Bioinformatics, Proteomics and Immaging Analysis

OMMEGA Publishers
Answering Intelligence
www.ommegaonline.com

Editorial

OPEN ACCESS

Searching Genome-wide Disease Associations in Case-Control Studies

Jing Zhang[1*], Xuan Guo[2], Yi Pan[2]

[1]Department of Mathematics and Statistics, Georgia State University, USA
[2]Department of Computer Science, Georgia State University, Atlanta, Georgia, USA

Corresponding Author: Zhang, J. Department of Mathematics and Statistics, Georgia State University, USA.
E-mail: jzhang47@gsu.edu

Received Date: Sep 08, 2015
Accepted Date: Sep 09, 2015
Published Date: Sep 15, 2015

Citation: Zhang, J., et al. Searching Genome-wide Disease Associations in Case-Control Studies (2015) Bioinfo Proteom Img Anal 1(2): 36-37.

Introduction

With the high-throughput genotyping technology of Single Nucleotide Polymorphism (SNP), Genome-Wide Association Studies (GWASs) are considered carrying hope for resolving complex connections between genotype and phenotype[1]. GWASs intend to recognize genetic variants associated with disease by assaying and analyzing numbers of SNPs. Although traditional single-locus-based and two-locus-based methods have been standardized and led to various exciting findings, recently, a substantial number of GWASs show that, for most disorders, joint genetic effects (epistatic interaction) across the entire genome are broadly existing in complex traits[2]. At present, identifying high-order epistatic interactions from GWASs is computationally and methodologically challenging.

Our lab's research interest focuses on the problem of searching high-order genome-wide association with considering two frequently encountered situations, i.e. one case one control and multi-cases multi-controls. Existing approaches for exploring epistatic interactions for the first situation can be classified into four general categories, exhaustive search, stepwise search, stochastic search and heuristic approaches. In the review of recent works of literature, we recognize 47 methods applied to detect epistasis, excluding specializations, tweaks, and merely paralleled methods[2,3]. The naive solution to tackle the problem of detecting epistatic interaction is exhaustive exploration employing χ^2 test, exact likelihood ratio test or entropy-based test for any module of multiple-locus. However, finding higher order (more than two loci) disease-related associations are extremely computationally costly to be feasible, particularly for a GWAS project with millions of SNPs. Instead of explicitly enumerating all possible combinations of k-locus, stepwise search approaches first select a subset of SNPs based on single-locus tests or model-free measures, then conduct tests for multi-locus interactions on the selected subset of SNPs. Similarly, stochastic methods use random sampling procedures to search the space of interactions. Likewise, heuristics approaches adopt machine learning techniques, such as neural networks and predictive rules, to explore the space of epistatic interactions rather than explicitly enumerating and testing all the combinations of k-locus. Since non-exhaustive methods will exclude a substantial portion of SNPs, they may not be able to detect interactions involving loci with small or no marginal effects.

To address the time-consuming issue and improve the accuracy as well, we provide a novel approach, named "Dynamic Clustering for High-order genome-wide Epistatic interactions detecting" (DCHE)[4]. DCHE uses an elegant dynamic clustering scheme to maximize statistical significance for SNP combinations and ranks top ones as results. DCHE applies statistic tests on merged groups of genotypes determined by the dynamic clustering. Each grouped genotype category may share a similar effect associating with corresponding phenotypes. Truly disease-related joint genetic effects will win higher ranking scores with the condition that genotype combinations have been correctly clustered together. Systematic analyses on simulated two- and three-locus disease models datasets confirm that DCHE is more powerful in detecting epistatic interactions than some recently developed methods including TEAM[5], SNPRuler[6], BOOST[1] and EDCF[7]. Our analyses on two real genome-wide case/control data sets, Age-related macular degeneration (AMD) and Rheumatoid arthritis (RA)[8] show that DCHE is possible for the full-scale investigations of multi-locus associations on large GWAS datasets, and it enriches many novels, significant high-order epistatic interactions that have not been reported in the literature.

For the second situation, we presented our follow-up study entitled

Copy rights: ©2015 Zhang, J. This is an Open access article distributed under the terms of Creative Commons Attribution 4.0 International License.

"DAM: A Bayesian Method for Genome-wide Associations Detecting on Multiple Diseases" at the 11th International Symposium on Bioinformatics Research and Applications (ISBRA) in Norfolk, Virginia, US on June 9, 2015[9]. We designed and implemented a Bayesian inference method for Detecting genome-wide Association on Multiple diseases, named DAM, to deal with multiple cases in a GWAS dataset. DAM employs the Markov Chain Monte Carlo (MCMC) sampling based on the Bayesian Variable Partition (BVP) model[10], and also makes the use of a stepwise condition evaluation procedure to identify significant disease(s)-specific interactions. It first produces a candidate set of SNPs by applying the BVP model, which can capture the disease-specific associations, with the Metropolis-Hastings (MH) algorithm. Following that, a stepwise association evaluation procedure is engaged in detecting the genetic effect types and removing redundant SNPs in a module. Experiments on both simulated and two real GWAS datasets, i.e. Rheumatoid Arthritis (RA) and, Type 1 Diabetes (T1D), show that our method is feasible for identifying multi-locus interaction on multiple GWAS datasets, and it also reports some significant high-order epistatic interactions with specialties on various diseases. For instance, rs1230649 dwells within the coding region of the gene, PHTF1 (putative homeo domain transcription factor 1). PHTF1 can recruit FEM1B to the endoplasmic reticulum membrane, and FEM1B belonging to the death receptor-associated family of proteins presents a significant role in mediating apoptosis[11]. The associated SNP with rs1230649 is rs11984645, which resides near the gene, MRPL15, mitochondrial ribosomal protein L15. By utilizing the LD plot from HapMap and the NCBI dbSNP, we found rs1230647 and MRPL15 are both inside a block caused by LD effect. MRPL15 is encoded by nuclear genes and helps in protein synthesis within the mitochondrion[12].

A large number of SNPs genotyped in genome-wide association studies poses a significant computational challenge in the identification of gene-gene interactions. During the last few years, there have been fast-growing interests in developing and applying computational and statistical approaches to detecting gene-gene interactions. By comparing to other popular tools, our methods show significant improvement in terms of the discrimination power. Experimental results on real data prove that our two methods can discover remarkable novel biologically significant associations.

References

1. Wan, X., Yang, C., Yang, Q., et al. BOOST: A fast approach to detecting gene-gene interactions in genome-wide case-control studies. (2010) The Am J Hum Genet 87(3): 325-340.
2. Guo, X., Yu, N., Gu, F., et al. Genome-wide interaction-based association of human diseases-a survey. (2014) Tsinghua Science and Technolo 19(6): 596-616.
3. Shang, J., Zhang, J., Sun, Y., et al. Performance analysis of novel methods for detecting epistasis. (2011) BMC bioinformatics 12(1): 475.
4. Guo, X., Meng, Y., Yu, N., et al. Cloud computing for detecting high-order genome-wide epistatic interaction via dynamic clustering. (2014) BMC bioinformatics 15: 102.
5. Zhang, X., Huang, S., Zou, F., et al. TEAM: Efficient two-locus epistasis tests in human genome-wide association study. (2010) Bioinformatics 26(12): i217-i227.
6. Wan, X., Yang, C., Yang, Q., et al. Predictive rule inference for epistatic interaction detection in genome-wide association studies. (2010) Bioinformatics 26(1): 30-37.
7. Xie, M., Li, J., Jiang, T. Detecting genome-wide epistases based on the clustering of relatively frequent items. (2012) Bioinformatics 28(1): 5-12.
8. Ding, X., Wang, J., Zelikovsky, A., et al. Searching high-order SNP combinations for complex diseases based on energy distribution difference. (2015) IEEE/ACM Trans Comput Biol Bioinform 12(3): 695-704.
9. Guo, X., Zhang, J., Cai, Z., et al. DAM: A Bayesian Method for Detecting Genome-wide Associations on Multiple Diseases. (2015) Bioinform Res App 96-107.
10. Zhang, J., Hou, T., Wang, W., et al. Detecting and understanding combinatorial mutation patterns responsible for HIV drug resistance. (2010) Proc Natl Acad Sci 107(4): 1321-1326.
11. Zhang, J., Zhang, Q., Lewis, D., et al. A Bayesian Method for Disentangling Dependent Structure of Epistatic Iteraction. (2011) Ameri J Biostat 2(1): 1-10.
12. Zhang, Y., Zhang, J., Liu, J. S. Block-based Bayesian epistasis association mapping with application to WTCCC type 1 diabetes data. (2011) The Ann Appl stat 5(3): 2052-2077.