Hum Genet (2002) 111 : 573–574
DOI 10.1007/s00439-002-0819-y

FOOD FOR THOUGHT

Jianfeng Xu · Aubrey Turner · Joy Little
Eugene R. Bleecker · Deborah A. Meyers

Positive results in association studies are associated with departure from Hardy-Weinberg equilibrium: hint for genotyping error?

Received: 7 June 2002 / Accepted: 29 July 2002 / Published online: 13 September 2002

Case-control studies have been widely used to test for association between DNA sequence variants and complex diseases. The premise of genetic association studies is that the increased allele or genotype frequencies in cases compared with controls implicates sequence variants that either increase risk to a disease or are in strong linkage disequilibrium (LD) with a disease-causal mutation. However, many other factors can also lead to an observed difference in allele or genotype frequencies between cases and controls. While much attention has been devoted to the potential impact of incomparability between cases and controls in terms of sources of cases and controls, environmental exposures, and genetic background (population stratification), there is a clear lack of comprehension of the impact of genotyping error on the results of association studies. The accuracy and precision of genotyping becomes more critical in case-control studies of complex diseases because: (1) the effect of a specific risk allele under study is usually small, therefore even a low frequency of genotyping error may lead to a false positive or false negative finding; (2) no Mendelian inheritance check can be performed due to lack of family genotype data; (3) a large number of genotypes are usually generated. However, the degree of genotyping error in case-control studies remains unclear, due to the lack of direct measures of this type of error.

To indirectly assess the prevalence and magnitude of genotyping error in case-control studies, we systematically reviewed reported association studies from PUBMED and performed Hardy-Weinberg equilibrium (HWE) tests in control subjects for each reported single nucleotide polymorphism (SNP). A significant difference between the observed and expected genotype frequencies under HWE may indicate genotyping error, because the conditions of HWE are generally applicable to the control subjects in any well-designed study population, i.e. (1) mating takes place at random with respect to genotype, (2) allelic frequencies are the same in males and females, and (3) mutation, selection, and migration are negligible. Although exceptions to the conditions of HWE may explain deviation, it is critical that investigators recognize the need to perform a test of HWE, and then evaluate the reason(s) for any observed deviation.

We searched for articles with the keywords "association genotyp* genetic case control", and limited the search to articles in English, with human subjects, and 2000 publication date. The search yielded 157 articles, which we then limited to the 101 articles available among 1,721 journals received at the Wake Forest Baptist Medical Center Coy C. Carpenter Library. We disregarded 26 articles that did not include SNPs. This search and selection scheme resulted in a total of 75 articles describing 133 SNPs (the list of these articles is at www.wfubmc.edu/genomics). It is worth noting that limiting journals to our local library could bias toward higher profile journals and our findings may not be representative of all reports. A goodness-of-fit χ^2 test was used to test for HWE by comparing the observed number of subjects for each genotype with the expected number of subjects assuming HWE. The proportion of SNPs that deviated from HWE was then calculated.

Two major findings were observed. First, 12% of the SNPs (n=16) were found to be inconsistent with HWE in control subjects, and the ranges of these P values were from 10^{-30} to 0.049. This rate is significantly higher than the expected 5% type I error (χ^2=4.22, P=0.04). Eleven of the 16 SNPs that departed from HWE were from different studies, while the remaining five SNPs were in two different studies (two SNPs in one gene and three SNPs in another gene). It is surprising that a HWE test was not even mentioned in the articles describing nine of these 16 SNPs. To make things even worse, five of the remaining seven SNPs that did mention HWE tests were incorrectly reported to be consistent with HWE. Interestingly, our test for relationship between the logarithm of P values (HWE test) for each of the 133 SNPs and the ISI impact factor for each journal revealed a negative, but not statistically significant,

J. Xu (✉) · A. Turner · J. Little · E.R. Bleecker · D.A. Meyers
Center for Human Genomics,
Wake Forest University School of Medicine,
Winston-Salem, NC27157, USA
e-mail: jxu@wfubmc.edu,
Tel.: +1-336-7165700, Fax: +1-336-7167575

correlation (r=–0.14, P=0.11). Second, the proportion of SNPs that deviated from HWE was higher among the 53 SNPs where positive association with a trait of interest was reported (18.9%, n=10) than in the 80 SNPs where null association was reported (7.5%, n=6). The difference between the two groups was marginally significant (χ^2=3.89, P=0.048; Fisher exact test P=0.059, two-sided). The four worst P values ($<10^{-5}$) were all in papers showing positive association.

It is possible that a departure from HWE for some SNPs in control subjects is due to some unknown factors other than genotyping errors, especially when multiple SNPs that are in strong LD deviate from HWE. As deviation from HWE has been shown to inflate the chance of a false-positive association (Schaid and Jacobsen 1999), a statistical method that does not assume HWE should be used to test for association between a SNP and a disease of interest. For example, the Armitage test for trend in proportions should be used (does not assume HWE), instead of the Pearson χ^2 test (assumes HWE) (Sasieni 1997). However, it is extremely important to note that such analytical treatments should only be applied after possible genotyping error has already been seriously examined and can be practically excluded. The following example from one of the articles sampled in our study clearly exemplifies this point. The numbers of "11", "12", and "22" genotypes for a SNP were 37, 101, and 48 in cases, and 44, 127, and 29 in controls. The frequency of "22" was reported to be significantly higher in cases (25.8%) than in controls (14.5%). However, the authors did not perform a HWE test for this SNP. We performed HWE tests in cases and controls, and found that the SNP was consistent with HWE in cases (χ^2=1.5, P=0.22), but not in controls (χ^2=15.37, P=0.00009). The number of "22" genotype in controls was deficient and would be closer to that of cases (21.5%) if it was in HWE. Clearly, the cause of departure from HWE for this SNP should be examined before considering alternative analytical methods.

The generality of our findings could be limited because this study was based on a small sample. However, these results effectively demonstrate a widespread problem of under appreciation for the HWE test and for genotyping quality among population-based association studies. The outcome of such a problem could be serious, as it can lead to either false positive or false negative findings for association. So then the critical questions become, "what are the potential sources for genotyping error and how could genotyping error vary between cases and controls?" Obviously, genotyping assays are susceptible to DNA contamination from plates, tubes, primers, and other environmental components of a laboratory, especially for the most sensitive new technologies. However, other types of systematic errors may be particularly troubling because they affect an entire research process and the degree of the problem could be different between cases and controls. For example, it is not uncommon that DNA plates are grouped separately for cases and controls, or that data are coded by a structured numbering system that allows for convenient identification of cases and controls. Although these approaches provide for identification of sample sources, members of the research team ranging from lab technicians to statisticians are not blinded to case-control status. This may lead to bias during the important steps of genotyping and scoring the alleles, particularly for ambiguous allele calls. These un-blinded study designs may be further exacerbated when the inherent failure rates of SNP genotyping are different for homozygous or heterozygous genotypes, leading to skewed gene frequencies. Any one or a combination of these potential problems may lead to artificially different genotype frequencies between cases and controls. Although sporadic errors may occur when even the most cautious laboratory practices are observed, the influence of systematic flaws may be an under-appreciated contributor to erroneous study findings. However, there are a number of ways to control for these problems.

Therefore, we would like to suggest the following practices for SNP genotyping: (1) blind the researchers to case-control status (which could be achieved by including cases and controls on each plate and an unstructured sample numbering system); (2) include blanks in each plate, in different well positions; (3) include multiple and duplicate control subjects in each plate in different well positions; (4) check scoring of alleles (for example: if done manually, double score each genotype); (5) determine an acceptable amount of missing data and rerun assays if there is more missing data in either cases or controls; (6) perform an HWE test for each SNP before testing any hypothesis.

Genetic association studies of complex diseases have proven to be daunting. While there is no doubt that association studies are one of the useful approaches to understand the etiology of complex diseases, attention should be paid to multiple aspects of these studies, including the quality of genotyping. We believe quality genotyping decreases the potential for false findings (positive or negative) and increases our ability to identify small but real associations between SNPs and complex diseases.

Acknowledgement We thank Drs. Siqun Lilly Zheng and Timothy D. Howard for discussion on the quality of SNP genotyping.

References

Sasieni PD (1997) From genotypes to genes: doubling the sample size. Biometrics 53:1253–1261

Schaid DJ, Jacobsen SJ. (1999) Biased tests of association: comparisons of allele frequencies when departing from Hardy-Weinberg proportions. Am J Epidemiol 149:706–711