

Gene-gene and gene-environment interactions underlying complex traits and their detection

Abbreviations: MDR, multifactor dimensionality reduction method; CART, classification and regression trees; MARS, multivariate adaptive regression spline; GPNN, genetic programming optimized neural network

Introduction

No genes or environmental factors are isolated from the interactive genomic and epigenomic networks in shaping a biological phenotype. Non intuitivity and nonlinearity are a natural property of the network's architecture.¹⁻⁴ Consequently, the existence of interactions among genes, called gene-gene (also known as epistatic) interactions, and between genes and environmental factors (broadly defined as all non-genetic exposures), called gene-environment (GE) interactions, is the normal rather than an exception.⁵⁻⁸ Several converging lines of evidence have pointed to the dominant role of interactions in the inherited traits;⁶⁻⁹ in particular, epistatic and GE interactions are considered as one of the primary culprits for missing heritability,^{10,11} referred to the majority of the genetic variation that is not yet identified by the more than a decade's practice of genome-wide association studies.¹²⁻¹⁴ Identification of background-specific factors among genes in combination with lifestyles and environmental exposures is an important scientific topic in genetics, breeding, and genetic epidemiology.

A high degree of context dependence of genetic architecture likely results in a relatively weak marginal genotype-phenotype correlations for complex traits, making traditional univariate approaches that test for association one factor at a time futile.^{5,11} The multi factorial strategies are thus critical in hunting highly mutually dependent factors underlying a trait. However, such a search has to face a significant obstacle called "the curse of dimensionality", a problem caused by the exponential increase in volume of possible interactions with the number of factors to consider.¹⁵ The conventional regression methods, established by the extension under the concept of single factor-based approaches, are hardly appropriate for tackling ubiquitous yet elusive interactions because of several problems, heavy computational burden (usually computationally intractable), increased Type I and II errors, and reduced robustness and potential bias as a result of highly sparse data in a multi factorial model.¹⁶ Diverse novel approaches such as data mining and machine learning have been explored recently for various kinds of phenotypes,¹⁷⁻¹⁹ namely, Bayesian belief network,^{20,21} tree-based algorithms including multivariate adaptive regression spline (MARS),²² classification and regression trees (CART) or recursive partitioning methods²³⁻²⁵ and random forests approach,^{26,27} pattern recognition approaches including neural network strategies such as the parameter decreasing method (PDM)²⁸ and genetic programming optimized neural network (GPNN),²⁹ genetic algorithm strategies,³⁰ and cellular automata (CA) approach,³¹ support vector machine (SVM),³² penalized regression,³³ and Bayesian methods.^{34,35}

Among these methods emerged recently, data reduction approaches (a constructive induction strategy) such as the multifactor dimensionality reduction method (MDR),^{36,37} the combinatorial partitioning method,³⁸ and the restricted partition method,³⁹ are promising to address the multidimensionality problems. Rather than modeling the interaction term per se as with regression methods,

Volume 1 Issue 2 - 2014

Xiang-Yang Lou

Department of Biostatistics, University of Alabama at Birmingham, USA

Correspondence: Xiang-Yang Lou, Department of Biostatistics, University of Alabama at Birmingham 1665 University Boulevard, RPHB 327, Birmingham, Alabama 35294-0022, USA, Tel 205-975-9145, Fax 205-975-2541, Email xyLou@uab.edu

Received: September 18, 2014 | **Published:** October 07, 2014

a data reduction strategy seeks for a pattern in a combination of factors/attributes of interest that maximizes the phenotypic variation it explains. It treats the joint action as a whole, coinciding to the very original epistasis coined by Bateson,⁴⁰ offering a solution that avoids decomposition as in regression methods where the number of interaction parameters grows exponentially as each new variable is added. It also has a straightforward correspondence to the concept of the phenotypic landscape that unifies biological, statistical genetics and evolutionary theories.⁴¹⁻⁴⁵ Notably the pioneering MDR method has sustained its popularity in detection of interactions since its launch.⁴⁶

Several extensions of the MDR have been made for analyzing different traits, e.g., binary, count, continuous, polytomous, ordinal, time-to-onset, multivariate and others, as well as combinations of those, and also entertaining various study designs including homogeneous and admixed unrelated-subject and family as well as mixtures of them.⁴⁷ Such extensions include to inclusion of covariates,^{48,49} to continuous traits,⁴⁹ to survival data,^{50,51} to multivariate phenotypes,^{52,53} to multi-categorical or ordinal phenotypes,^{47,54} to case-control study in structured populations,^{55,56} to family study,^{57,58} and to unified analysis of both unrelated and related samples.⁵⁹ With these extensions, the MDR-type methods offer a powerful tool for handling the breadth of data types and addressing statistical issues associated with study design and sampling scheme.

Despite the methodological progresses in detection of multifactor interactions, there are still difficult computational challenges and multiple hypothesis testing problems in practice, especially for detecting high-order interactions for the large-scale such as whole genome data. Further theoretical and computational work is required for effective identification of interacting factors underlying the complex traits.

Acknowledgement

The author thanks Guo-Bo Chen, Hai-Ming Xu, Xi-Wei Sun, and Lei Yan for their contributions to the development of GMDR. This project was supported in part by NIH Grant DA025095 to X.-Y.L.

Conflict of interest

The author declares no conflict of interest.

Funding

None.

References

- Barry P. No gene is an Island: Even as biologists catalog the discrete parts of life forms, an emerging picture reveals that life's functions arise from interconnectedness. *Science News*. 2008;174(12):22–26.
- Szathmary E, Jordan F, Pal C. Molecular biology and evolution. Can genes explain biological complexity? *Science*. 2001;292(5520):1315–1316.
- Hartwell L. Genetics, Robust interactions. *Science*. 2004;303(5659):774–775.
- Nijhout HF. On the association between genes and complex traits. *Journal of Investigative Dermatology Symposium Proceedings*. 2003;8:162–163.
- Carlson CS, Eberle MA, Kruglyak L, et al. Mapping complex disease loci in whole-genome association studies. *Nature*. 2004;429(6990):446–452.
- Shaoa H, Lindsay CB, David SS, et al. Genetic architecture of complex traits: Large phenotypic effects and pervasive epistasis. *PNAS*. 2008;105(50):19910–19914.
- Huang W, Richards S, Carbone MA, et al. Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. *Proc Natl Acad Sci*. 2012;109(39):15553–15559.
- Zuka O, Eliana H, Shamil RS, et al. The mystery of missing heritability: Genetic interactions create phantom heritability. *PNAS*. 2011;109(4):1193–1198.
- Stratton MR, Rahman N. The emerging landscape of breast cancer susceptibility. *Nat Genet*. 2008;40(1):17–22.
- Frazer KA, Murray SS, Schork NJ, et al. Human genetic variation and its contribution to complex traits. *Nat Rev Genet*. 2009;10(4):241–251.
- Phillips PC. Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet*. 2008;9(11):855–867.
- Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461(7265):747–753.
- Eichler EE, Flint J, Gibson G, et al. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet*. 2010;11(6):446–450.
- Maher B. Personal genomes: The case of the missing heritability. *Nature*. 2008;456(7218):18–21.
- Moore JH, Ritchie MD. STUDENT JAMA: The challenges of whole-genome approaches to common diseases. *The JAMA*. 2004;291(13):1642–1643.
- Carlborg O, Haley CS. Epistasis: too often neglected in complex trait studies? *Nat Rev Genet*. 2004;5(8):618–625.
- Cordell HJ. Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet*. 2009;10(6):392–404.
- Heidema GA, Jolanda MA B, Nico N, et al. The challenge for genetic epidemiologists: how to analyze large numbers of SNPs in relation to complex diseases. *BMC Genet*. 2006;7:23.
- Motsinger AA, Ritchie MD, Reif DM. Novel methods for detecting epistasis in pharmacogenomics studies. *Pharmacogenomics*. 2007;8(9):1229–1241.
- Sebastiani P, Ramoni MF, Nolan V, et al. Genetic dissection and prognostic modeling of overt stroke in sickle cell anemia. *Nat Genet*. 2005;37(4):435–440.
- Hornig JT, Hu KC, Wu LC, et al. Identifying the combination of genetic factors that determine susceptibility to cervical cancer. *IEEE Trans Inf Technol Biomed*. 2004;8(1):59–66.
- Cook NR, Zee RY, Ridker PM. Tree and spline based association analysis of gene-gene interaction models for ischemic stroke. *Stat Med*. 2004;23(9):1439–1453.
- Tracy JC, Michael DS, Mahyar S, et al. Use of tree-based models to identify subgroups and increase power to detect linkage to cardiovascular disease traits. *BMC Genet*. 2003;4(S1):S66.
- Province MA, Shannon WD, Rao DC. Classification methods for confronting heterogeneity. *Adv Genet*. 2001;42:273–286.
- Shannon WD, Province MA, Rao DC. Tree-based recursive partitioning methods for subdividing sibpairs into relatively more homogeneous subgroups. *Genet Epidemiol*. 2001;20(3): 293–306.
- Lunetta KL, Hayward LB, Segal J, et al. Screening large-scale association study data: exploiting interactions using random forests. *BMC Genet*. 2004;5:32.
- Xiang C, Ching-Ti Liu, Meizhuo Z, et al. A forest-based approach to identifying gene and gene-gene interactions. *Proc Natl Acad Sci*. 2007;104(49):19199–19203.
- Tomita Y, Tomida S, Hasegawa Y, et al. Artificial neural network approach for selection of susceptible single nucleotide polymorphisms and construction of prediction model on childhood allergic asthma. *BMC Bioinformatics*. 2004;5:120.
- Ritchie MD, White BC, Parker JS, et al. Optimization of neural network architecture using genetic programming improves detection and modeling of gene-gene interactions in studies of human diseases. *BMC Bioinformatics*. 2003;4:28.
- Jason HM, Lance WH, Marylyn DR, et al. Routine discovery of complex genetic models using genetic algorithms. *Appl Soft Comput*. 2004;4(1):79–86.
- Moore JH, Hahn LW. A cellular automata approach to detecting interactions among single-nucleotide polymorphisms in complex multifactorial diseases. *Pac Symp Biocomput*. 2002;53–64.
- Chen SH, Sun J, Dimitrov L, et al. A support vector machine approach for detecting gene-gene interaction. *Genet Epidemiol*. 2008;32(2):152–167.
- Park MY, Hastie T. Penalized logistic regression for detecting gene interactions. *Biostatistics*. 2008;9(1): 30–50.
- Zhang Y, Liu JS. Bayesian inference of epistatic interactions in case-control studies. *Nat Genet*. 2007;39(9):1167–1173.
- Nengjun Yi, Brian SY, Gary AC, et al. Bayesian model selection for genome-wide epistatic quantitative trait loci analysis. *Genetics*. 2005;170(3):1333–1344.
- Ritchie MD, Hahn LW, Roodi N, et al. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet*. 2001;69(1):138–147.
- Hahn LW, Ritchie MD, Moore JH. Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. *Bioinformatics*. 2003;19(3):376–382.
- Nelson MR, Kardia SLR, Ferrell RE, et al. A combinatorial partitioning method to identify multilocus genotypic partitions that predict quantitative trait variation. *Genome Res*. 2001;11(3):458–470.
- Culverhouse R, Klein T, Shannon W. Detecting epistatic interactions contributing to quantitative traits. *Genet Epidemiol*. 2004;27(2):141–152.
- Bateson W. *Mendel's principles of heredity*. UK: Cambridge University Press; 1909.

41. Nijhout HF. Developmental phenotypic landscapes. *Evol Biol.* 2008;35(2):100–103.
42. Wright S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proceedings of the Sixth International Congress on Genetics.* 1933;1:356–366.
43. Rice SH. A general population genetic theory for the evolution of developmental interactions. *Proc Natl Acad Sci.* 2002;99(24):15518–15523.
44. Wolf JB. The geometry of phenotypic evolution in developmental hyperspace. *Proc Natl Acad Sci.* 2002;99(25):15849–15851.
45. Nijhout HF. The nature of robustness in development. *Bio essays.* 2002;24(6):553–563.
46. Motsinger AA, Ritchie MD. Multifactor dimensionality reduction: an analysis strategy for modeling and detecting gene-gene interactions in human genetics and pharmacogenomics studies. *Hum Genomics.* 2006;2(5):318–328.
47. Lou XY. UGMDR: A unified conceptual framework for detection of multifactor interactions underlying complex traits. *Heredity.* 2014.
48. Lee SY, Chung Y, Elston RC, et al. Log-linear model-based multifactor dimensionality reduction method to detect gene-gene interactions. *Bioinformatics.* 2007;23(19):2589–2595.
49. Xiang Yang L, Guo Bo C, Lei Yan, et al. A generalized combinatorial approach for detecting gene-by-gene and gene-by-environment interactions with application to nicotine dependence. *Am J Hum Genet.* 2007;80(6):1125–1137.
50. Gui J, Moore JH, Kelsey KT, et al. A novel survival multifactor dimensionality reduction method for detecting gene-gene interactions with application to bladder cancer prognosis. *Human Genetics.* 2011;129(1):101–110.
51. Lee S, Kwon MS, Oh JM, et al. Gene-gene interaction analysis for the survival phenotype based on the Cox model. *Bioinformatics.* 2002;28(18):i582–i588.
52. Choi J, Park T. Multivariate generalized multifactor dimensionality reduction to detect gene-gene interactions. *BMC Syst Biol.* 2016;7(S6): S15.
53. Xu HM, Sun XW, Ting Qi, et al. Multivariate dimensionality reduction approaches to identify gene-gene and gene-environment interactions underlying multiple complex traits. *PLoS One.* 2014.
54. Kim K, Kwon MS, Oh S. Identification of multiple gene-gene interactions for ordinal phenotypes. *BMC Med Genomics.* 2013;6(S2): S9.
55. Niu A, Zhang S, Sha Q. A Novel Method to Detect Gene-Gene Interactions in Structured Populations: MDR-SP. *Ann Hum Genet.* 2011;75(6):742–754.
56. Lou XY. A PCA-based Generalized Multifactor Reduction Method for Correcting Population Stratification. *Genetic Epidemiology.* 2012;36(7):753.
57. Martin ER, Ritchie MD, Hahn L, et al. A novel method to identify gene-gene effects in nuclear families: the MDR-PDT. *Genet Epidemiol.* 2016;30(2):111–123.
58. Lou XY, Chen GB, Yan L, et al. A combinatorial approach to detecting gene-gene and gene-environment interactions in family studies. *Am J Hum Genet.* 2008;83(4):457–467.
59. Chen GB, Liu N, Klimentidis YC, et al. A unified GMDR method for detecting gene-gene interactions in family and unrelated samples with application to nicotine dependence. *Hum Genet.* 2014;133(2):139–150.