The Roles of *PAX6* and *SOX2* in Myopia: Lessons from the 1958 British Birth Cohort

Claire L. Simpson,[1,2] *Pirro Hysi,*[1] *Shomi S. Bhattacharya,*[2] *Chris J. Hammond,*[3] *Andrew Webster,*[2] *Catherine S. Peckham,*[1] *Pak C. Sham,*[4] *and Jugnoo S. Rahi*[1,2]

Purpose. Myopia is a common complex trait that affects up to 60% of some populations. Its development is influenced by multiple genes and environmental factors. *PAX6* and *SOX2* are genes with fundamental roles in ocular growth and development, and they have been linked with myopia in a recent linkage study. The authors investigated the roles of *PAX6* and *SOX2* in common myopia as part of a broader association study of refractive error.

Methods. Five hundred ninety-six persons from the 1958 British Birth Cohort, a nationally representative population, were randomly selected from the outer tertiles of the refractive error (RE) distribution and were genotyped using 25 tagSNPs across *PAX6* and 3 tagSNPs across *SOX2* and their putative control regions. This experiment had 80% power to exclude either gene contributing more than 10% of the variance of refractive error.

Results. All SNPs were in Hardy-Weinberg equilibrium, and the genotyping failure rate was less than 5%. Accounting for multiple testing, no significant association ($P < 0.05$) was found between any of the SNPs or haplotypes and refractive error.

Conclusions. *PAX6* and *SOX2* are obvious candidates in RE genetic studies because of their biological roles and prior linkage studies. The present findings strongly suggest refractive error is not directly affected in this population by variants in either gene or by their known promoters/enhancers. The authors suggest that neither *PAX6* nor *SOX2* should be prioritized in the international search for genetic modifiers of refractive error. Their findings contribute to broader understanding of the pathophysiology of refractive error and highlight the critical role of replication in genetic research on complex disorders. (*Invest Ophthalmol Vis Sci.* 2007;48:4421–4425) DOI:10.1167/iovs.07-0231

From the [1]MRC Centre of Epidemiology for Child Health, Institute of Child Health, and [2]Institute of Ophthalmology, University College London, London, United Kingdom; [3]Twin Research and Genetic Epidemiology Unit, Kings College London, London, United Kingdom; and [4]Genome Research Center, The University of Hong Kong, Hong Kong, SAR, China.

Supported by Medical Research Council Grants G0301069 and G0000934; the Wellcome Trust; The Great Ormond Street Hospital Special Trustees; and the NHS Executive (for research and development).

Submitted for publication February 22, 2007; revised April 23, 2007; accepted August 13, 2007.

Disclosure: **C.L. Simpson**, None; **P. Hysi**, None; **S.S. Bhattacharya**, None; **C.J. Hammond**, None; **A. Webster**, None; **C.S. Peckham**, None; **P.C. Sham**, None; **J.S. Rahi**, None

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

Corresponding author: Jugnoo S. Rahi, MRC Centre of Epidemiology for Child Health, Institute of Child Health, University College London, London WC1N 1EH, United Kingdom; j.rahi@ich.ucl.ac.uk.

Global genetic research is focusing increasingly on complex traits or disorders, such as obesity, that are common and that contribute significantly to morbidity and mortality in proportions of the population. Such complex diseases, by definition, are influenced by a combination of genes and environmental factors, with each gene contributing a relatively small effect on disease susceptibility and pathogenesis. The common disease/common variant (CD/CV) hypothesis proposes that the genetic risk for common, complex traits results from disease loci containing one or a few common variants.[1] This hypothesis is the basis of both the HapMap project[2,3] and the concept of genomewide association studies. Until recently, genomewide association had been hampered by the lack of high-throughput genotyping platforms to analyze enough markers to attain sufficient power. Thus, association studies investigating candidate genes, selected on the basis of experimental evidence and knowledge of relevant biological pathways, have been the prevalent approach in the study of common complex diseases.

Refractive error (myopia, hyperopia, or astigmatism) is an archetypal complex quantitative trait; its development is considered to be influenced by multiple genes, interactions between genes, and interactions between genes and environmental factors. There is a particular and longstanding scientific interest in myopia—written accounts date from at least the time of Aristotle[4]—that reflects its impact on the lives of affected persons and on the associated societal costs of lost productivity and of comorbidity from vision impairment.[5] Recognizing this, the World Health Organization has identified the prevention and treatment of myopia as a priority within its current global initiative against avoidable blindness.[6]

The prevalence of myopia shows wide geographic variation. Mild to moderate (primary physiological or common) myopia affects approximately 25% of the population in Europe and North America[7–9] but only approximately 5%[9] in Africa and as many as 61% in Asia.[10] The risk for sight loss increases with increasing severity of myopia. High myopia (pathologic or syndromic), which may be a distinct entity, affects approximately 2% to 3% of most populations.[10] Recent marked increases in the prevalence and the severity of common myopia in many populations[11] have been attributed to changing environmental influences. However, there has been a renewed interest in its genetic basis fueled by twin studies,[12–14] indicating high heritability (60%–90%), and by familial aggregation studies.[15–18]

Most genetic studies to date have focused on rare forms of pathologic or syndromic myopia, which may have underlying causes different from those of common myopia. The first genomewide linkage scan to investigate common refractive error was reported in a sample of dizygotic twins drawn from a volunteer twin register.[19] Of four linked loci reported, the two of the highest LOD scores were on 11p13 (LOD 6.1) and 3q26 (LOD 3.7). *PAX6* underlies the highest point of the peak on 11p13. It plays an essential role in oculogenesis. Mutations in *PAX6* cause familial and sporadic aniridia and a number of other severe ocular phenotypes.[20] The purpose of its continued postnatal expression in the eye is unknown, but dosage of

PAX6 in transgenic mice influences eye size.[21] Excessive eye elongation is a key feature of myopia, and this has led to the hypothesis that common, more benign, polymorphisms in *PAX6* may be associated with the development of refractive error, though no association has been found to date.[19] Thus, the findings of the genomewide linkage scan described earlier have generated a great deal of recent interest in *PAX6* as a candidate gene for myopia, and much research effort is being directed toward it.

The 3q26 locus contains the *SOX2* gene, a member of the family of sex-determining region Y-box transcription factor genes. *SOX2* also has a key role in eye development, and mutations in *SOX2* are responsible for syndromic microphthalmia. Although mutations in *SOX2* (as with those in *PAX6*) produce a severe phenotype, its important role in eye development supports the idea that because of common variation, this gene might be a good candidate for study in refractive error.

Thus, we have investigated common variants selected from HapMap across *PAX6* and *SOX2* and tested their association with refractive error in a well-characterized and nationally representative population.

Methods

The 1958 British birth cohort initially comprised all 17,000 persons born in Britain during one week in March 1958. The surviving members have been followed up subsequently at intervals, enabling the compilation of a complex and diverse data set of biomedical and social information.[22] The most recent broad biomedical assessment was undertaken when cohort members were 44 to 45 years of age. This included noncycloplegic autorefraction of each eye using a handheld auto refractometer (Retinomax 2; Nikon, Tokyo, Japan) in 23% of the subjects chosen randomly (details previously reported[23]), providing a reliable and objective quantitative measure of refractive error phenotype at an age when all primary myopia has become manifest.[23] Blood samples were also taken for DNA extraction and for creation of immortalized cell lines. Thus, this population offers unique opportunities for investigating genetic and environmental risk factors for refractive error. Given that the cohort was based on those born in the United Kingdom in 1958, 97% of participants are white British. This research followed the tenets of the Declaration of Helsinki. All participants gave informed written consent to participate in genetic association studies, and the present study was approved by the South East MultiCentre Research Ethics Committee (MREC) and the Oversight Committee for the biomedical examination of the British 1958 cohort. Biomedical examination protocols were approved by the South East MREC.

Refractive error is quantified using the summary measure of spherical equivalent of refraction (SER) in diopters. Although subjects can be divided using thresholds of SER set arbitrarily into categories of myopia (negative SER), hypermetropia (positive SER), and emmetropia (no refractive error), greater power is obtained by treating refractive error as a continuous variable. From autorefraction readings, spherical equivalent of refraction (phenotype) was calculated in the conventional way ($SER = S + C/2$), where S is the sphere and C is the cylinder. Data on subjects whose discordance between the spherical equivalents in each eye was within the worst 5% of the sample were discarded. This difference was defined as

$$\max\left(\frac{(SER_R - SER_L)^2}{SER_R}, \frac{(SER_R - SER_L)^2}{SER_L}\right)$$

where SER_R is the spherical equivalent of the right eye and SER_L is the spherical equivalent of the left eye. If the denominator was 0, that value was excluded, and judgment over including or excluding the subject was made on the basis of the other ratio whose denominator was different from 0.

Five hundred ninety-six persons were selected at random from the lowest and highest tertiles (myopic versus hypermetropic) of the mean SER distribution because comparison of extremely opposed phenotypes was considered more beneficial than analysis of the entire trait spectrum.[24,25]

Refractive error has a leptokurtotic distribution and is skewed toward the myopia end of the distribution (Fig. 1). This means the hypermetropic tertile includes all those with hypermetropia and some with emmetropia, whereas the myopic tertile covers the full spectrum of myopia severity. Importantly, in analyzing RE as a quantitative trait, it is not the absolute value of SER but the ranking within the distribution that is the variable of interest.

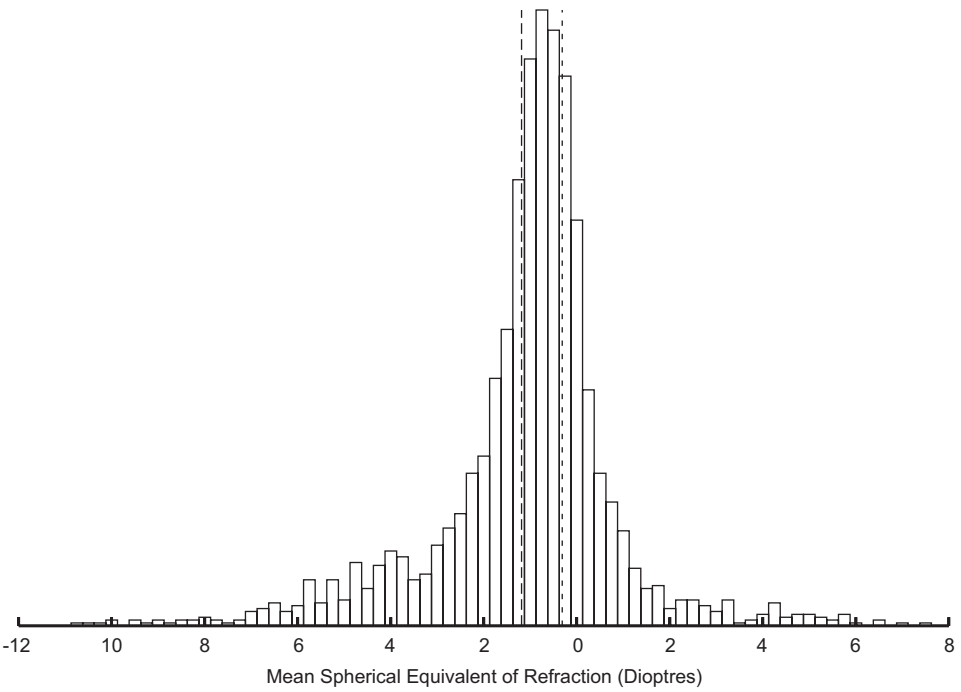

Figure 1. Refraction distribution of the 1958 cohort, with tertiles marked.

Tagging SNPs were selected from phase 2 of the HapMap project using the Tagger algorithm[26] with an r^2 cutoff of 0.8. Genotyping was outsourced to Illumina Inc. and was performed on their genotyping technology (GoldenGate; Illumina Inc., San Diego, CA). Qualitative trait analysis was performed on individual SNPs and 3 SNP sliding window haplotypes using the likelihood ratio test in *Whap*[27] under the assumption of the additive genetic model. Subjects were assigned as myopic or nonmyopic based on their designated tertile. For quantitative trait analysis, the mean spherical equivalent between the two eyes was standardized to a mean of 0 and a variance of 1 for the entire population before selection. Statistical analysis was performed on individual SNPs and 3 SNP sliding window haplotypes using the conditional analysis in *Whap* and individual SNP regression in statistical software (Stata; Stata Corp., College Station, TX).[28] Empiric *P* values were generated using a Monte Carlo method in *Whap,* which permuted the trait values in the sample.

The study presented here was part of a broader association study of refractive error, in which 1536 tag SNPs were chosen across 111 candidate genes. Of these, 25 SNPs were selected across a 530-kb region that included *PAX6* and putative control regions. Three SNPs were selected across a 10.8-kb region centered on *SOX2* and potential control regions. Correction for multiple testing was performed using the false discovery rate.[29] This study had 80% power to exclude either *PAX6* or *SOX2*, contributing to more than 10% of the variance of the refractive error in this cohort. Power calculations were performed using the Genetic Power Calculator[30] based on an additive genetic variance of 0.1. In the cohort ($N = 596$) we analyzed and whose results we are reporting, the variance of refraction was 6.25 and the SD from the mean was 2.50. Using these data, we modeled the power to detect, at $\alpha = 0.05$, any association between genotype and phenotype. The model can be characterized by two key parameters, regression slope and SD, of regression errors.

Results show that our panel was powered more than 80% to detect even minor effects, not only when, as generally expected, deviation from the predicted refraction value was no larger than the SD for the phenotype we observed (2.5) but also when it was twice as large. Our sample has more than 80% power to detect association at a liberal significance level, even in circumstances of small effects from the locus (changes in refraction for each additional copy of susceptibility allele are as small as 0.2–0.3 D). Power to confidently rule out association is more elusive to calculate using most of the standard statistical methods geared at assessing positive association. Power calculations reported assumed an aggressive significance level ($\alpha < .01$) and did not provide sufficient information for lower contribution to the genetic variance.

Table 1. Single SNP *P* Values for Qualitative Trait Analysis in *PAX6* and *SOX2*

SNP	Position (Mb)	*P*-Value: Likelihood Ratio Test	*P*-Value: Permuted
PAX6			
rs17248764	31.36	0.99	0.98
rs509628	31.49	0.71	0.77
rs2996464	31.55	0.16	0.12
rs11031423	31.58	0.04	0.05
rs986527	31.59	0.96	0.98
rs7125966	31.76	0.77	0.69
rs2177482	31.76	0.08	0.10
rs3026401	31.76	0.45	0.54
rs3026398	31.77	0.41	0.38
rs662702	31.77	0.55	0.58
rs1506	31.77	0.75	0.76
rs3026393	31.77	0.26	0.21
rs2071754	31.77	0.78	0.82
rs17646359	31.84	0.55	0.58
rs677874	31.85	0.13	0.20
rs11825821	31.85	0.54	0.57
rs11031505	31.86	0.57	0.59
rs7106566	31.86	0.60	0.65
rs16922551	31.86	0.81	0.83
rs604900	31.87	0.94	0.93
rs17719728	31.87	0.46	0.50
rs621420	31.88	0.84	0.74
rs10488687	31.88	0.77	0.73
rs16922585	31.88	0.60	0.56
rs586662	31.89	0.26	0.29
SOX2			
rs12497248	182.91	0.77	0.76
rs11915160	182.91	0.10	0.08
rs4459940	182.92	0.54	0.52

Results

All SNPs were in Hardy-Weinberg equilibrium and had a genotyping failure of 5% or less. Analysis of refraction as a qualitative trait did not show significant differences between allele or genotype frequencies after correction for multiple testing. Only one of the *uncorrected P* values was marginally significant (rs11031423, $P = 0.04$) and less significant by permutation testing ($P = 0.05$; Table 1). Haplotype analysis also did not show any significant association (data not shown). Quantitative trait analysis using the likelihood ratio test failed to show any association between mean SER and any of the SNPs tested, even before correction for multiple testing (Table 2), and haplotype analysis was also nonsignificant (data not shown). One SNP was marginally significant by regression analysis before correction for multiple testing (rs2996464, $P = 0.04$).

Discussion

Our findings strongly suggest that refractive error is not directly affected in this population by variants in *PAX6*, *SOX2*, or their known promoters and enhancers. This finding is important for understanding the pathophysiology of refractive error given the fundamental role each gene is known to play in the development and growth of the human eye.

We think it unlikely that potential errors in our study account for this unexpected finding. Unlike many of the early candidate gene studies that were underpowered or overinterpreted and whose findings were irreproducible,[31] the design and size of our study are sufficiently robust to allow us to exclude these as explanations for our finding. Despite the undisputed practical importance of the conclusions of the HapMap project, it has been suggested that HapMap SNPs do not necessarily capture variants that have not been genotyped by the HapMap project.[32] Therefore, it remains possible that other polymorphisms not typed by HapMap will not be in significant enough linkage disequilibrium (LD) with any of the tagSNPs chosen to be sufficiently captured. Most of the SNPs reported by dbSNP but not typed by the HapMap around *PAX6* and *SOX2* have low minor allele frequencies or are not polymorphic in Caucasians because both genes are highly conserved. The CD/CV hypothesis suggests that such rare variants are not likely to be of relevance to common traits. Given the extensive research literature on *PAX6* and *SOX2*, it seems unlikely that the failure to identify an association between either of these genes and refractive error results from the existence of yet unidentified common variants. The theoretical alternative to the CD/CV hypothesis is the common disease, rare variant hypothesis (CD/RV),[33] where multiple rare variants within genes are responsible for the development of complex traits. This hypothesis was not tested and is not ruled out by this study. There is little empiric evidence for rare variants in most complex traits to date, and we do not think this is a likely underlying paradigm in refractive error.

TABLE 2. Single SNP *P* Values for Quantitative Trait Analysis in *PAX6* and *SOX2*

SNP	Position (Mb)	*P*-Value	
		Likelihood Ratio Test	Regression
PAX6			
rs17248764	31.36	0.39	0.40
rs509628	31.49	0.25	0.44
rs2996464	31.55	0.29	0.04
rs11031423	31.58	0.29	0.27
rs986527	31.59	0.33	0.51
rs7125966	31.76	0.29	0.16
rs2177482	31.76	0.33	0.05
rs3026401	31.76	0.33	0.89
rs3026398	31.77	0.34	0.09
rs662702	31.77	0.30	0.59
rs1506	31.77	0.27	0.61
rs3026393	31.77	0.26	0.23
rs2071754	31.77	0.56	0.59
rs17646359	31.84	0.86	0.96
rs677874	31.85	0.92	0.81
rs11825821	31.85	0.69	0.85
rs11031505	31.86	0.72	0.29
rs7106566	31.86	0.81	0.71
16922551	31.86	0.91	0.96
rs604900	31.87	0.87	0.55
rs17719728	31.87	0.73	0.28
rs621420	31.88	0.78	0.62
rs10488687	31.88	0.77	0.61
rs16922585	31.88	0.93	0.58
rs586662	31.89	0.92	0.34
SOX2			
rs12497248	182.91	0.89	0.92
rs11915160	182.91	0.13	0.38
rs4459940	182.92	0.41	0.56

Our study serves to highlight the critical role of replication in the interpretation of genetic research. Unfortunately, good evidence indicates significant publication bias in favor of "positive" findings if they are novel and contrary to findings that either rule out important hypotheses ("negative" findings) or fail to replicate earlier work and thereby demonstrate prior false discovery. Notably, it is generally agreed[34,35] that regardless of the relative quality of the science, it takes longer to achieve publication of negative results, and it is difficult to get them published in journals with higher impact factors. Indeed such findings are sometimes not even submitted for publication. Instead they are labeled "file drawer problem"[36] in anticipation of the difficulties. Promotion of false gene discovery in this way is certain to mean that research efforts and funds are being directed to areas unlikely to be fruitful.

PAX6 and *SOX2* were obvious candidates for refractive error, based on knowledge of biological roles and linkage data. We recognize that the lack of association in our study does not preclude *PAX6* or *SOX2* involvement in refractive error caused by mechanisms such as regulation of expression by environmental factors or undiscovered *cis*- or *trans*-acting elements. However, from the findings of the present study, we suggest that neither *PAX6* nor *SOX2* should be prioritized in the active international search for risk and phenotype modifiers of refractive error.

Acknowledgments

The authors thank the Great Ormond Street Hospital Special Trustees for funds for 10 Retinomax 2 autorefractors. They also thank the Centre for Longitudinal Studies, IOE, and the ALSPAC DNA Bank at the University of Bristol for help with accessing data and DNA samples.

References

1. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science*. 1996;273:1516–1517.
2. International HapMap Consortium. A haplotype map of the human genome. *Nature*. 2005;437:1299–1320.
3. de Bakker PI, Burtt NP, Graham RR, et al. Transferability of tag SNPs in genetic association studies in multiple populations. *Nat Genet*. 2006;38:1298–1303.
4. Goldschmidt E. [On the etiology of myopia: an epidemiological study.] *Acta Ophthalmol (Copenh)*. 1968;98(suppl):1+.
5. Javitt JC. The cost-effectiveness of restoring sight. *Arch Ophthalmol*. 1993;111:1615.
6. Dandona R, Dandona L. Refractive error blindness. *Bull World Health Organ*. 2001;79:237–243.
7. Katz J, Tielsch JM, Sommer A. Prevalence and risk factors for refractive errors in an adult inner city population. *Invest Ophthalmol Vis Sci*. 1997;38:334–340.
8. Kempen JH, Mitchell P, Lee KE, et al. The prevalence of refractive errors among adults in the United States, Western Europe, and Australia. *Arch Ophthalmol*. 2004;122:495–505.
9. Kleinstein RN, Jones LA, Hullett S, et al. Refractive error and ethnicity in children. *Arch Ophthalmol*. 2003;121:1141–1147.
10. Saw SM, Katz J, Schein OD, Chew SJ, Chan TK. Epidemiology of myopia. *Epidemiol Rev*. 1996;18:175–187.
11. Tay MT, Au Eong KG, Ng CY, Lim MK. Myopia and educational attainment in 421,116 young Singaporean males. *Ann Acad Med Singapore*. 1992;21:785–791.
12. Guggenheim JA, Kirov G, Hodson SA. The heritability of high myopia: a reanalysis of Goldschmidt's data. *J Med Genet*. 2000;37:227–231.
13. Hammond CJ, Snieder H, Gilbert CE, Spector TD. Genes and environment in refractive error: the twin eye study. *Invest Ophthalmol Vis Sci*. 2001;42:1232–1236.
14. Lyhne N, Sjolie AK, Kyvik KO, Green A. The importance of genes and environment for ocular refraction and its determiners: a population based study among 20–45 year old twins. *Br J Ophthalmol*. 2001;85:1470–1476.
15. Hui J, Peck L, Howland HC. Correlations between familial refractive error and children's non-cycloplegic refractions. *Vision Res*. 1995;35:1353–1358.
16. Keller JT. A comparison of the refractive status of myopic children and their parents. *Am J Optom Arch Am Acad Optom*. 1973;50:206–211.
17. Krause UH, Rantakallio PT, Koiranen MJ, Mottonen JK. The development of myopia up to the age of twenty and a comparison of refraction in parents and children. *Arctic Med Res*. 1993;52:161–165.
18. Sorsby A, Leary GA, Fraser GR. Family studies on ocular refraction and its components. *J Med Genet*. 1966;3:269–273.
19. Hammond CJ, Andrew T, Mak YT, Spector TD. A susceptibility locus for myopia in the normal population is linked to the *PAX6* gene region on chromosome 11: a genomewide scan of dizygotic twins. *Am J Hum Genet*. 2004;75:294–304.
20. Hever AM, Williamson KA, van Heyningen V. Developmental malformations of the eye: the role of *PAX6*, *SOX2* and *OTX2*. *Clin Genet*. 2006;69:459–470.
21. Schedl A, Ross A, Lee M, et al. Influence of PAX6 gene dosage on development: overexpression causes severe eye abnormalities. *Cell*. 1996;86:71–82.
22. Power C, Elliott J. Cohort profile: 1958 British birth cohort (National Child Development Study). *Int J Epidemiol*. 2006;35:34–41.
23. Cumberland PM, Peckham CS, Rahi JS. Inferring myopia over the lifecourse from uncorrected distance visual acuity in childhood. *Br J Ophthalmol*. 2007;91:151–153.

24. Abecasis GR, Cookson WO, Cardon LR. The power to detect linkage disequilibrium with quantitative traits in selected samples. *Am J Hum Genet*. 2001;68:1463–1474.
25. Schork NJ, Nath SK, Fallin D, Chakravarti A. Linkage disequilibrium analysis of biallelic DNA markers, human quantitative trait loci, and threshold-defined case and control subjects. *Am J Hum Genet*. 2000;67:1208–1218.
26. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. *Nat Genet*. 2005;37:1217–1223.
27. Purcell S, Daly MJ, Sham PC. WHAP: haplotype-based association analysis. *Bioinformatics*. 2007;23:255–256.
28. Stata Statistical Software. Release 9. College Station, TX: StataCorp.; 2006. *http://www.stata.com/*
29. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B*. 1995;57:289–300.
30. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*. 2003;19:149–150.
31. Calnan M, Smith GD, Sterne JA. The publication process itself was the major cause of publication bias in genetic epidemiology. *J Clin Epidemiol*. 2006;59:1312–1318.
32. Tantoso E, Yang Y, Li KB. How well do HapMap SNPs capture the untyped SNPs? *BMC Genomics*. 2006;7:238.
33. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet*. 2001;69:124–137.
34. Chalmers TC, Frank CS, Reitman D. Minimizing the three stages of publication bias. *JAMA*. 1990;263:1392–1395.
35. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet*. 1991;337:867–872.
36. Kennedy D. The old file-drawer problem. *Science*. 2004;305:451.