9-1-2014

Re: Association of polymorphism in cytochrome P450 2C9 with susceptibility to head and neck cancer and treatment outcome: Pragmatic use of Hardy-Weinberg equilibrium and statistical interaction analysis

Sandeep Kumar Singh
Environmental and Occupational Health,Robert Stempel College of Public Health and Social Work, Florida International University, sksingh@fiu.edu

Applied & Translational Genomics 3 (2014) 48–49

Contents lists available at ScienceDirect

Applied & Translational Genomics

journal homepage: www.elsevier.com/locate/atg

Letter to the Editor

Re: Association of polymorphism in cytochrome P450 2C9 with susceptibility to head and neck cancer and treatment outcome: Pragmatic use of Hardy-Weinberg equilibrium and statistical interaction analysis

To the Editors:

In a recent issue of the journal, the research article "Association of polymorphism in cytochrome P450 2C9 with susceptibility to head and neck cancer and treatment outcome", by Yadav et al. (2014) reported that cytochrome P450 2C9 single nucleotide polymorphisms (SNPs), CYP2C9*2 (rs1799853) and CYP2C9*3 (rs1057910), showed a risk association with head and neck squamous cell carcinoma (HNSCC). They also reported the statistical interaction of these polymorphisms with tobacco smoking, tobacco chewing, and alcohol consumption. In this study, the author underscored several important points, however, failed to address several limitations.

The author reported significant deviation from the Hardy-Weinberg equilibrium (HWE) for controls for both polymorphisms (Yadav et al., 2014). The Hardy-Weinberg disequilibrium (HWD) in controls is a major limitation and results should be interpreted cautiously (Salanti et al., 2005). Use of the Cochran-Armitage test of trend has been suggested if there is a significant deviation (P-value < 0.05) from HWE (Lewis and Knight, 2012). HWE elucidates that in a large population, alleles segregate randomly (Lewis and Knight, 2012; Lunetta, 2008). Allele and genotype frequencies in the population do not change until there is selection, mutation, or migration. Departure from HWE in controls may be due to several issues, including cryptic population structure (Salanti et al., 2005; Lewis and Knight, 2012; Lunetta, 2008). It has been recommended to exclude SNPs which show significant deviation from HWE in controls to avoid spurious associations. However, the criterion of exclusion depends on several factors (Lunetta, 2008).

The Indian population is very heterogeneous, with several subpopulation structures among ethnic groups existing across India (Basu et al., 2003; Indian Genomic Variation Consortium, 2008; Rich et al., 2009). The North Indian population is also not homogenous. Even pooling samples by state of origin can mask population substructure (Indian Genomic Variation Consortium, 2008). Counting them as a homogenous population will increase false-positive disease associations due to systematic ancestry difference between cases and controls (Indian Genomic Variation Consortium, 2008; Rich et al., 2009). In this study, the author did not account for population stratification as a major confounder. Population stratification can be avoided by classifying individuals into known subgroups (e.g., by ethnicity or caste) in cases and controls, to avoid confounding (Lewis and Knight, 2012; Lunetta, 2008). Most importantly, the results should be adjusted for any covariates related to population structure (Lunetta, 2008). Genomic controls can also be added to reduce confounding by alleles.

The second limitation of this study is the use of different genetic models (Fig. 1) (Lewis and Knight, 2012) for the same polymorphism. In this study, the author used the Cochran-Armitage test of trend (odds ratio per allele), as suggested in the case of HWE deviation for association analysis (Lewis and Knight, 2012), and the dominant model (allele positivity) for the interaction analysis, which may suggest data dredging (Smith, 2002). The genetic model should be same for all statistical analyses pertaining to the same polymorphism. For example, in this study, interaction analysis should be done by using the Cochran-Armitage test of trend as well. The dominant model is also not suggested in the case of HWD (Lewis and Knight, 2012). Another major issue is that data provided in the text do not match the provided tables. Odds ratios (ORs) and confidence interval (CI) in tables II and III have some ambiguity with the text.

The interaction analysis conducted in this study is actually a stratified analysis by exposure variables (exposure vs non-exposure) (Yadav et al., 2014). Stratified analysis is used to judge confounding or effect modification. Stratified analysis should also be adjusted for major confounders to avoid false-positive associations. It is not clear whether there is any statistical adjustment for major confounders. To compare the two strata (exposure vs non-exposure), an overall analysis combining the ORs for exposure and non-exposure for the dominant model is necessary for any interpretation in this study.

$$OR_{(interaction)} = OR_{GE}/(OR_G * OR_E). \quad (1)$$

In the case of effect modification, stratified analysis gives different ORs for each stratum. A statistically significant difference (P-value < 0.05) between these two ORs suggests statistical interaction ($OR_{interaction}$) (Table 1) (Liu et al., 2004; Clayton and McKeigue, 2001) and can easily be done by adding an interaction term in a logistic regression model (Kraft et al., 2007). Here G denotes genotype, E denotes environmental variable and G × E denotes interaction term. Kraft et al. (2007) have elucidated a detailed explanation of gene-environment interaction in genetic association studies. It is not pragmatic to claim an interaction without proper statistical interaction analysis.

$$Y = \beta_0 + \beta_1 G + \beta_2 E + \beta_3 G \times E. \quad (2)$$

In any genetic association study, cautious interpretation is pivotal to avoid spurious associations. Any methodological error during the study design or statistical analysis may introduce bias which may reduce the validity of the study (Salanti et al., 2005; Price et al., 2006), and of course replication failure. HWD raises several questions for additional thinking about a study. Statistical adjustment for subpopulation and genomic controls are not a panacea, but play a critical role to reduce type I error.

References

Basu, A., et al., 2003. Ethnic India: a genomic view, with special reference to peopling and structure. Genome Res. 13, 2277–2290.

http://dx.doi.org/10.1016/j.atg.2014.05.002

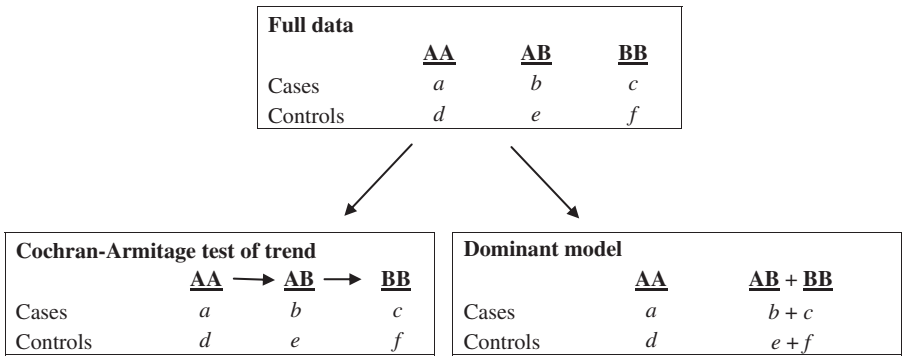

Fig. 1. Analysis methods in a case-control genetic association study used in Yadav et al. Cochran-Armitage test was used for association analysis and dominant model was used for interaction analysis.

Table 1
Complete frequency data for cases and controls for interaction analysis.

Disease Status	Genotype present		Genotype absent	
	Exposed	Unexposed	Exposed	Unexposed
Case	*a*	*b*	*e*	*f*
Control	*c*	*d*	*g*	*h*
Odds ratio	OR_{GE} = ah/cf	OR_G = bh/df	OR_E = eh/gf	OR_{00} = 1 (ref)

Clayton, D., McKeigue, P.M., 2001. Epidemiological methods for studying genes and environmental factors in complex diseases. Lancet 358, 1356–1360.
Indian Genomic Variation Consortium, 2008. Genetic landscape of the people of India: a canvas for disease gene exploration. J. Genet. 87 (3–20).
Kraft, P., et al., 2007. Exploiting gene-environment interaction to detect genetic associations. Hum. Hered. 63, 111–119.
Lewis, C.M., Knight, J., 2012. Introduction to genetic association studies. Cold Spring Harb. Protoc. 2012 (3), 297–306.
Liu, X., et al., 2004. Genetic dissection methods: designs used for tests of gene-environment interaction. Curr. Opin. Genet. Dev. 14, 241–245.
Lunetta, K.L., 2008. Genetic association studies. Circulation 118, 96–101.
Price, A.L., et al., 2006. Principal component analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38 (8), 904–909.
Rich, D., et al., 2009. Reconstructing Indian population history. Nature 461, 489–494.
Salanti, G., et al., 2005. Hardy-Weinberg equilibrium in genetic association studies: an empirical evaluation of reporting, deviations and power. Eur. J. Hum. Genet. 13, 840–848.
Smith, G.D., 2002. Data dredging, bias, or confounding. BMJ 325 (7378), 1437–1438.
Yadav, S.S., et al., 2014. Association of polymorphism in cytochrome P450 2C9 with susceptibility to head and neck cancer and treatment outcome. App. Tran. Gen. 3, 8–13.

Sandeep Kumar Singh
Environmental and Occupational Health,
Robert Stempel College of Public Health and Social Work,
Florida International University, Miami, FL, United States