



Novel approaches to the analysis of family data in genetic epidemiology

Nathan Morris^{1,2}, Robert C. Elston^{1,3}, Jill S. Barnholtz-Sloan^{1,3} and Xiangqing Sun^{1*}

¹ Department of Epidemiology and Biostatistics, Case Western Reserve University, OH, USA

² Center for Clinical Investigation, Case Western Reserve University, OH, USA

³ Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, OH, USA

*Correspondence: x.sun@case.edu

Edited and reviewed by:

Anthony Gean Comuzzie, Texas Biomedical Research Institute, USA

Keywords: genome-wide association, family studies, study designs, genetic factors, environmental factors

THE IMPORTANCE OF FAMILY DATA

The study of Genetic Epidemiology has historically focused on the inheritance of genetic factors and phenotypes within families. In fact, much of genetics involves the study of patterns of familial resemblance and identifying the factors that explain the observed patterns. However, in recent years the most common study design for investigating the genetic determinants of diseases has become that of genome wide association studies (GWAS) utilizing samples of *unrelated* individuals. The popularity of this approach has been driven primarily by a flood of ever improving technologies. Unfortunately, while GWAS using unrelated individuals have revealed a great many interesting disease associated variants, these variants are typically of small effect and cannot explain the observed patterns of heritability for many traits. In contrast there are numerous examples of highly penetrant rare segregating alleles that have been discovered using family based approaches. Furthermore, family based approaches have other advantages: the ability to overcome confounding factors such as population stratification, and the numerous studies that have collected large amounts of family data and which should continue to be leveraged. Unfortunately, family based approaches to genetics have an added layer of complexity at all stages from design to analysis.

This editorial introduces the Frontiers in Genetics Research Topic and Ebook: “Novel approaches to the analysis of family data in genetic epidemiology.” The papers in this issue reveal that, even with easy access to high-throughput genotyping tools such as SNP arrays and next generation sequencing, family based study designs still play an important role in untangling the complex web of environmental and genetic factors that lead to disease.

FAMILY BASED STUDY DESIGNS

A number of articles in this issue shed light on unique study designs and approaches to analyzing family data. Stein et al. (2013) describe a household contact study design which involves collecting data on households that may include both related and unrelated individuals. They argue that this research study design may be a powerful approach for jointly studying genetic and environmental exposures. Similarly, Estus et al. (2013) describe an approach to combining family based and population based data by utilizing a combined association test. Wang et al. (2013)

describe an approach of using only the independent probands from a family based study of autism to investigate genetic factors that account for IQ differences in autism patients. Nelson et al. (2013) describe a unique population based registry in Utah that contains pedigree information for all residents of the state and dates back many decades. Using this information they show that certain subsets of prostate cancer, such as early onset, high BMI, and lethal prostate cancer, cluster in families more strongly than other forms of prostate cancer. They further suggest that future studies should focus on families that display a clear clustering of a more carefully defined cancer phenotype to reduce the signal to noise ratio. Uemoto et al. (2013) discuss the power of regional heritability mapping with a mixed model approach applicable to both related and unrelated persons. This approach leverages the fact that even distantly related individuals share small regions of the genome that are inherited from a common ancestor.

ANALYSIS OF FAMILY DATA

The analysis of family data is generally more complex than the analysis of unrelated samples, and, thus, specialized statistical methods and software are often needed. Huang et al. (2013) propose a novel method of linkage analysis using sequence data on large pedigrees. This method, which uniquely combines MCMC based approximations with non-stochastic approaches, can be used to map disease genes using linkage and/or association evidence. Song and Elston (2013a) investigate the distributional properties of a commonly used linkage analysis statistic. These authors also describe a new web based software package which, among other things, plots pedigrees, calculates genetic similarity coefficients and performs visualization of the relatedness among family members (Song and Elston, 2013b). Similarly, Lutz et al. (2013) describe a method of using data from family based studies to test for a direct genetic effect, an extension of a method previously used for analysis of unrelated individuals. Additionally, Lutz et al. (2014) describe an approach to look at secondary phenotypes in case-control genetic association studies that circumvents the computational issues of a former approach.

CONCLUSION

Although GWAS with unrelated samples have become one of the most common study designs currently used in human genetics,

utilizing a family based design has many advantages. If a variant can be observed to co-segregate with a phenotype within a family, the evidence for its association with the disease is greatly strengthened. Family data provide excellent opportunities to find highly penetrant rare variants, and thus discover important biology informing us about disease. The articles in this issue illustrate how family based genetic designs remain a foundational part of human genetics.

REFERENCES

- Estus, J. L., Family Investigation of Nephropathy and Diabetes Research Group, and Fardo, D. W. (2013). Combining genetic association study designs: a GWAS case study. *Front. Genet.* 4:186. doi: 10.3389/fgene.2013.00186
- Huang, Y., Thomas, A., and Vieland, V. J. (2013). Employing MCMC under the PPL framework to analyze sequence data in large pedigrees. *Front. Genet.* 4:59. doi: 10.3389/fgene.2013.00059
- Lutz, S. M., Hokanson, J. E., and Lange, C. (2014). An alternative hypothesis testing strategy for secondary phenotype data in case-control genetic association studies. *Front. Genet.* 5:188. doi: 10.3389/fgene.2014.00188
- Lutz, S. M., Vansteelandt, S., and Lange, C. (2013). Testing for direct genetic effects using a screening step in family-based association studies. *Front. Genet.* 4:243. doi: 10.3389/fgene.2013.00243
- Nelson, Q., Agarwal, N., Stephenson, R., and Cannon-Albright, L. A. (2013). A population-based analysis of clustering identifies a strong genetic contribution to lethal prostate cancer. *Front. Genet.* 4:152. doi: 10.3389/fgene.2013.00152
- Song, Y. E., and Elston, R. C. (2013a). The null distribution of likelihood-ratio statistics in the conditional-logistic linkage model. *Front. Genet.* 4:244. doi: 10.3389/fgene.2013.00244
- Song, Y. E., and Elston, R. C. (2013b). PedWiz: a web-based tool for pedigree informatics. *Front. Genet.* 4:189. doi: 10.3389/fgene.2013.00189
- Stein, C. M., Hall, N. B., Malone, L. L., and Mupere, E. (2013). The household contact study design for genetic epidemiological studies of infectious diseases. *Front. Genet.* 4:61. doi: 10.3389/fgene.2013.00061
- Uemoto, Y., Pong-Wong, R., Navarro, P., Vitart, V., Hayward, C., Wilson, J. F., et al. (2013). The power of regional heritability analysis for rare and common variant detection: simulations and application to eye biometrical traits. *Front. Genet.* 4:232. doi: 10.3389/fgene.2013.00232
- Wang, H. Z., Qin, H. D., Guo, W., Samuels, J., and Shugart, Y. Y. (2013). New insights into the genetic mechanism of IQ in autism spectrum disorders. *Front. Genet.* 4:195. doi: 10.3389/fgene.2013.00195

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 18 December 2014; accepted: 19 January 2015; published online: 06 February 2015.

Citation: Morris N, Elston RC, Barnholtz-Sloan JS and Sun X (2015) Novel approaches to the analysis of family data in genetic epidemiology. *Front. Genet.* 6:27. doi: 10.3389/fgene.2015.00027

This article was submitted to *Applied Genetic Epidemiology*, a section of the journal *Frontiers in Genetics*.

Copyright © 2015 Morris, Elston, Barnholtz-Sloan and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.