Cancer Epidemiology, Biomarkers & Prevention

Null Results in Brief

Single-Nucleotide Polymorphism Data Support the General Unrelatedness of the Males in the Agricultural Health Study

John R. Jack[1], Alison A. Motsinger-Reif[1], Stella Koutros[2], Michael C. Alavanja[2], Laura E. Beane Freeman[2], and Jane A. Hoppin[3]

Abstract

Background: Farming is often a family and multigenerational business. Relatedness among farmers could bias gene–environment interaction analysis. To evaluate the potential relatedness of farmers, we used data from a nested case–control study of prostate cancer conducted in the Agricultural Health Study (AHS), a prospective study of farmers in Iowa and North Carolina.

Methods: We analyzed the genetic data for 25,009 SNPs (single-nucleotide polymorphisms) from 2,220 White participants to test for cryptic relatedness among these farmers. We used two software packages: (i) PLINK, to calculate inbreeding coefficients and identity-by-descent (IBD) statistics and (ii) EIGENSOFT, to perform a principal component analysis on the genetic data.

Results: Inbreeding coefficients estimates and IBD statistics show that the subjects are overwhelmingly unrelated, with little potential for cryptic relatedness in these data.

Conclusions: Our analysis rejects the hypothesis that individuals in the case–control study exhibit cryptic relatedness.

Impact: These findings are important for all subsequent analyses of gene–environment interactions in the AHS. *Cancer Epidemiol Biomarkers Prev; 23(10); 2192–5. ©2014 AACR.*

Background

The Agricultural Health Study (AHS; ref. 1) is a cohort study that collected longitudinal data on farmers and their spouses to understand the health risks of exposure to pesticides and other agricultural agents. From 1993 to 1997, a total of 57,310 licensed pesticide applicators enrolled in the study, representing 82% of the licensed private pesticide applicators in Iowa and North Carolina. The AHS has evaluated a number of health outcomes, including cancer, respiratory disease, neurologic outcomes, and diabetes. In addition to the questionnaire data collected for the main study, a nested case–control genetic association study was conducted to identify genetic factors that modified pesticide risks for prostate cancer. In this study, men with or without prostate cancer were selected for genotyping. Results of the gene–environment analyses have previously been reported (2–4).

These gene–environment interaction studies suggested that specific single-nucleotide polymorphisms (SNP) may interact with use of specific pesticides by these farmers to contribute to prostate cancer risk. As farming is often a family-based business that is passed from father to son, we hypothesized that some of the individuals in the case–control study could potentially be related. As relatedness among the farmers could contribute bias in evaluating gene–environment interactions, we used the SNP data to estimate the overall genomic sharing to test to see whether cryptic relatedness was present.

Materials and Methods

The SNP data were previously described in ref. (2). Briefly, genotyping was performed with the Custom InifiniumR BeadChip Assays from Illumina (iSelect) as part of an array of 26,512 SNPs. Quality control procedures, including STRUCTURE, linkage disequilibrium, and principal components analysis (PCA), identified 25,009 SNPs across 2,220 individuals for further analysis. For this analysis, we performed additional quality control steps. Markers and individuals were removed if they had low genotyping efficiency (missing data >1%). SNPs showing significant deviation from Hardy–Weinberg proportions ($P < 0.00001$ after a χ^2 test of associations) were also removed. A total of 22,706 SNPs and 1,837 total individuals passed the entire quality control scheme.

[1]Department of Statistics, Bioinformatics Research Center, Center for Human Health and the Environment, Center for Comparative Medicine and Translational Research, North Carolina State University, Raleigh, North Carolina. [2]Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, Department of Health and Human Services, National Cancer Institute, NIH, Bethesda, Maryland. [3]Department of Biological Sciences, North Carolina State University, Raleigh, North Carolina.

Corresponding Author: John R. Jack, North Carolina State University, Bioinformatics Research Center, Campus Box 7566, 2601 Stinson Drive, Raleigh, NC 27695. Phone: 919-515-3499; Fax: 919-515-7315; E-mail: jrjack@ncsu.edu

doi: 10.1158/1055-9965.EPI-14-0276

Next, we used all of the control data (1196 individuals), along with a random sampling of 10% of the cases (64 of 641 individuals), to ensure sufficient representation of the male AHS population in general, and to avoid over-representing patients with prostate cancer (5) during the kinship analysis. Hence, 1,260 total individuals were used in this analysis.

The genetic data allowed us to directly test for potential relatedness in the cohort by assessing the overall genomic sharing across the genome. We used PLINK (6), to estimate inbreeding coefficients for each individual and genome-wide identity-by-descent (IBD) estimates for all pairs of individuals. In addition, PCA via EIGENSOFT (7) was performed to visualize the overall patterns of sharing for a high number of principal components.

Results

The results of our PLINK analyses for potential cryptic relatedness are shown in Fig. 1: The inbreeding coefficients for each individual and the mean inbreeding coefficient (solid red line) are shown in Fig. 1A. The maximum inbreeding coefficient for all of the individuals in the study was 0.1433 whereas the majority (99.52%) of individuals had an inbreeding coefficient less than 0.0625, which indicates that the individuals in this study are overwhelmingly unrelated.

Using IBD estimates, we generally saw no evidence for potential cryptic relatedness. In Fig. 1B, the probabilities of IBD for all possible pairs of individuals are shown—P(IBD = 0) and P(IBD = 1) on the *y*- and *x*-axis, respectively. The vast majority of individual pairs have P(IBD = 0) close to 1.0, which indicates an unrelated population. In fact, only six individual–individual pairs scored P(IBD = 1) > 0.38 of 793,170 total pairwise combinations.

In Fig. 2, we show the first three as well as the ninth principal components to visualize the variance across the data. The first three principal components show no significant subpopulations. The ninth component is the first display of separation for the individuals with relatively higher P(IBD = 1). Cryptic relatedness is often not revealed in lower components, but is evident in higher components (8).

Conclusions

Overall, the results of our analysis indicate that there is generally no cryptic relatedness among this subset of male applicators in the AHS. The inbreeding and IBD results indicate that levels of genetic sharing are not higher than

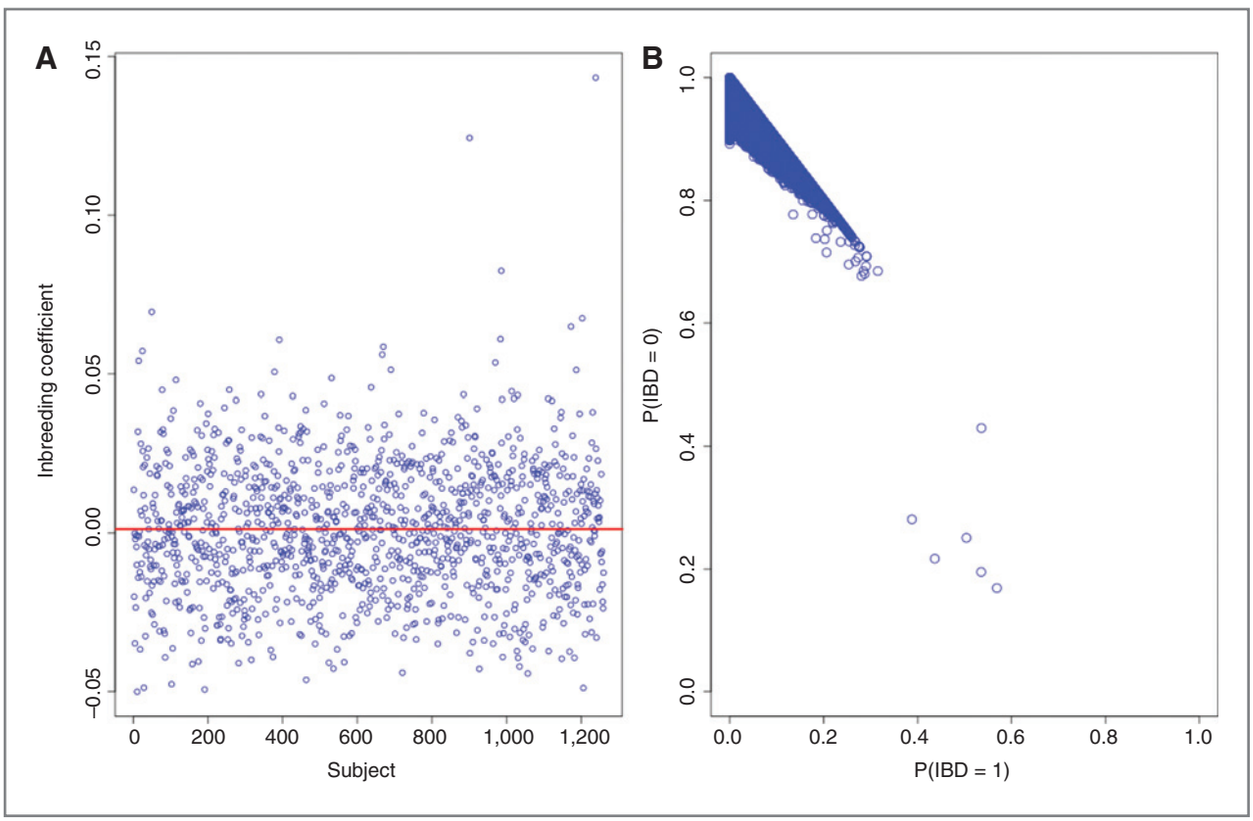

Figure 1. Inbreeding coefficient and IBD. A, a plot of the inbreeding coefficients per individual with the mean (red line) across all individuals. The data clearly represent a population with negligible inbreeding across the 1,196 controls and 64 cases (10% random sample). B, the probabilities of IBD equal to zero or one based on all possible pairs of individuals.

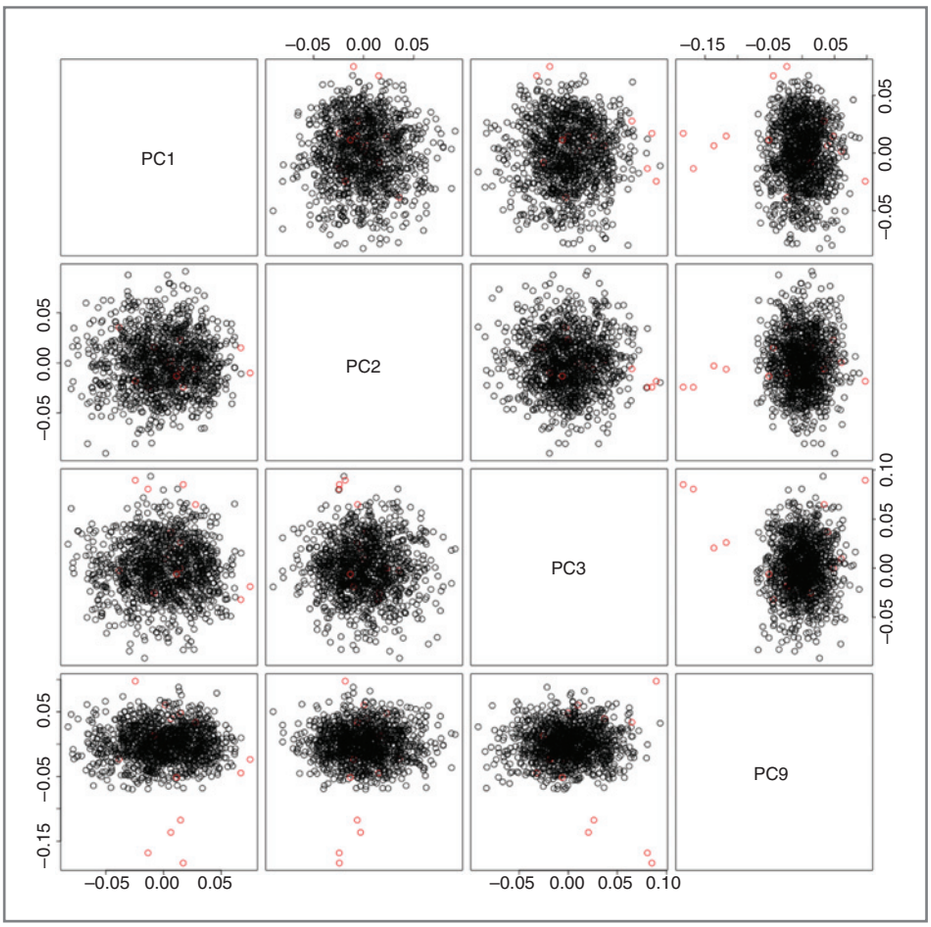

Figure 2. PCA, comparison of the first four principal components provides visualization of the population. Each point, an individual, whereas the red points indicate individuals with relatively higher inbreeding coefficients ($F > 0.1$) or P(IBD = 1) > 0.38.

expected for an unrelated cohort in an unstratified population. This is an important null result, as it justifies treating individuals in the AHS as independent observations for studies of gene–environment interactions.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: A.A. Motsinger-Reif, M.C. Alavanja, J.A. Hoppin
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): M.C. Alavanja, J.A. Hoppin
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): J.R. Jack, A.A. Motsinger-Reif, S. Koutros, J.A. Hoppin
Writing, review, and/or revision of the manuscript: J.R. Jack, A.A. Motsinger-Reif, S. Koutros, M.C. Alavanja, L.E. Beane Freeman, J.A. Hoppin
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S. Koutros, L.E. Beane Freeman, J.A. Hoppin
Study supervision: A.A. Motsinger-Reif

Grant Support

This work was supported in part by the intramural research program of the NIH, National Cancer Institute (Z01-ES049030; PIs, M.C. Alavanja and L.E. Beane Freeman), National Institute of Environmental Health Sciences (Z01-CP010119; PI, D.P. Sandler), and Pilot Project Program (PI, Alison Motsinger-Reif) of the North Carolina State University's Center for Human Health and the Environment.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 2, 2014; accepted June 4, 2014; published OnlineFirst July 21, 2014.

References

1. Alavanja MC, Sandler DP, McMaster SB, Zahm SH, McDonnell CJ, Lynch CF, et al. The Agricultural Health Study. Environ Health Perspect 1996;104:362–9.
2. Koutros S, Beane Freeman LE, Berndt SI, Andreotti G, Lubin JH, Sandler DP, et al. Pesticide exposure modifies the association between variants on chromosome 8q24 and prostate cancer. Cancer Res 2010;70:9224–33.
3. Barry KH, Koutros S, Berndt SI, Andreotti G, Hoppin JA, Sandler DP, et al. Genetic variation in base excision repair pathway genes, pesticide exposure, and prostate cancer risk. Environ Health Perspect 2011;119:1726–32.
4. Karami S, Andreotti G, Koutros S, Barry KH, Moore LE, Han SS, et al. Pesticide exposure and inherited variants in vitamin D pathway genes in relation to prostate cancer. Cancer Epidemiol Biomarkers Prev 2013;22:1557–66.
5. Howlader N, Noone AM, Krapcho M, Garshell J, Neyman N, Altekruse SF, et al. SEER Cancer Statistics Review, 1975–2010, National Cancer Institute. Bethesda, MD. Based on November 2012 SEER data

submission, posted to the SEER web site, April 2013. Available from: http://seer.cancer.gov/csr/1975_2010/.

6. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81: 559–75.
7. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006 38:904–9.
8. Thornton T, McPeek MS. Roadtrips: case–control association testing with partially or completely unknown population and pedigree structure. Am J Hum Genet 2010;86:172–84.