**Hypothesis/Commentary**

**Family-Based Samples Can Play an Important Role in Genetic Association Studies**

Ethan M. Lange,1,2 Jielin Sun,3,4 Leslie A. Lange,1 S. Lilly Zheng,3,4 David Duggan,5 John D. Carpten,5 Henrik Gronberg,6 William B. Isaacs,7 Jianfeng Xu,3,4 and Bao-Li Chang3,4

Departments of Genetics and Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; Center for Cancer Genomics and Center for Human Genomics, Wake Forest University School of Medicine, Winston-Salem, North Carolina; Translational Genomics Research Institute (TGen), Phoenix, Arizona; Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; and Johns Hopkins Medical Institutions, Baltimore, Maryland

**Abstract**

Over the past 2 decades, DNA samples from thousands of families have been collected and genotyped for linkage studies of common complex diseases, such as type 2 diabetes, asthma, and prostate cancer. Unfortunately, little success has been achieved in identifying genetic susceptibility risk factors through these considerable efforts. However, significant success in identifying common disease risk-associated variants has been recently achieved from genome-wide association studies using unrelated case-control samples. These genome-wide association studies are typically done using population-based cases and controls that are ascertained irrespective of their family history for the disease of interest. Few genetic association studies have taken full advantage of the considerable resources that are available from the linkage-based family collections despite evidence showing cases that have a positive family history of disease are more likely to carry common genetic variants associated with disease susceptibility. Herein, we argue that population stratification is still a concern in case-control genetic association studies, despite the development of analytic methods designed to account for this source of confounding, for a subset of single nucleotide polymorphisms in the genome, most notably those single nucleotide polymorphisms in regions involved with natural selection. We note that current analytic approaches designed to address the issue of population stratification in case-control studies cannot definitively distinguish between true and false associations, and we argue that family-based samples can still serve an invaluable role in following up findings from case-control studies. (Cancer Epidemiol Biomarkers Prev 2008;17(9):2208–14)

**Introduction**

Recent successes in identifying disease risk-associated variants using genome-wide association (GWA) studies have shown the power of this approach in discovering novel risk factors. Interestingly, many of these genetic risk variants were identified in study populations consisting of patients with specific diseases and a common set of controls from the general population (1), rather than in well-matched case-control populations where cases and controls are ascertained from the same well-defined populations. The fact that many of these associations can be consistently replicated in other study populations seems to relieve concern about population stratification, a phenomenon in genetic association studies where differences in allele or genotype frequencies at a particular polymorphism between samples of cases and controls are due to differences in ethnic origins between the two samples, rather than a real effect of the variation at the polymorphism on disease risk. However, by carefully examining genetic association findings of several recent studies, we found that population stratification may manifest in different forms and still be a concern in genetic association studies for a subset of single nucleotide polymorphisms (SNP) in the genome. We believe that proper recognition of the potential problem, use of appropriate study designs and analytic tools to address this potential confounding, and cautious interpretation of association results are still critical issues in genetic association studies. We argue that family-based association studies, which take advantage of previously collected family samples, can play an important role in assessing the relationship between genetic variants and disease and may be particularly valuable for following up results from case-control studies.
Large Geographic Variation in Allele Frequencies for a Subset of SNPs in the Genome

Considerably heterogeneous allele frequencies are observed among people from different geographic regions for a subset of SNPs in the genome. As shown in the Wellcome Trust Case Control Consortium GWA study where allele frequencies of SNPs across the genome were compared among ~17,000 subjects from 12 broad geographic regions across the United Kingdom (1), SNPs in several genomic regions had highly significant differences in allele frequencies (Fig. 1A), including several regions not previously implicated by past studies. Some of the differences reached genome-wide significance levels ($P < 10^{-8}$), for example, SNPs at 2q21.3, 4p14, and 6p21. Spurious statistical differences in allele frequencies between cases and controls can occur for SNPs at these regions if they are not well matched in terms of ancestry. Although the effect of ancestral variation on genetic association studies can be minimized with appropriate study designs and analytic adjustment, it may be difficult to remove this confounder in all situations, such as when there are strong differences in allele frequencies among SNPs in genomic regions under selective pressure between geographically neighboring populations or when the same factors play a role in contributing to both geographic variation in allele frequencies and disease risk.

For example, allele frequencies of multiple SNPs at 4p14 near the TLR6-1-10 gene cluster were found to be significantly different among control subjects from the 12 broad geographic regions across the United Kingdom in the Wellcome Trust Case Control Consortium study; $P$ values of $<10^{-40}$ were found when using 11-degree-of-freedom tests (1). It is hypothesized that these differences may reflect the outcome of natural selection due to the roles of these genes in preventing a variety of infectious diseases (2). In the meantime, sequence variants in the TLR6-1-10 gene cluster have been reported to be associated with prostate cancer risk in a well-matched case-control study population from Sweden by our group (3). Several SNPs were significantly associated with prostate cancer risk in a population-based case-control study from Sweden that included 2,887 cases and 1,715 controls. For example, we found that a SNP at the 5’-untranslated region of the TLR1 gene (rs5743551) was significantly associated with prostate cancer risk; the minor allele frequency was 0.254 in cases and 0.221 in controls ($P = 0.0009$; Fig. 1B). Further examination of this SNP revealed that the significant association was primarily driven by subjects from the central part of Sweden ($P = 0.002$). The association was in the same direction but was not statistically significant among subjects from the northern part of Sweden. The difference in strength of association for this SNP in these two regions could be partially explained by the large observed geographic variation in minor allele frequency. The difference in the minor allele frequency observed in these two regional populations was significantly different both before and after accounting for prostate cancer ($P < 0.001$ and $P = 0.01$, respectively). Although the SNP remained significant after adjustment for geographic region ($P = 0.004$), we cannot completely remove the potential confounding effect of geographic variation because there may be additional hidden population structure within these regions. Considering that TLR6-1-10 sequence variants may play a role in both prostate cancer risk (4, 5) and natural selection (2), it is difficult to dissect true prostate cancer association from confounding using frequency tests, even in this relatively homogeneous and well-matched Swedish study population.

Figure 1. Geographic variation of allele frequencies in the genome. **A**, copy of Fig. 2A in a published article (1) showing $P$ values for the 11-degree-of-freedom test for differences in SNP allele frequencies between 12 broad geographic regions in the United Kingdom. Green dots indicate SNPs with a $P$ value of $<1 \times 10^{-8}$. Cluster of highly significant SNPs is found in several genomic regions, including 4p14 that contains TLR6-1-10 gene cluster. **B**, association test between prostate cancer risk and a SNP at 5’-untranslated region of TLR1 gene (rs5743551) in the Cancer of Prostate in Sweden study population. Statistical association was found in Cancer of Prostate in Sweden ($P = 0.0009$). However, a significant difference in allele frequencies between Region 1 (northern part of Sweden) and Region 2 (central part of Sweden) was also found, although the minor allele was always higher in cases than in controls in both regions.
Variable Ancestry Proportions in Cases and Controls within a Race Group

It is well known that false-positive associations may occur in genetic association studies if samples of cases and controls are from different populations and if the rate of disease is different in the two populations. The problem likely also exists when cases and controls are ostensibly from the same race and ethnic group but have different proportions of subrace or subethnic ancestry. For example, African-Americans from different parts of the United States have various levels of west African, east African, and European ancestry, whereas European Americans in the United States have various levels of southern or northern European ancestry (6). The magnitude of this potential problem depends on the complexity and degree of admixture among subgroups, differences in disease risk and prevalence between the subgroups, and differences in allele frequencies for SNPs in the genome (global) and in a specific region of genome (local) between these subgroups. Recent results of association studies of multiple independent sequence variants at 8q24 and prostate cancer risk exemplify the complexity of this potential problem.

Association of prostate cancer risk with 8q24 variants was initially discovered in a fine mapping study of a prostate cancer linkage region at 8q24 among two case-control study populations from Iceland (7). It includes a SNP, rs1447295, and multiple other variants that are in strong linkage disequilibrium with this SNP. This locus was independently confirmed in all published study populations, including those from Sweden and the United States (7), Multiethnic Cohort (8), Australia (9), Mayo Clinics (10), Breast and Prostate Cancer Cohort Consortium (11), and King County of Washington (12). This locus was also implicated as one of the strongest prostate cancer associated regions in three GWA studies (13-16). Interestingly, in addition to the locus rs1447295 at 128,554,220 bp (referred to as Region 1 of 8q24), two additional loci that are within 500 kb proximal to Region 1 were found to be independently associated with prostate cancer risk (13, 14, 17). One of these two regions includes the SNP rs6983267 at 128,510,352 bp (referred to as Region 3 of 8q24) and the other includes a SNP rs16901979 at 128,194,098 bp (referred to as Region 2 of 8q24). Independent support for association of these three regions at 8q24 with prostate cancer risk was also observed in a hospital-based case-control study from the Johns Hopkins Hospital (18). Although these consistent findings strongly suggest the presence of true disease causal variants in this region, several interesting observations suggest that we need to take residual population stratification into consideration to appropriately interpret the observed associations at multiple independent regions of 8q24.

One interesting observation is that all of the “risk alleles” for SNPs at the three 8q24 regions have higher estimated frequencies in African study populations; allele A of rs1447295 at Region 1, allele T of rs6983267 at Region 3, and allele A of rs16901979 at Region 2 were all found to be more common in prostate cancer cases than in controls of European Americans (13-18). All of these risk alleles have considerably higher frequencies in the African population (YRI) than the European population (CEU); their frequencies (YRI/CEU) for these three SNPs in Regions 1, 3, and 2 are 0.54/0.02, 0.98/0.46, and 0.34/0.07, respectively. The implication of these large differences in allele frequencies between YRI and CEU samples is that if prostate cancer cases in general have a higher proportion of African ancestry than controls, these SNPs would be particularly susceptible to be erroneously statistically associated with prostate cancer risk due to population stratification given the increased risk for prostate cancer in men of African descent.

Data from an admixture mapping study in African-American prostate cancer cases and controls do indicate that African-American prostate cancer cases have a
higher proportion of African ancestry than controls. As shown in Fig. 2, the proportion of African ancestry was estimated at each region of the genome among African-American prostate cancer cases and controls in the United States (8, 17). In the samples considered, African-American prostate cancer cases have a higher proportion of African descent across the entire genome than do African-American controls. The mean proportion of African ancestry in the genome (global) was significantly higher in these cases (78.2%) than in controls (74.8%; \( P < 0.001 \)). The difference was largest (>7%) at a local region of 8q24 (126-129 Mb). These results, if confirmed, may have several important implications.

The systematic difference in estimated proportion of African ancestry suggests that these African-American samples are not ideally matched to control the effects of population stratification. The higher African ancestry proportion across the entire genome in cases than in controls suggests that many SNPs may have different allele frequencies between cases and controls due to population stratification. Correcting for this consistently observed difference in proportion of African ancestry across the genome using appropriate analytic techniques on available genetic marker data is necessary to maintain unbiased association tests given the systematic differences in allele frequencies between European and African populations.

Another observation is that there is considerable variation in allele frequencies for the 8q24 risk variants among European subjects. As shown in Table 1, the allele frequencies for rs1447295 at Region 1 differed considerably among different European study populations. The frequency of the risk allele A ranges from 0.07 in a French population (southern Europe) to 0.17 in a Finnish population (northern Europe). However, it is striking to note that prostate cancer cases have a consistently higher frequency than controls within each of these study populations. One of the most favorable interpretations of these results is that the consistent association finding is due to true prostate cancer risk loci at 8q24. Alternatively, the observed association could also be due to systematic population stratification. If prostate cancer cases in each of these study populations have higher proportions of African ancestry than the controls, then we would expect to observe consistently higher frequencies of these risk alleles in cases than in controls. Unfortunately, there is a lack of data at this time for estimating African proportions among European prostate cancer cases and controls.

### Analytic Adjustment for Case-Control Association Tests

We have presented two scenarios where population stratification may confound results and compromise the interpretation of genetic association studies. As our knowledge about genetic variations among different human populations improves, we will have a better appreciation of the various forms and extent of population stratification on genetic association studies. Analytic methods and study designs that address subtle and complex population stratification are needed to address these concerns.

Currently, several methods are available to adjust for potential population stratification. These methods include genomic control (19), structured association (20, 21), and principal components analysis (22). All three methods have been shown to reduce the effect of population stratification in genetic association studies. In addition, all three procedures require the additional genotyping of a sizable number of genetic markers, preferably not in linkage disequilibrium, which are hypothesized not to be associated with the trait under study. Genomic control corrects for stratification by adjusting association test statistics at each marker by a uniform inflation factor determined by evaluating the distribution of test statistics across all markers. However, this approach does not factor in that some markers will have greater differences in allele frequencies across different ancestral populations than others. Thus, the uniform adjustment would undercorrect for markers with large differences in allele frequencies across populations and overcorrect for markers with relatively small differences in allele frequencies across populations. Structured association is based on assigning individuals to subpopulations using a model-based Bayesian clustering algorithm implemented in the program STRUC-TURE and then carrying out all analyses conditional on the inferred assignments. Given the computational complexity of this approach, structured association is limited to a relatively small number of ancestral informative markers—markers that have large allele frequencies across measured populations. The primary concerns about this approach are the reliance on a relatively small number of markers to differentiate between populations, the adequacy of a modest number of markers to differentiate between populations that have not been previously extensively studied, and the sensitivity of the approach to the assumed number of clusters (which is somewhat arbitrarily chosen by the user). The use of principal components, a data reduction method

### Table 1. Reported frequency of allele A at rs1447295 in men of European ancestry

<table>
<thead>
<tr>
<th>Study (reference)</th>
<th>No. subjects</th>
<th>Frequency</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases Controls</td>
<td>Frequency</td>
<td>Note</td>
</tr>
<tr>
<td>deCODE (7)</td>
<td>1,291 997</td>
<td>0.17 0.11</td>
<td>Iceland</td>
</tr>
<tr>
<td>CAPS (7)</td>
<td>1,435 779</td>
<td>0.16 0.13</td>
<td>Sweden</td>
</tr>
<tr>
<td>Chicago (7)</td>
<td>458 247</td>
<td>0.13 0.08</td>
<td>United States</td>
</tr>
<tr>
<td>Australia (9)</td>
<td>821 732</td>
<td>0.15 0.11</td>
<td>Australia</td>
</tr>
<tr>
<td>Mayo (10)</td>
<td>435 545</td>
<td>0.12 0.10</td>
<td>United States</td>
</tr>
<tr>
<td>ACS (11)</td>
<td>1,150 1,151</td>
<td>0.12 0.08</td>
<td>United States</td>
</tr>
<tr>
<td>ATBC (11, 13)</td>
<td>894 896</td>
<td>0.21 0.17</td>
<td>Finland</td>
</tr>
<tr>
<td>EPIC (11, 13)</td>
<td>732 1,114</td>
<td>0.13 0.12</td>
<td>Europe</td>
</tr>
<tr>
<td>HPFS (11, 13)</td>
<td>625 636</td>
<td>0.13 0.09</td>
<td>United States</td>
</tr>
<tr>
<td>MEC (11, 17)</td>
<td>1,168 938</td>
<td>0.13 0.09</td>
<td>United States</td>
</tr>
<tr>
<td>PHS (11)</td>
<td>969 1,264</td>
<td>0.12 0.09</td>
<td>United States</td>
</tr>
<tr>
<td>PLCO (11, 13)</td>
<td>1,172 1,157</td>
<td>0.14 0.10</td>
<td>United States</td>
</tr>
<tr>
<td>Washington (12)</td>
<td>630 564</td>
<td>0.14 0.11</td>
<td>United States</td>
</tr>
<tr>
<td>FPC (13)</td>
<td>455 489</td>
<td>0.12 0.07</td>
<td>France</td>
</tr>
<tr>
<td>Baltimore (18)</td>
<td>1,545 576</td>
<td>0.13 0.08</td>
<td>United States</td>
</tr>
</tbody>
</table>

Abbreviations: deCODE, deCODE Genetics; CAPS, Cancer of Prostate in Sweden; ACS, American Cancer Society; ATBC, Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study; EPIC, European Prospective Investigation into Cancer and Nutrition Cohort; HPFS, Health Professionals Follow-up Study; MEC, Multiethnic Cohort Study; PHS, Physicians’ Health Study; PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; FPC, French Prostate Case-Control Study.
-designed to capture most of the variability across all SNPs using a relatively small number of independent continuous variables, allows the inclusion of a genome-wide panel of markers to measure and account for systematic differences in ancestry and is becoming an increasingly popular analytic approach for identifying population substructure and for controlling for it in genetic association studies. This method has the benefit of being computationally efficient and allows the inclusion of hundreds of thousands of genetic markers, making the a priori choice of ancestry informative markers in GWA studies unnecessary.

Critical to the validity of case-control genetic association study results is the ability of analytic methods to account for population stratification, particularly for SNPs in regions of selection. Despite the recognition of this problem, empirical and simulation-based studies designed to evaluate the effect of population stratification, after application of methods designed to control for it, have been somewhat limited. It should be noted that all of the described analytic approaches failed to resolve the observed population stratification in a case-control sample of Americans of European origin that were selected for extreme values of height (6, 22, 23). For example, in the study by Campbell and colleagues (6), height and allele frequencies at the lactase gene \( \text{LCT} \) (MIM 603202) varied considerably from northwestern to southeastern Europe and a false-positive association was found linking the lactase gene with height in this sample. Results across several studies suggest that analytic methods designed to address confounding will dramatically reduce the number of false positives due to population stratification but will not completely eliminate the problem (6). A recent European GWA study on rheumatoid arthritis found evidence for spurious association at \( \text{LCT} \) and \( \text{IRF4} \), two regions that are highly selected in European populations (24). The application of both principal components and STRUCTURE, in this study, reduced the confounding at these two regions but did not remove the effects entirely. Regardless, if the causal polymorphisms are highly correlated with underlying population structure (e.g., because of natural selection), it will not be possible to distinguish between true and false positives statistically, and any attempt to remove the population structure will reduce the power to detect truly associated markers.

The recent Wellcome Trust Case Control Consortium GWA study found only modest evidence of over-dispersion (\( \lambda \) ranging from 1.03 to 1.11 for the seven diseases evaluated) in their association trend test statistics, despite strong evidence for highly discrepant allele frequencies at SNPs in 13 different chromosomal regions between control samples collected from 12 different geographic regions in the United Kingdom (1). In this study, both cases and controls were ascertained throughout the United Kingdom; there was little evidence for confounding due to population stratification when comparing case allele frequencies with that of controls even before incorporating analytic methods to control for this effect. On the other hand, the observed evidence for numerous genomic regions under apparent selective pressure in the Wellcome Trust Case Control Consortium study could cause some concern about the validity of case-control genetic association study results for a subset of SNPs in the genome, especially when cases and controls are not ascertained from the same geographic region.

Family-Based Samples and Family-Based Genetic Association Studies

For many complex diseases, a large number of pedigrees with multiple affected relatives have been recruited in the past for genetic linkage studies. Although the ability of genetic linkage studies to identify regions that contain susceptibility variants for complex diseases may be limited due to heterogeneity, reduced penetrance, and phenocopies (25), these families may still be extremely useful for association testing. Of note, DNA samples from thousands of families with hereditary prostate cancer have already been collected for linkage studies. These families have extensive clinical information with respect to prostate and other cancers in the probands and other family members that may prove invaluable when trying to understand the complex genetic etiology of the disease and offer the unique opportunity to evaluate parent-of-origin effects, or imprinting, that cannot be evaluated in studies of unrelated cases and controls.

Typical genetic case-control association studies use unrelated population-based cases and controls. Several theoretical and simulation studies (26-29) have shown that greater power can be achieved in genetic case-control association studies by selecting independent cases from families with multiple cases over selecting cases with no family history of the disease. Empirical results from prostate cancer genetic association studies support these conclusions. Specifically, we observed stronger effects of \( 8q24 \) variants in our prostate cancer cases from high-density hereditary prostate cancer families than in unrelated prostate cancer cases collected independent of family history (30). A recently completed GWA study for prostate cancer using cases diagnosed before age 61 or with a family history of disease was tremendously successful not only in strongly replicating the results from previous GWA studies (i.e., variants associated with prostate cancer on \( 8q \) and \( 17q \)) but also in identifying seven new regions that harbor genetic variants significantly associated with the disease (16). This study shows the considerable power that can be gained by studying cases, such as those collected through linkage studies, which are enriched for carrying genetic susceptibility variants.

Although considerable power can be gained using familial cases in case-control association studies, the use of familial cases in the case-control study design does not protect the conclusions from the effect of population stratification. An alternative study design to case-control designs is family-based association. Family-based association methods evaluate whether particular alleles are transmitted from parents down to affected offspring in a proportion that is different than expectation under the null hypothesis of no association between marker and disease. Because these methods use nontransmitted alleles from the same parents as the control sample, these methods are not susceptible to population stratification. The transmission disequilibrium test was originally proposed for a parent-parent–affected offspring design (31) but has been extended to include siblings.
discordant for the disease under study and general family designs (32, 33). Because many of the diseases of interest typically have a late age of onset, the consequential unavailability of parents has made these more general forms of the transmission disequilibrium test popular.

Although family-based study designs offer protection from false positives due to population stratification, they have not been the design of choice for many due to the relative inefficiency of the design. Specifically, it is widely accepted that collecting family-based samples is considerably more expensive than collecting a sample of unrelated cases and controls. In addition, the matching of transmitted with nontransmitted alleles from the same family reduces statistical power. Consequently, to get the same power as a case-control study design, a larger number of family-based samples will have to be ascertained and genotyped. This increased cost has motivated the development of statistical techniques to account for population stratification in case-control designs. However, as noted above, these statistical methods are not perfect and cannot completely remove doubt about the validity of any single association result in a case-control design. Many family-based samples have already been collected through linkage studies; thus, the additional cost of collecting these family-based samples is often not a major concern. The additional costs associated with having to genotype a significantly larger number of samples to achieve equivalent power still likely render doing a GWA study on family-based samples (even on previously collected samples) inefficient. Unique to nuclear families when both parents are available for genotyping, the cost of doing a GWA study can be mitigated by genotyping some offspring on a significantly reduced number of SNPs and imputing (with a high degree of accuracy) the remaining SNPs in the offspring by using the linkage phase on the subset of SNPs typed for all family members (34).

Family-based studies may be ideal for validating previously identified genetic risk factors. Follow-up case-control studies typically have to do extra genotyping for a large number of ancestry informative markers (which would not need to be genotyped for a family-based study) in an attempt to control for population stratification. In addition, whereas GWA studies require a strict correction for multiple testing to account for the hundreds of thousands of genetic markers analyzed, follow-up studies require a much less stringent significance threshold given the focused a priori hypotheses. Combining these factors with the fact that large family-based samples have already been collected suggests that family-based association tests can be an efficient and useful strategy for validating previously identified genetic susceptibility markers. Further studies need to be done to evaluate the relative power of family-based association analyses compared with case-control designs when using methods that correct for population stratification. In particular, it is important to evaluate the effect of poorly matched case-control samples, over correction for ancestry, and causal markers that are strongly associated with population structure. Regardless of these findings, family-based association methods are clearly useful to dissect true association from false association due to population stratification.

We did a transmission disequilibrium test for SNPs at Regions 1, 2, and 3 of 8q24 in 168 hereditary prostate cancer families of European ancestry using the FBAT program (30). As shown in Table 2, risk alleles of one SNP in Region 2 was significantly overtransmitted from parents to affected men ($P = 0.003$). However, this was not the case for risk alleles of SNPs in Regions 1 and 3, despite significant differences in allele frequencies observed for these SNPs between these same familial cases and a sample of unrelated screened controls. Results from the family-based transmission tests suggested that SNPs at Region 2 confer prostate cancer risk while creating some concern that the association between prostate cancer and SNPs in the other two regions may be due to population stratification (perhaps a consequence of higher African ancestry at 8q24 in prostate cancer cases than controls).

From a clinical standpoint, common susceptibility markers found in primarily nonfamilial cases need to be evaluated in familial cases. Individuals with a strong family history of a disease are most likely to seek genetic counseling and subsequent genotyping for genetic risk factors for that disease. Although family-based samples collected for linkage analyses are a nonrepresentative collection of cases that are more likely to carry genetic susceptibility than population-based cases, these existing family data can be useful to assess how important these risk factors are in individuals most likely to seek medical prevention. Finally, for diseases such as prostate cancer, genetic heterogeneity has made the identification of rare highly penetrant mutations through linkage analysis difficult. Evaluating the role of common variants in hereditary prostate cancer families will be critical in determining whether an accumulation of common risk variants is responsible for hereditary prostate cancer or whether there exist rare high penetrant mutations that explain the significant clustering of the disease within families.

### Conclusion

We believe that cautious interpretation and continued attention to population stratification as a potential confounder remains a critical issue in case-control genetic association studies in spite of the recent advances in analytic methods designed to address this potential problem. Family-based association study designs are

<table>
<thead>
<tr>
<th>SNP</th>
<th>Position*</th>
<th>Region</th>
<th>Allele</th>
<th>Frequency</th>
<th>No. family</th>
<th>S</th>
<th>E(S)</th>
<th>Var(S)</th>
<th>Z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs16901979</td>
<td>128,194,098</td>
<td>2</td>
<td>A</td>
<td>0.07</td>
<td>20</td>
<td>37</td>
<td>28.00</td>
<td>9.34</td>
<td>2.95</td>
<td>0.0032</td>
</tr>
<tr>
<td>rs6983267</td>
<td>128,482,487</td>
<td>3</td>
<td>G</td>
<td>0.50</td>
<td>70</td>
<td>262</td>
<td>258.86</td>
<td>47.53</td>
<td>0.46</td>
<td>0.65</td>
</tr>
<tr>
<td>rs1447295</td>
<td>128,554,220</td>
<td>1</td>
<td>A</td>
<td>0.12</td>
<td>47</td>
<td>89</td>
<td>84.00</td>
<td>25.09</td>
<td>1.00</td>
<td>0.32</td>
</tr>
</tbody>
</table>

*Position is based on National Center for Biotechnology Information Build 35.

Table 2. Results from family-based association tests
lately impervious to the effects of population stratification in terms of controlling the rate of false-positive tests. Family-based samples are readily available from linkage studies that have collected DNA samples from thousands of pedigrees that have multiple cases of disease. Despite some clear advantages, including higher rates of genetic susceptibility variants in familial cases, these family samples have been largely ignored in GWA studies. Family-based association study designs have not been routinely used because of concerns that they are not as statistically powerful or cost effective as case-control designs. These concerns argue that family-based association methods are not ideal for SNP discovery. However, family-based association studies may be vital in teasing out which variants identified in case-control studies are truly associated with disease and not an artifact of population stratification.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Acknowledgments
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.

References
Family-Based Samples Can Play an Important Role in Genetic Association Studies

Ethan M. Lange, Jielin Sun, Leslie A. Lange, et al.

Cancer Epidemiol Biomarkers Prev 2008;17:2208-2214.