

EDITORIAL

When Is a Replication Not a Replication? Or How to Spot a Good Genetic Association Study

Tim D. Spector, Kourosh R. Ahmadi, and Ana M. Valdes

Published genetic association studies are becoming an epidemic. For example, whereas only 53 PubMed entries are found for the keywords “polymorphism” and “arthritis” in 1998, the number of entries is nearly 3-fold higher for 2004. With the recent completion of phase I of the haplotype map (HapMap) project and the continuing drop in genotyping costs, genome-wide association studies using up to 500,000 markers are becoming feasible. Soon we can expect a flourish of genetic association studies with positive findings to appear in the leading journals. Although the findings of these studies will be exciting at first glance, the vast amount of genes and polymorphisms tested can leave both readers and reviewers confused as to their interpretation. If already published studies serve as predictors of what is to come, a high percentage of the to-be-published genetic association studies may yield one-off results that will fail to be consistently replicated. In fact, according to recent meta-analyses, only 16–30% of initially reported significant associations have been consistently replicated (1,2).

Encouragingly, we know some of the reasons for this inconsistency. First, the attitude of the modern gene hunter has not yet evolved to keep pace with the recent developments in genetic epidemiology, and the optimism is still largely based on findings for Mendelian disease (3). To combat this, useful guidelines that addressed this issue were produced nearly 2 years ago in *Arthritis & Rheumatism* (4) (see Table 1). Given the advances in technology and methodology, however, the current dilemma involves 3 questions: What do results actually mean? Can they be relied upon? And, most

important, is it worth investing more funds to follow them up?

A number of factors can explain the lack of confirmation of an initial positive finding. These include 1) overestimation of the genetic effect in the first report, 2) inconsistent coverage of genetic variation in the gene, 3) inherent genetic differences in the populations studied, and 4) differences in the clinical definition of phenotype.

Overestimation of the genetic effect in the first report

Meta-analyses have revealed that the estimated genetic effect in the first report of positive findings in the vast majority of cases is larger than the true effect (2), which is explained by a phenomenon called the “winner’s curse.” That is, the genetic effect (odds ratio [OR]) in an association study will be biased upward as a result of selection due to the pressure on the investigators to be the first to report a statistically significant result. Ideally, therefore, a study aiming to confirm a reported genetic association should have a sample size large enough to detect genetic effects likely to be smaller than the original association. Indeed, low-powered attempts at replication in which investigators conclude no evidence for association can be extremely misleading.

Inconsistent coverage of genetic variation in the gene

One way of attempting to replicate a reported result has been to genotype only the most significant single-nucleotide polymorphisms (SNPs) from the original study. This incurs a high risk of a false-negative result because of the failure to include or indirectly “tag” the relevant functional variant (5) within the replication sample. The testing of only known functional variants has the advantage of biologic plausibility, but such testing assumes a full knowledge of gene function and how this is influenced by genetic variation (6). The utility of haplotype-tagging SNPs (via the HapMap project) will go some way toward alleviating this problem.

Tim D. Spector, MD, MSc, FRCP, Kourosh R. Ahmadi, PhD, Ana M. Valdes, PhD: St. Thomas’ Hospital, Kings College London, London, UK.

Address correspondence and reprint requests to Tim D. Spector, MD, MSc, FRCP, Twin Research and Genetic Epidemiology Unit, St. Thomas’ Hospital, Lambeth Palace Road, London SE1 7EH, UK. E-mail: tim.spector@kcl.ac.uk.

Submitted for publication December 9, 2005; accepted in revised form December 20, 2005.

Table 1. Summary of recommendations for genetic association studies made by Huizinga et al (4)

Recommendation for selecting populations for study
Demonstrate adequate sample size/statistical power
Avoid hidden population stratification
Consider the use of family-based controls (such as parents), if available
Confirm allele frequencies of single-nucleotide polymorphisms
Recommendation for selecting the gene and genetic variant
Is study hypothesis testing or hypothesis generating?
Are there positional data for the gene?
If the study is hypothesis testing, assess the adequacy of the functional data for markers, including replication by other groups
Consider additional markers to define haplotypes
Recommendation to ensure plausible biologic context
Consider gene-dose effects
Consider properties of the phenotype logically associated with the genotype
Recommendation for quality measures
Specify the quality measures used for the genotyping analysis
Provide information on the reproducibility between quality control replicates

Inherent genetic differences in the populations studied

Even if the exact same markers analyzed by the initial association study are evaluated, the replication may fail to detect a true association. The problem is that due to sampling, the allele frequencies and haplotype structure in the region may not be the same in the 2 cohorts even if they are from the same underlying population (e.g., Caucasians). This problem can be further exacerbated for genes that span a very large physical region when relatively similar populations are compared, and this is even more true when comparing very ethnically diverse groups. Neale and Sham (6) have recently proposed a gene-based replication approach in which the gene identified by the initial study is subsequently reexamined for association using directly or indirectly all genetic variants in the associated gene. This approach is costly, but it avoids some of the above-mentioned pitfalls.

Differences in the clinical definition of phenotype

The phenotypes studied might be related (e.g., osteoarthritis [OA]) but not the same (e.g., hip OA versus generalized OA). A good example appears in this issue of *Arthritis & Rheumatism*, in which Lane and coworkers report a replication study of the *FRZB* gene in OA (7). The *FRZB* gene encodes for a Frizzled-related protein, which helps to maintain articular cartilage, and the associated alleles at *FRZB* reduce the activity of this important protein.

There is a strong biologic rationale for this gene's involvement in OA, and it was originally implicated via fine mapping of the 2q linkage region. Investigators subsequently found an association of the *FRZB* gene with hip OA in women in a case-control study in the UK (8). In particular, the haplotype composed of substitutions at 2 highly conserved arginine residues (Arg200Trp and Arg324Gly) in *FRZB* was highlighted as a strong risk factor for primary hip OA in the original study (OR 4.1, $P = 0.004$). Later, in their Rotterdam study, Min et al (9) attempted to replicate the association of the 2 *FRZB* variants with OA in a large sample of subjects from a population-based cohort selected for the presence of primary symptomatic OA at multiple sites. The authors found no evidence for association with hip OA, but they did find an association of the variant allele at Arg324Gly with generalized radiographic OA (OR 1.41, $P < 0.02$) (Table 2) in the Rotterdam study alone, which turned out not to be statistically significant ($P < 0.10$) when the Rotterdam data were combined with those from the Genetics, osteoARthritis and Progression (GARP) study. In the GARP study, this same variant was associated with familial symptomatic OA at multiple sites (OR 1.6, $P < 0.02$). More recently, Lane and coworkers (7) studied the same *FRZB* variants in the Study of Osteoporotic Fractures (SOF), and, unlike the investigators in the Dutch study, they did find an association between hip OA and the Arg324Gly SNP (OR 1.33, $P = 0.05$). Lane et al also found an association between hip OA (adjusted for confounding factors) and the original haplotype reported by the UK study investigators, but with a smaller OR of 1.5 ($P < 0.05$).

For readers of *Arthritis & Rheumatism*, the 3 studies can be difficult to pick apart individually; in the future, reviewers should, perhaps, demand more useful contrasts. Nevertheless, it appears that Lane et al have essentially replicated the earlier findings of Loughlin et al (8), while both groups' findings contrast with those reported by Min et al (9) (Table 2). Two aspects of the above-mentioned studies, however, deserve further discussion. First, the Rotterdam study was carried out using a mixed cohort of men and women, whereas the UK and SOF studies showed an association with hip OA only among women. Second, the estimated haplotype frequency was 4-fold higher in SOF controls than in UK controls, since each group used a different method to derive haplotype counts. These discrepancies in study design could partly explain the differing results from the 3 studies.

As discussed above, the prevailing view is that replication should be gene based and not allele based.

Table 2. *FRZB* association in 3 different studies*

Study (ref.), trait/sex	No. of OA patients/no. of controls	R200W frequency,			R234G frequency,			WG haplotype frequency, OA		
		OA patients/ controls	OR†	<i>P</i>	OA patients/ controls	OR†	<i>P</i>	patients/ controls	OR†	<i>P</i>
UK (8)										
Hip OA/F only	558/399	0.136/0.115	1.21	<0.17	0.104/0.071	1.52	<0.02	0.026/0.006	4.1	0.004
Hip OA/M only	378/361	0.128/0.137	0.92	<0.60	0.067/0.056	0.75	<0.16	NA	–	–
Rotterdam (9)										
Hip OA/M + F	148/917	0.125/0.129	0.96	<0.84	0.075/0.076	0.99	<0.96	NA	–	–
Generalized OA/M + F	283/917	0.113/0.129	0.86	<0.26	0.103/0.076	1.41	<0.02	NA	–	–
SOF (7)										
Hip OA JSN/F only‡	569/1,317	0.142/0.117	1.25	0.14	0.109/0.083	1.36	0.21	0.046/0.025	1.9	<0.01

* OA = osteoarthritis; OR = odds ratio; NA = not available; SOF = Study of Osteoporotic Fractures; JSN = joint space narrowing.

† All ORs are unadjusted and are for alleles/haplotypes.

‡ The JSN definition of OA was used for comparison for greater consistency with the study by Loughlin and coworkers (8).

For known functional disease genes, it is common for subtle differences in alleles or haplotypes to appear in different populations. In contrast, there are no clear examples of risk alleles having genuinely opposite effects in different studies.

This leads to the questions of what we mean by replication and what constitutes a replication sample. Ideally, a replication sample should be an adequately powered, independent sample (cases and controls) ascertained from the same source population and matched for the same disease definition and other epidemiologic parameters to the original sample. Some studies are well equipped to deal with this, since they are large enough to plan for these requirements a priori. However, for practical reasons, often slightly different populations are used. An often-ignored distinction and source of confusion arises when an associated variant is tested in an independent sample not matched epidemiologically to the original sample. In this case, the original association can be replicated in the new sample only if the underlying genetic make-up and epidemiologic characteristics of the second sample are identical to those of the original population. For example, if a genetic variant that has been associated with a phenotype in a Caucasian sample is tested and shown to be associated in the Japanese as well, we would label this a “confirmation” of the association in another population, rather than a replication. If the variant and phenotype were not associated in the Japanese, one would hardly call this a nonreplication.

Whether a replication study is designed around your own or someone else’s findings, it is imperative that it be designed and implemented to a high standard. This point takes on even more importance when designing a replication study around genome-wide association stud-

ies, such as one by our group (10), in which 25,000 “random” SNPs were tested in a DNA pooling analysis. After an initial replication, the association of the gene *LRCHI* with knee OA was replicated in 2 separate populations. For studies involving large numbers of SNPs, several replications are the only way to deal with multiple testing problems, although it is difficult to interpret results involving genes such as *LRCHI* that have no known function. On the one hand, a clear balance has to be struck between reporting results of flawed studies with no replication and taking the strong chance of reporting a false-positive result. On the other hand, there are no association studies that are perfect.

Very stringent *P* value guidelines are not necessarily the solution. The validity of the *P* value depends on the quality of the data put into a study and on the quality of the analyses employed (11). Moreover, some “true” polymorphisms consistently associated with complex diseases have relatively modest effects, such as *PPARγ* with type 2 diabetes or *CTLA4* with type 1 diabetes (both with ORs <1.3). Requiring extremely small *P* values as proof would result in the exclusion of these true genetic associations.

If replication is difficult, one balancing factor is the ability to demonstrate functional importance of the gene through in vitro studies (12). A group of Japanese studies has shown this effectively (e.g., see refs. 13 and 14) for the genes *ASPN* and *CALMI*, showing replication and demonstrating function, although the genetic associations may not replicate in Caucasians. Unfortunately, we are not yet at a stage where we can accurately test for the functional effect of all associated variants.

In conclusion, genetic association studies and the assembly of large case-control groups should continue to be encouraged, although sufficient size and at least a

second replication group are crucial. Meanwhile, the results of association studies should be examined critically, since different populations may have different SNPs that contribute to genetic risk. The final point is that, contrary to our traditional emphasis on research novelty, the larger, second, “less novel” replication study is often the hardest to perform and the most scientifically important—and the one about which you should read.

REFERENCES

1. Ioannidis JP. Genetic associations: false or true? *Trends Mol Med* 2003;9:135–8.
2. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 2003;33:177–82.
3. Goldstein DB, Cavalleri GL, Ahmadi KR. The genetics of common diseases: 10 million times as hard. *Cold Spring Harb Symp Quant Biol* 2003;68:395–401.
4. Huizinga TW, Pisetsky DS, Kimberly RP. Associations, populations, and the truth: recommendations for genetic association studies in *Arthritis & Rheumatism*. *Arthritis Rheum* 2004;50:2066–71.
5. Ahmadi KR, Weale ME, Xue ZY, Soranzo N, Yarnall DP, Briley JD, et al. A single-nucleotide polymorphism tagging set for human drug metabolism and transport. *Nat Genet* 2005;37:84–9.
6. Neale B, Sham PC. The future of association studies: gene-based analysis and replication. *Am J Hum Genet* 2004;75:353–62.
7. Lane NE, Lian K, Nevitt MC, Zmuda JM, Lui L, Li J, et al. Frizzled-related protein variants are risk factors for hip osteoarthritis. *Arthritis Rheum* 2006;54:1246–54.
8. Loughlin J, Dowling B, Chapman K, Marcelline L, Mustafa Z, Southam L, et al. Functional variants within the secreted frizzled-related protein 3 gene are associated with hip osteoarthritis in females. *Proc Natl Acad Sci U S A* 2004;101:9757–62.
9. Min JL, Meulenbelt I, Riyazi N, Kloppenburg M, Houwing-Duistermaat JJ, Seymour AB, et al. Association of the Frizzled-related protein gene with symptomatic osteoarthritis at multiple sites. *Arthritis Rheum* 2005;52:1077–80.
10. Spector TD, Reneland RH, Mah S, Valdes AM, Hart DJ, Kammerer S, et al. Association between a variation in LRCH1 and knee osteoarthritis: a genome-wide single-nucleotide polymorphism association study using DNA pooling. *Arthritis Rheum* 2006;54:524–32.
11. Page GP, George V, Go RC, Page PZ, Allison DB. “Are we there yet?”: deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits. *Am J Hum Genet* 2003;73:711–9.
12. Rebbeck TR, Spitz M, Wu X. Assessing the function of genetic variants in candidate gene association studies. *Nat Rev Genet* 2004;5:589–97.
13. Kizawa H, Kou I, Iida A, Sudo A, Miyamoto Y, Fukuda A, et al. An aspartic acid repeat polymorphism in asporin inhibits chondrogenesis and increases susceptibility to osteoarthritis. *Nat Genet* 2005;37:138–44.
14. Mototani H, Mabuchi A, Saito S, Fujioka M, Iida A, Takatori Y, et al. A functional single nucleotide polymorphism in the core promoter region of CALM1 is associated with hip osteoarthritis in Japanese. *Hum Mol Genet* 2005;14:1009–17.