically facilitate discovery of the mechanism underlying multiple traits.

Although their application to genome-wide genotype-phenotype network construction is difficult due to computational limitations, the "score and search" based causal inference methods are suitable to the phenome-wide association studies where starting phenomics, defined as the unbiased study of a large number of phenotypes in a population. We study the complex networks between multiple expressed phenotypes and genetic variants. Since the number of genetic variants in the phenome-wide association is quite limited and hence the size of the genotype-phenotype network is limited, the required computational time of construction of genotype-phenotype networks using causal inference is in the range the current computer system can reach. Advances in biosensors and sequencing technologies generate large amounts of phenotype and genetic data. Causal genetic and epigenetic network analysis may emerge as a new paradigm of genetic studies of complex traits. The main purpose of this editorial is to stimulate discussion about what are the optimal strategies to facilitate the development of a new generation of genetic analysis. I hope that more and more real data analysis in the future will greatly increase the confidence in causal inference for genotype-phenotype studies.

References

1. van der Sijde MR, Ng A, Fu J (2014) Systems genetics: From GWAS to disease pathways. Biochim Biophys Acta 1842: 1903-1909.
2. Björkegren JL, Kovacic JC, Dudley JT, Schadt EE (2015) Genome-wide significant loci: how important are they? Systems genetics to understand heritability of coronary artery disease and other common complex disorders. J Am Coll Cardiol 65: 830-845.
3. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW (2013) Pleiotropy in complex traits: challenges and strategies. Nat Rev Genet 14: 483-495.
4. Wagner GP, Zhang J (2011) The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. Nature Rev. Genet 12: 204-213.