SOURCE OF HETEROGENEITY

A number of reasons have been advanced for heterogeneity in the genetic effects across the results of various studies.[8,13,14,47] False-positive results in the initial studies and false-negative results in small replication studies are implicated as the most likely reasons for non-replications.[8–10,13,14] Inconsistency and between-study heterogeneity may be caused because of biases or genuine differences in the genetic effects across populations. We review briefly in this article.

Biases

Differential biases due to population stratification, misclassification of clinical outcome, genotyping error and overestimation of genetic effect in the first study can be sources of between-study heterogeneity.

The presence of population stratification tends to spurious associations. It can be caused when there are undetected genetically different subgroups within a study population and disease prevalence differs among these subgroups.[11,62] The effect of population stratification on the results of genetic association studies is debatable.[62–66] According to systematic reviews of meta-analyses of genetic association studies, it is not so much frequent that difference in racial or ethnic groups could explain heterogeneity.[9,67]

Inadequate assignment of cases and controls may cause misclassification bias. Although there is a possibility that misclassification of cases and controls would weaken the gene–disease association, the results of misclassification bias may be modest unless the trait is common.[13,32]

Ioannidis *et al.*[10] conducted a systematic review of 36 meta-analyses including a total of 370 genetic association studies. Statistically significant between-study heterogeneity was observed in 14 meta-analyses. Restricting to meta-analyses with at least 15 studies, 7 of 9 meta-analyses showed significant heterogeneity. In 25 or 26 meta-analyses, the first study showed more predisposing or protective OR than subsequent replication studies. Using cumulative meta-analysis plots, the authors depicted the process that strong associations claimed in the first study were regressed toward null associations, as subsequent replication studies were accumulated over time. Similar findings were reported in Lohmueller *et al.*[9] Associations passing predetermined thresholds of statistical significance tend to overestimate the size of the genetic effect, especially when the sample size of the study is small and the threshold is stringent in multiple testing situations.[68–74] Such an upward bias is called as *winner's curse* phenomenon.[9,69]

Genuine differences

Differences in the pattern of LD structure over chromosomal regions of interest across populations are implicated as a cause of between-study heterogeneity in the genetic effects. Zondervan and Cardon[75] show that marker allelic OR can vary according to the extent of LD between marker and true disease allele in terms of D' and according to mismatch between disease allele frequency and marker allele frequency. This issue may be especially pronounced in the GWA settings because the SNPs that most efficiently surrogate the other SNPs in a genomic region with high LD (that is, tag SNPs) rather than putative functional SNPs have been used to increase genome coverage. When the extent of LD between tag SNP and true disease allele varies across studied populations, the observed ORs could vary across studies.

Many common diseases are implicated to have a complex etiology involving multiple genetic and environmental factors including their interactions. Gene–disease associations can be modified when the gene–gene or gene–environment interaction exists. If these interactions are not identified and controlled for, the gene–disease associations would be heterogeneous across populations according to distribution of a genetic variant or prevalence of a particular environmental exposure. It is needed to conduct a consortium-based meta-analysis of individual patient data in large scale to account for gene–gene or gene–environment interactions.[47]

SIMULATION STUDY

We conducted a simulation study to illustrate (i) the power of Cochran's Q test, (ii) the properties of measures of between-study heterogeneity (I^2 and H_M^2) and (iii) the type I error rate and the power of meta-analysis for detecting the gene–disease association in the presence of between-study heterogeneity.

We consider meta-analysis of k case–control association studies to estimate the overall genetic effect (θ; log OR) of disease outcome. The exposure status (*AA*, *Aa* and *aa*) of subjects included in each case–control study are ascertained in the sampling manner outlined below.[70] The values $y \in \{1, 0\}$ are labels encoding case (1) or control (0). Let A denote the susceptibility allele, we assume the dominant model and then the SNP genotype predictor value x was designed as 1=*AA* or *Aa*, 0=*aa*. Under the assumption of HWE, the frequency of x written as f_x is calculated based on the disease allele frequency f_A: $f_1 = 1-(1-f_A)^2$. The logistic regression model for ith study ($i = 1, 2, \ldots, k$) is produced as follows:

$$\log\left(\Pr(Y=1|x)/(1-\Pr(Y=1|x))\right) = \alpha_i + \theta_i x$$

where α_i is the intercept and θ_i is the log OR for ith study. θ_i is drawn from $N(\theta, \tau^2)$. τ^2 is the between-study variance. α_i can be calculated by using the equation for the prevalence of the disease $\pi = \sum_x \frac{\exp(\alpha_i+\theta_i x)}{1+\exp(\alpha_i+\theta_i x)} \times f_x$. The genotypes of case and control subjects are generated based on the conditional probabilities of x given by y as follows:

$$\Pr(X = x|Y = 1) = \frac{f_x}{\pi} \times \frac{\exp(\alpha_i+\theta_i x)}{1+\exp(\alpha_i+\theta_i x)},$$

$$\Pr(X = x|Y = 0) = \frac{f_x}{1-\pi} \times \frac{1}{1+\exp(\alpha_i+\theta_i x)}$$

For each study, the genotypes of case–control samples were generated and then the OR and its variance were calculated. Then, the ORs for k studies were combined by FEM and REM meta-analyses. Cochran's Q test was conducted and the I^2 and H_M^2 were measured.

We considered simple five simulation scenarios of meta-analyses. The description of simulation scenarios is shown in Table 1. The scenarios I, II and III were designed to be same in sample size within each study but different in the number of included studies. In scenarios III, IV and V, numbers of studies were different but total number of case–control samples included in meta-analysis was fixed at 20 000. The pairs of scenarios I and V or II and IV were designed to have the same number of studies but differ in sample size within each study.

We examined 126 parameter combinations for each scenario. The between-study variance (τ^2) varied from 0.0 to 0.02 with increments of 0.001. The true summary OR ($\exp(\theta)$) was set to be 1.0, 1.4 or 2.0. The disease allele frequency f_A was assigned to be 0.1 or 0.3. The disease prevalence π was fixed at 0.01. The values of τ^2 were based on the literature values reported by Moonesinghe *et al.*[76] for the confirmed 10 loci in a meta-analysis of three GWA studies of type 2 diabetes.[77] Therefore, our simulation would reflect the possible range of between-study variance. For each scenario and parameter combination, 100 000 simulations were carried out.

Table 1 Description of five simulation scenarios of meta-analysis

Scenario	k	$n_{case}/n_{control}$
I	5	500/500
II	10	500/500
III	20	500/500
IV	10	1000/1000
V	5	2000/2000

k denotes the number of included studies and n_{case} and $n_{control}$ are the number of cases and controls within each study, respectively.

The empirical power of Cochran's *Q* test was evaluated by the proportion of the simulation runs crossing the significance level of 0.1 when $\tau^2 > 0.0$. The top row of Figure 2 shows the powers of Cochran's *Q* test obtained with five scenarios as the function of τ^2 when the overall OR=1.0 and f_A=0.1 or 0.3. For each scenario, the power increased as τ^2 increased. Comparing among scenarios I, II and III, the power increased as the number of studies increased. When total number of case–control samples was fixed (that is, comparing among scenarios III, IV and V), the powers were similar but scenarios with smaller number of studies showed higher power

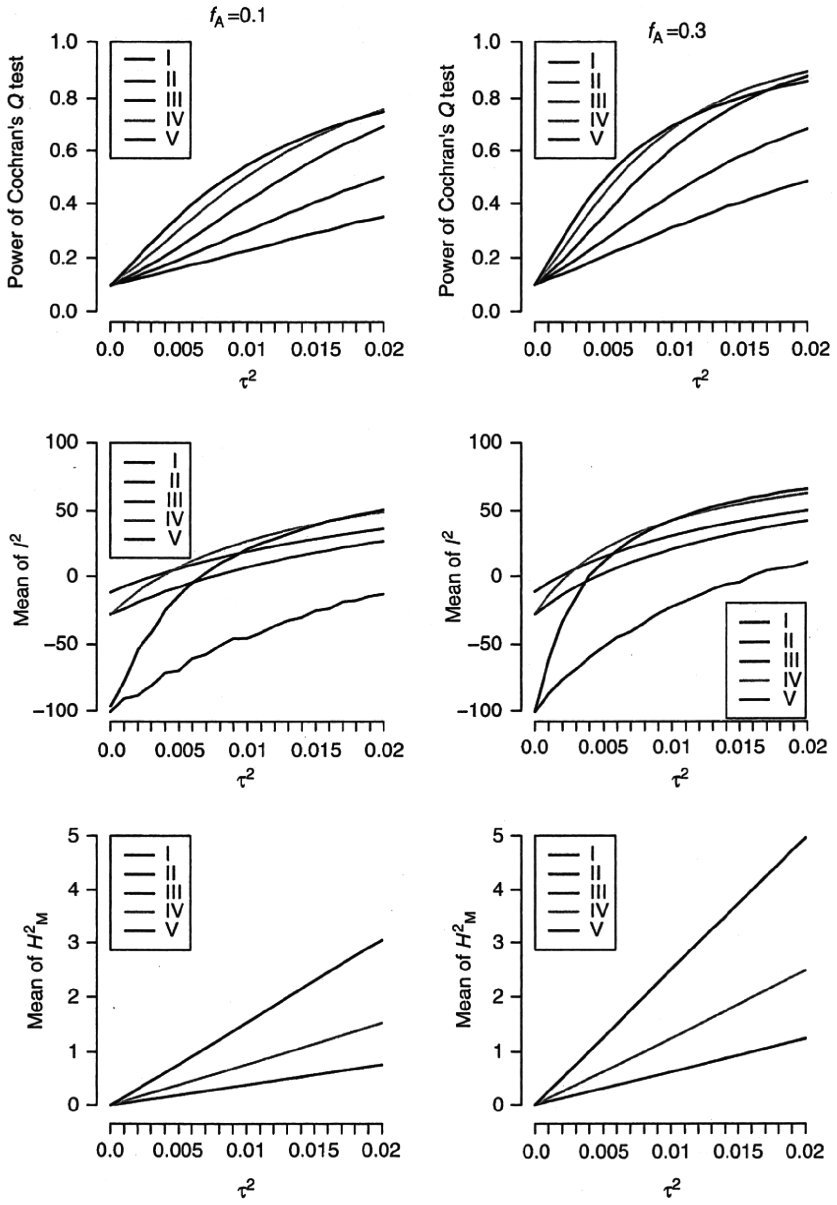


Figure 2 Behaviors of test for and measures of between-study heterogeneity for five simulation scenarios as the function of τ^2, the disease allele frequency f_A=0.1 or 0.3, and the overall odds ratio (OR)=1.0. The top row shows the power of the Cochran's *Q* test at the significance level of 0.1. The middle and bottom rows show the means of I^2 and H^2_M, respectively. The lines of H^2_M for scenarios I, II and III are overlapping. The description of each simulation scenario is in Table 1.

when τ^2 was small. When numbers of studies were identical (that is, two pairwise comparisons of scenarios I versus V or II versus IV), meta-analyses with larger sample size showed higher power for the same τ^2. The powers obtained with f_A=0.3 were higher than those with f_A=0.1. For most of our parameter settings, the powers of Cochran's Q test did not reach at 0.8, although the significance level was set to be 0.10.

The means of 100 000 simulated values for the measures of heterogeneity (I^2 and H_M^2) are shown as the function of τ^2 when the overall OR=1.0 and f_A=0.1 or 0.3 (the middle and bottom rows of Figure 2). In practice, max{0, I^2} and max{0, H_M^2} are used to restrict the ranges of these measures as positive. As the simulation study of Mittlbock and Heinzl,[59] unrestricted values of I^2 and H_M^2 were used to obtain unbiased distributions for these measures in this study. These two measures presented monotonic increases as τ^2 increased. I^2 and H_M^2 increased as the sample size per study increased (scenarios I versus V or II versus IV). The two measures obtained with f_A=0.3 were higher than those with f_A=0.1. These results indicate that I^2 and H_M^2 increased as within-study variance, $k/(\sum_{i=1}^{k} w_i)$, decreased. Comparing scenarios I, II and III shows the important difference between I^2 and H_M^2: whereas I^2 increased as the number of studies increased, H_M^2 did not change (the lines of H_M^2 for scenarios I, II and III are overlapping in the bottom rows of Figure 2). This suggests that H_M^2 may be a good indicator of comparing the extent of between-study heterogeneity across meta-analyses. Similar results and further discussion are provided by Mittlbock and Heinzl.[59] The 95% intervals of simulated I^2 and H_M^2 were large, especially when the number of studies is small (Supplementary Figure S1).

The type I error rate in meta-analysis was assessed as the proportion of the simulation runs showing significant summary OR at the significance level of 0.05 when the null hypothesis was true (that is, the true overall OR=1.0). Figure 3 shows the type I error rates of five scenarios when f_A=0.1 or 0.3. When there was no between-study variance (τ^2=0.0), the type I error rates under FEM were well controlled at 0.05, but REM showed slightly conservative results (the type I error rate≈0.04). As τ^2 increased, the type I error rates under FEM rapidly inflated, but those under REM slightly increased. The type I error rates under both models for the same τ^2 increased when sample size per study was large or f_A=0.3. We should note that the use of FEM could increase the type I error rate even to the extent that the between-study heterogeneity could not be fully identified by Cochran's Q test and two measures I^2 and H_M^2. For example, in case of τ^2=0.005 and f_A=0.3, the type I error rate under FEM for five scenarios were 8.5–19.2% (Figure 3). For the parameter setting, the powers of Cochran's Q-test were 20.6–48.3%, the means of I^2 were −51.9 to 20.8% and the means of H_M^2 were 0.31–1.25 (Figure 2).

The power of detecting a gene–disease association was evaluated as the proportion of simulation runs reaching the significance level of 5.7×10^{-7}, assuming the consortium-based meta-analysis of GWA data sets. As shown in Figure 3, applying FEM meta-analysis to heterogeneous genetic associations could lead to false-positive findings; therefore, we considered only REM when assessing the power of

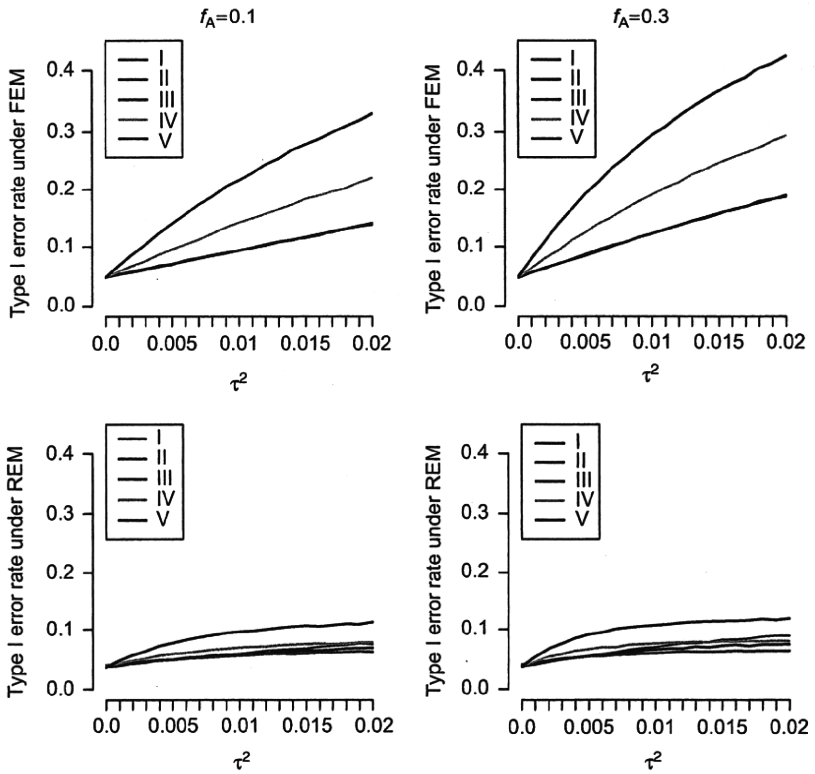


Figure 3 The type I error rate in fixed effects model (FEM) and random effects model (REM) meta-analyses at the significance level of 0.05 for five scenarios as the function of τ^2 and the disease allele frequency f_A=0.1 or 0.3. The top and bottom rows show the type I error rates when applying FEM and REM, respectively. The lines of the type I error rate under FEM for scenarios I, II and III are overlapping. The description of each simulation scenario is in Table 1.

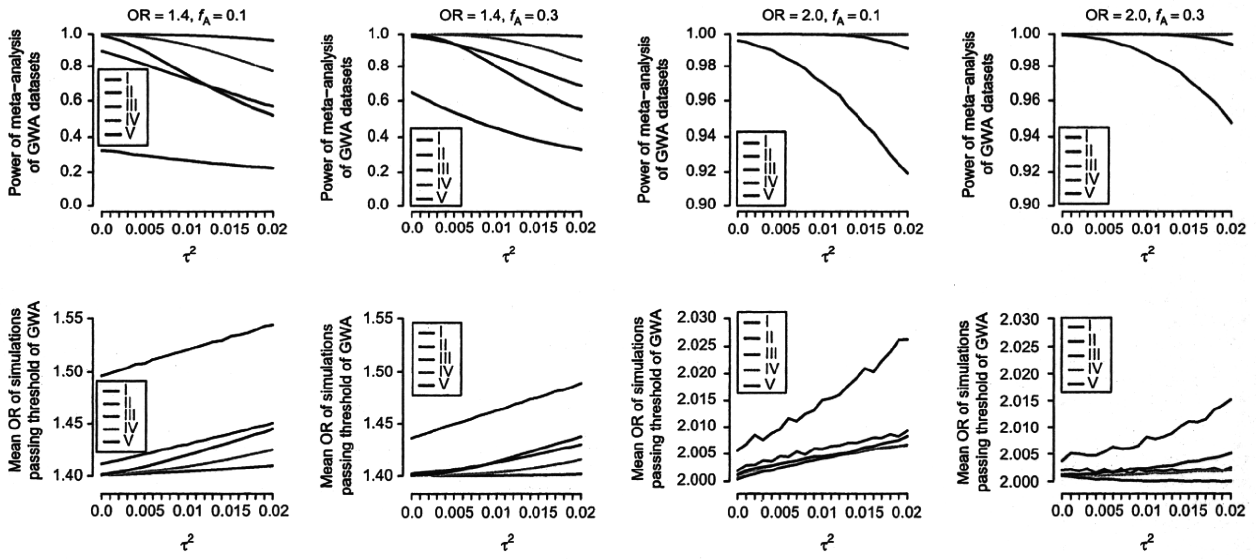


Figure 4 Simulations for the powers in random effects model (REM) meta-analyses of detecting a gene–disease association at the significance level of 5.7×10^{-7} (the top row) and the mean odds ratio (OR) of the simulations passing the threshold (the bottom row) as the function of τ^2, the disease allele frequency f_A=0.1 or 0.3, and the overall OR=1.4 or 2.0. When the overall OR=2.0, the lines of the powers for scenarios II, III and IV are overlapping. The description of each simulation scenario is in Table 1.

meta-analysis. The top row of Figure 4 shows the result, assuming the dominant model and f_A=0.1 or 0.3. When the true overall OR=1.4, the power for each scenario gradually decreased as τ^2 increased. Comparing scenarios III, IV and V, the decreases in the power for the same τ^2 were larger in the scenarios with large sample size per study. While the values of v_{FEM} for scenarios III, IV and V were not different, the values of v_{REM} for scenarios III, IV and V varied when between-study heterogeneity was present. For the same τ^2 (>0), the following inequality was true: v_{REM} for scenario V$>v_{REM}$ for scenario IV$>v_{REM}$ for scenario III. When $\theta\neq0$, the mean of the distribution of the Z-test under REM is $\lambda=\theta/\sqrt{v_{REM}}$. The power of detecting gene-disease association of effect size of θ is[78]

$$\text{Power} = 1-\Phi(C_{\alpha/2}-\lambda)+\Phi(-C_{\alpha/2}-\lambda)$$

where Φ is the cumulative distribution function of the standard normal and $C_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the standard normal distribution. Along with the inequality described above, the decrease in the power for the same τ^2 is larger in the scenarios with large sample size per study when the total sample sizes are equal across scenarios. When the overall OR was set to be 2.0, the powers did not so much decrease in the simulated range of τ^2. Furthermore, we calculated the mean OR of the simulations passing the genome-wide significance threshold (P-value$<5.7\times10^{-7}$). The estimates of mean OR were upwardly biased, especially in scenarios whose powers of detecting gene–disease associations were low (the bottom row of Figure 4). On the other hand, if the meta-analyses were sufficiently powered (for example, the true overall OR=2.0), upward biases were not so pronounced in the simulated range of τ^2.

Our simulation suggests that the power of meta-analysis of GWA data sets to detect small genetic effect would decrease due to between-study heterogeneity ($\tau^2\sim0.02$). As a result, the discovered gene–disease association could have inflated effect (*winner's curse* phenomenon). Such a *winner's curse* phenomenon can be seen even to the extent that the between-study heterogeneity could not be fully identified. Similar results were obtained when different genetic models (that is, recessive and additive in log-odds scale models) were examined (data not shown).

CONCLUSION

We reviewed the process and the methods of meta-analysis of genetic association studies. To conduct and report a transparent meta-analysis, the search strategy, the inclusion or exclusion criteria of studies and the statistical procedures should be fully described. Assessment of HWE and determination of genetic model are methodological issues relevant to meta-analysis of genetic association studies.

In genetic association studies of common disease, effect size of consistently replicated gene–disease associations were found to be small (OR=1.2–1.5);[15] therefore, meta-analysis of GWA data sets is the most important approach to increase the power to detect such gene–disease associations.[35]

Our simulation shows that the power of REM meta-analysis of GWA data sets to detect a small genetic effect could decrease due to between-study heterogeneity and then the mean OR of the simulated meta-analyses that passing the genome-wide significance threshold would be upwardly biased. Recently, Moonesinghe *et al.*[76] show that the required sample size in meta-analysis to detect an overall association with adequate power at a significant level increases as between-study heterogeneity increases and when the between-study heterogeneity exceeds a threshold, meta-analysis cannot reach the power regardless of how large included studies are. At the same time, empirical evaluation of published meta-analyses[61] and our simulation study show the uncertainty of estimated between-study heterogeneity is large unless many studies are combined.

These findings suggest that when a meta-analysis of GWA data sets shows association signals reaching genome-wide significance with small between-study heterogeneity, the result should be cautiously reported and further replication studies by institutions other than GWA teams are required.[35] Moreover, when a large number of data sets are available, challenges to explain and reduce the observed

between-study heterogeneity may become important.[74,76] The knowledge about the potential causes of between-study heterogeneity may help. Such post-GWA research will enable us to map the causative variant finely[79] or to detect polymorphisms associated with clinically important subtypes of diseases.[80]

1 Lander, E. S. The new genomics: global views of biology. *Science* **274,** 536–539 (1996).
2 Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273,** 1516–1517 (1996).
3 The International HapMap Consortium. The International HapMap Project. *Nature* **426,** 789–796 (2003).
4 International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431,** 931–945 (2004).
5 Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
6 Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G. *et al.* The sequence of the human genome. *Science* **291,** 1304–1351 (2001).
7 Manolio, T. A., Brooks, L. D. & Collins, F. S. A HapMap harvest of insights into the genetics of common disease. *J. Clin. Invest.* **118,** 1590–1605 (2008).
8 Hirschhorn, J. N., Lohmueller, K., Byrne, E. & Hirschhorn, K. A comprehensive review of genetic association studies. *Genet. Med.* **4,** 45–61 (2002).
9 Lohmueller, K. E., Pearce, C. L., Pike, M., Lander, E. S. & Hirschhorn, J. N. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat. Genet.* **33,** 177–182 (2003).
10 Ioannidis, J. P., Ntzani, E. E., Trikalinos, T. A. & Contopoulos-Ioannidis, D. G. Replication validity of genetic association studies. *Nat. Genet.* **29,** 306–309 (2001).
11 Cardon, L. R. & Bell, J. I. Association study designs for complex diseases. *Nat. Rev. Genet.* **2,** 91–99 (2001).
12 Freely associating. *Nat. Genet.* **22,** 1–2 (1999).
13 Colhoun, H. M., McKeigue, P. M. & Davey Smith, G. Problems of reporting genetic associations with complex outcomes. *Lancet* **361,** 865–872 (2003).
14 Ioannidis, J. P. Non-replication and inconsistency in the genome-wide association setting. *Hum. Hered.* **64,** 203–213 (2007).
15 Khoury, M. J., Little, J., Gwinn, M. & Ioannidis, J. P. On the synthesis and interpretation of consistent but weak gene–disease associations in the era of genome-wide association studies. *Int. J. Epidemiol.* **36,** 439–445 (2007).
16 NCI-NHGRI Working Group on Replication in Association StudiesChanock, S. J. Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D. J. *et al.* Replicating genotype–phenotype associations. *Nature* **447,** 655–660 (2007).
17 Elbaz, A., Nelson, L. M., Payami, H., Ioannidis, J. P., Fiske, B. K., Annesi, G. *et al.* Lack of replication of thirteen single-nucleotide polymorphisms implicated in Parkinson's disease: a large-scale international study. *Lancet Neurol.* **5,** 917–923 (2006).
18 Munafo, M. R. & Flint, J. Meta-analysis of genetic association studies. *Trends Genet.* **20,** 439–444 (2004).
19 Lau, J., Ioannidis, J. P. & Schmid, C. H. Summing up evidence: one answer is not always enough. *Lancet* **351,** 123–127 (1998).
20 Ioannidis, J. P., Patsopoulos, N. A. & Evangelou, E. Heterogeneity in meta-analyses of genome-wide association investigations. *PLoS ONE* **2,** e841 (2007).
21 Sagoo, G. S., Little, J. & Higgins, J. P. Systematic reviews of genetic association studies. Human Genome Epidemiology Network. *PLoS Med.* **6,** e28 (2009).
22 Egger, M. & Smith, G. D. Bias in location and selection of studies. *BMJ* **316,** 61–66 (1998).
23 Lin, B. K., Clyne, M., Walsh, M., Gomez, O., Yu, W., Gwinn, M. *et al.* Tracking the epidemiology of human genes in the literature: the HuGE Published Literature database. *Am. J. Epidemiol.* **164,** 1–4 (2006).
24 Tang, J. L. Selection bias in meta-analyses of gene–disease associations. *PLoS Med.* **2,** e409 (2005).
25 Kavvoura, F. K. & Ioannidis, J. P. Methods for meta-analysis in genetic association studies: a review of their potential and pitfalls. *Hum. Genet.* **123,** 1–14 (2008).
26 Attia, J., Thakkinstian, A. & D'Este, C. Meta-analyses of molecular association studies: methodologic lessons for genetic epidemiology. *J. Clin. Epidemiol.* **56,** 297–303 (2003).
27 Begg, C. B. & Mazumdar, M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* **50,** 1088–1101 (1994).
28 Egger, M., Davey Smith, G., Schneider, M. & Minder, C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* **315,** 629–634 (1997).
29 Lau, J., Antman, E. M., Jimenez-Silva, J., Kupelnick, B., Mosteller, F. & Chalmers, T. C. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N. Engl. J. Med.* **327,** 248–254 (1992).
30 Ioannidis, J. P., Contopoulos-Ioannidis, D. G. & Lau, J. Recursive cumulative meta-analysis: a diagnostic for the evolution of total randomized evidence from group and individual patient data. *J. Clin. Epidemiol.* **52,** 281–291 (1999).
31 Ioannidis, J. & Lau, J. Evolution of treatment effects over time: empirical insight from recursive cumulative metaanalyses. *Proc. Natl Acad. Sci. USA* **98,** 831–836 (2001).
32 McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* **9,** 356–369 (2008).
33 Seminara, D., Khoury, M. J., O'Brien, T. R., Manolio, T., Gwinn, M. L., Little, J. *et al.* The emergence of networks in human genome epidemiology: challenges and opportunities. *Epidemiology* **18,** 1–8 (2007).
34 Ioannidis, J. P., Bernstein, J., Boffetta, P., Danesh, J., Dolan, S., Hartge, P. *et al.* A network of investigator networks in human genome epidemiology. *Am. J. Epidemiol.* **162,** 302–304 (2005).
35 Zeggini, E. & Ioannidis, J. P. Meta-analysis in genome-wide association studies. *Pharmacogenomics* **10,** 191–201 (2009).
36 Zeggini, E., Scott, L. J., Saxena, R., Voight, B. F., Marchini, J. L., Hu, T. *et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.* **40,** 638–645 (2008).
37 Wellcome Trust Case Control Consortium. Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. *Nature* **447,** 661–678 (2007).
38 Evangelou, E., Maraganore, D. M. & Ioannidis, J. P. Meta-analysis in genome-wide association datasets: strategies and application in Parkinson disease. *PLoS ONE* **2,** e196 (2007).
39 Barrett, J. C., Hansoul, S., Nicolae, D. L., Cho, J. H., Duerr, R. H., Rioux, J. D. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.* **40,** 955–962 (2008).
40 Browning, S. R. Missing data imputation and haplotype phase inference for genome-wide association studies. *Hum. Genet.* **124,** 439–450 (2008).
41 Hosking, L., Lumsden, S., Lewis, K., Yeo, A., McCarthy, L., Bansal, A. *et al.* Detection of genotyping errors by Hardy–Weinberg equilibrium testing. *Eur. J. Hum. Genet.* **12,** 395–399 (2004).
42 Cox, D. G. & Kraft, P. Quantification of the power of Hardy–Weinberg equilibrium testing to detect genotyping error. *Hum. Hered.* **61,** 10–14 (2006).
43 Wittke-Thompson, J. K., Pluzhnikov, A. & Cox, N. J. Rational inferences about departures from Hardy–Weinberg equilibrium. *Am. J. Hum. Genet.* **76,** 967–986 (2005).
44 Minelli, C., Thompson, J. R., Abrams, K. R., Thakkinstian, A. & Attia, J. How should we use information about HWE in the meta-analyses of genetic association studies? *Int. J. Epidemiol.* **37,** 136–146 (2008).
45 Zintzaras, E. & Lau, J. Synthesis of genetic association studies for pertinent gene–disease associations requires appropriate methodological and statistical approaches. *J. Clin. Epidemiol.* **61,** 634–645 (2008).
46 Thakkinstian, A., McElduff, P., D'Este, C., Duffy, D. & Attia, J. A method for meta-analysis of molecular association studies. *Stat. Med.* **24,** 1291–1306 (2005).
47 Salanti, G., Sanderson, S. & Higgins, J. P. Obstacles and opportunities in meta-analysis of genetic association studies. *Genet. Med.* **7,** 13–20 (2005).
48 Lindley, D. Statistical inference concerning Hardy–Weinberg equilibrium. *Bayesian Stat.* **3,** 307–326 (1988).
49 Weir, B. S. in *Genetic Data Analysis II: Methods for Discrete Population Genetic Data* (Sinauer Associates, Sunderland, 1996).
50 Hernandez, J. L. & Weir, B. S. A disequilibrium coefficient approach to Hardy–Weinberg testing. *Biometrics* **45,** 53–70 (1989).
51 Minelli, C., Thompson, J. R., Abrams, K. R., Thakkinstian, A. & Attia, J. The choice of a genetic model in the meta-analysis of molecular association studies. *Int. J. Epidemiol.* **34,** 1319–1328 (2005).
52 Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.* **22,** 719–748 (1959).
53 Yusuf, S., Peto, R., Lewis, J., Collins, R. & Sleight, P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Prog. Cardiovasc. Dis.* **27,** 335–371 (1985).
54 Cochran, W. G. The combination of estimates from different experiments. *Biometrics* **10,** 101–129 (1954).
55 Hardy, R. J. & Thompson, S. G. Detecting and describing heterogeneity in meta-analysis. *Stat. Med.* **17,** 841–856 (1998).
56 Petitti, D. B. Approaches to heterogeneity in meta-analysis. *Stat. Med.* **20,** 3625–3633 (2001).
57 DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Control Clin. Trials* **7,** 177–188 (1986).
58 Higgins, J. P. & Thompson, S. G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21,** 1539–1558 (2002).
59 Mittlbock, M. & Heinzl, H. A simulation study comparing properties of heterogeneity measures in meta-analyses. *Stat. Med.* **25,** 4321–4333 (2006).
60 Higgins, J. P., Thompson, S. G., Deeks, J. J. & Altman, D. G. Measuring inconsistency in meta-analyses. *BMJ* **327,** 557–560 (2003).
61 Ioannidis, J. P., Patsopoulos, N. A. & Evangelou, E. Uncertainty in heterogeneity estimates in meta-analyses. *BMJ* **335,** 914–916 (2007).
62 Cardon, L. R. & Palmer, L. J. Population stratification and spurious allelic association. *Lancet* **361,** 598–604 (2003).
63 Wacholder, S., Rothman, N. & Caporaso, N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *J. Natl. Cancer Inst.* **92,** 1151–1158 (2000).
64 Wacholder, S., Rothman, N. & Caporaso, N. Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. *Cancer Epidemiol. Biomarkers Prev.* **11,** 513–520 (2002).
65 Marchini, J., Cardon, L. R., Phillips, M. S. & Donnelly, P. The effects of human population structure on large genetic association studies. *Nat. Genet.* **36,** 512–517 (2004).

66 Thomas, D. C. & Witte, J. S. Point: population stratification: a problem for case–control studies of candidate-gene associations? *Cancer Epidemiol. Biomarkers Prev.* **11**, 505–512 (2002).
67 Ioannidis, J. P., Ntzani, E. E. & Trikalinos, T. A. 'Racial' differences in genetic effects for complex diseases. *Nat. Genet.* **36**, 1312–1318 (2004).
68 Garner, C. Upward bias in odds ratio estimates from genome-wide association studies. *Genet. Epidemiol.* **31**, 288–295 (2007).
69 Zollner, S. & Pritchard, J. K. Overcoming the winner's curse: estimating penetrance parameters from case–control data. *Am. J. Hum. Genet.* **80**, 605–615 (2007).
70 Ghosh, A., Zou, F. & Wright, F. A. Estimating odds ratios in genome scans: an approximate conditional likelihood approach. *Am. J. Hum. Genet.* **82**, 1064–1074 (2008).
71 Ioannidis, J. P. Why most discovered true associations are inflated. *Epidemiology* **19**, 640–648 (2008).
72 Kraft, P. Curses—winner's and otherwise—in genetic epidemiology. *Epidemiology* **19**, 649–651 (2008); discussion 657–658.
73 Yu, K., Chatterjee, N., Wheeler, W., Li, Q., Wang, S., Rothman, N. *et al.* Flexible design for following up positive findings. *Am. J. Hum. Genet.* **81**, 540–551 (2007).
74 Ioannidis, J. P., Thomas, G. & Daly, M. J. Validating, augmenting and refining genome-wide association signals. *Nat. Rev. Genet.* **10**, 318–329 (2009).
75 Zondervan, K. T. & Cardon, L. R. The complex interplay among factors that influence allelic association. *Nat. Rev. Genet.* **5**, 89–100 (2004).
76 Moonesinghe, R., Khoury, M. J., Liu, T. & Ioannidis, J. P. Required sample size and nonreplicability thresholds for heterogeneous genetic associations. *Proc. Natl Acad. Sci. USA* **105**, 617–622 (2008).
77 Scott, L. J., Mohlke, K. L., Bonnycastle, L. L., Willer, C. J., Li, Y., Duren, W. L. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345 (2007).
78 Hedges, L. V. & Pigott, T. D. The power of statistical tests in meta-analysis. *Psychol. Methods* **6**, 203–217 (2001).
79 Helgason, A., Palsson, S., Thorleifsson, G., Grant, S. F., Emilsson, V., Gunnarsdottir, S. *et al.* Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. *Nat. Genet.* **39**, 218–225 (2007).
80 Garcia-Closas, M., Hall, P., Nevanlinna, H., Pooley, K., Morrison, J., Richesson, D. A. *et al.* Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS Genet.* **4**, e1000054 (2008).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)

The Winner's Curse

Hirofumi Nakaoka, *Tokai University, Isehara, Kanagawa, Japan*

Ituro Inoue, *Tokai University, Isehara, Kanagawa, Japan*

Advanced article

Article Contents

Online posting date: 15th November 2010

Winners in competitive bidding are losers in that they frequently pay too high a price. This phenomenon has recently been noted in genetic association studies of common diseases. The winner's curse in genetic association studies appears as upward bias in the estimated effect of a newly identified allele on disease risk when the study design lacks sufficient statistical power. The winner's curse manifests mostly in genome-wide association (GWA) studies in which 300 000–1 000 000 single-nucleotide polymorphisms are tested. The winner's curse also occurs in meta-analysis of several GWA data sets. To counter this effect, construction of a large-scale GWA study or a consortium-based meta-analysis of GWA studies that is sufficiently powered to account for the presence of between-study heterogeneity is required.

Introduction

'Winner's curse' was originally identified at competitive auctions, in which winners paradoxically are losers in that they pay too much (Thaler, 1988). The phenomenon was first conceptualised in the context of competitive bidding for oil-exploration concessions in the 1970s (Capen *et al.*, 1971). In the competitive bidding environment, the estimated value of a parcel will be extremely diverse among bidders due to the difficulty of calculating the oil reserves. According to the actual sales values at the time, the highest bid prices were 5 to 10 times higher than the lowest bids. In the most extreme case, it was 109 times higher. Furthermore, Capen *et al.* showed that various competitors made such dissimilar bids even having the same basic data to assess the property value, implying a large uncertainty underlying the estimation of value. Whereas the average value of the parcels' estimates may be roughly correct, the individual estimation for a parcel will be either high or low. The bidder who estimates the larger reserves for a tract will bid the higher price. Capen *et al.* draw the following conclusions:

> "In competitive bidding, the winner tends to be the player who most overestimates true tract value" and "He who bids on a parcel what he thinks it is worth will, in the long run, be taken for a cleaning".

They also performed a simulation study based on mathematical modelling of competitive biddings and formulated three conditions for more conservative bidding:

> "The less information one has compared with what his opponents have, the lower he ought to bid", "The more uncertain one is about his value estimate, the lower he should bid", and "The more bidders (above three) that show up on a given parcel, the lower one should bid".

The third condition might seem counterintuitive: we tend to think when there are more bidders, a bid lesser than contemplated will not be fruitful. However, the third condition, like all three conditions, is meant to maximise the rate of return rather than the area owned. In fact, when we underestimate or correctly estimate the value, the tract is more likely to be won by the opponent who most overestimates. Thus, when there are many bidders, the winner is more likely to be one who overestimates the value. Indeed, Capen *et al.* demonstrated by simulation that the larger the number of opponents, the larger the degree of the overestimation of the value offered by the winner.

Recently, 'the winner's curse' is frequently mentioned in genetic association studies (Ioannidis *et al.*, 2001; Ioannidis, 2008; Kraft, 2008; Lohmueller *et al.*, 2003; Nakaoka and Inoue, 2009). Population-based genetic association studies are conducted to identify susceptibility genes underlying common diseases such as diabetes, schizophrenia and coronary artery disease. Most genetic association studies are based on case–control study design, in which cases with a disease of interest and unaffected controls are selected from

ELS subject area: Genetics and Disease

How to cite:
Nakaoka, Hirofumi; and Inoue, Ituro (November 2010) The Winner's Curse. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.
DOI: 10.1002/9780470015902.a0022495

a population and allele or genotype frequency distributions of a genetic variant are compared between affected and unaffected subjects. If a correlation between genotype and disease status is statistically significant, association between the variant and the disease can be claimed. In case–control study, the strength of the association is quantified as an odds ratio (OR). For example, an allelic OR of 1.5 indicates that each additional risk allele increases the risk of the disease 1.5-fold. The winner's curse phenomenon in this field occurs when a new gene–disease association is identified and the genetic effect on disease risk quantifying the strength of the association is overestimated (Ioannidis, 2008; Nakaoka and Inoue, 2009; Zollner and Pritchard, 2007).

Although competitive bidding and genetic association study differ in many ways, similar characteristics occur with regard to the winner's curse.

Genome-wide Association Study for Common Diseases

Most common diseases are thought to be influenced by multiple genes and environmental factors. Mapping of susceptibility genes in common diseases such as diabetes, cancers and coronary artery disease is a major challenge in medical genetics. The common disease-common variant (CDCV) hypothesis is the most influential theory, in which common diseases are attributable to common allelic variants, each having a modest effect (Lander, 1996; Reich and Lander, 2001). On this hypothesis, genetic variants contributing to common diseases can be identified by indirect linkage disequilibrium (LD) mapping (Kruglyak, 1999). Single-nucleotide polymorphisms (SNPs) in a genomic region with high LD are highly correlated with each other. When an SNP marker and true disease variant are in LD, the disease variant can be detected by examination of the association between the SNP marker and the disease status. **See also**: An Evolutionary Framework for Common Disease; Linkage Disequilibrium; Population History and Linkage Disequilibrium; On Sequence Variants that Influence the Risk of Common Diseases

In April 2003, the International Human Genome Sequencing Project provided a finished-grade human genome sequence (International Human Genome Sequencing Consortium, 2004; Lander *et al.*, 2001; Venter *et al.*, 2001). The HapMap project provided information about human genome variation, including the patterns of LD across the human genome (The International HapMap Consortium, 2003). This data allowed researchers to select a subset of SNPs (i.e. tag SNPs) that most efficiently surrogate the many other SNPs across the whole genome. By genotyping hundreds of thousands SNPs of the entire 3 billion base pair genome, there is a reasonable possibility that such genome-wide association (GWA) studies can identify susceptibility genes under the CDCV hypothesis (Hirschhorn and Daly, 2005; McCarthy *et al.*, 2008). To date, several hundreds of gene–disease associations have been reported (Hindorff *et al.*, 2009; Manolio *et al.*, 2008). A typical process of detecting GWA signals is shown in **Figure 1a**. **See also**: Genome-Wide Association Studies; Genome-wide Association Studies: The Success, Failure and Future; HapMap Project; Human Genome Project: Importance in Clinical Genetics

Stringent Genome-wide Significance Threshold and Statistical Power to Detect GWA Signal

We describe two methodological issues in GWA studies, each of which is relevant to the occurrence of the winner's curse phenomenon: the stringent genome-wide significance threshold and the statistical power to detect a GWA signal. In GWA studies, a large number of SNPs (ranging from 300 000 to 1 million) are statistically tested; therefore, the issue of multiple comparisons or multiple testing must be addressed. The use of the conventional P-value of 0.05 is expected to generate 25 000 false-positive associations by chance alone when 500 000 SNPs are tested. This flood of false-positive associations can be reduced by controlling the family-wise error rate (FWER), which is the probability of one or more significant associations under the null hypothesis of no association at all. The simplest correction of P-values for multiple comparisons is a Bonferroni correction. The genome-wide significance level based on Bonferroni correction is $\alpha' = \alpha/n$, where α is the FWER and n the number of SNPs tested. For 500 000 SNPs, the genome-wide significance level should therefore be set at 1.0×10^{-7}, corresponding to an FWER of 0.05. Such a correction is somewhat conservative because the tests are not mutually independent due to LD between SNPs. In practice, one influential GWA study used a relaxed genome-wide significance level of 5.0×10^{-7} (Wellcome Trust Case Control Consortium, 2007).

On the contrary, a stringent genome-wide significance threshold may lead to lack of the statistical power to detect gene–disease associations with modest effects on disease risk. Indeed, effect sizes of consistently replicated gene–disease associations have been found to be small (ORs = 1.2–1.5) in genetic association studies of common disease (Khoury *et al.*, 2007). As shown in **Figure 2**, more than 2000 cases and the same number of controls are needed to detect a disease-susceptibility allele with frequency of 0.30 and OR of 1.3 at the genome-wide significance level of 5.0×10^{-7}. Therefore, most single GWA studies are underpowered to detect small genetic effects.

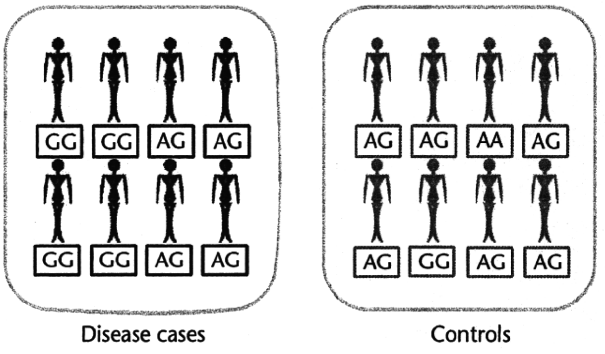
Consortium-based Meta-analysis of GWA Studies

Meta-analysis is a statistical tool to combine results from individual studies on the same research question.

DNA sample collection
Ascertainment based on disease status

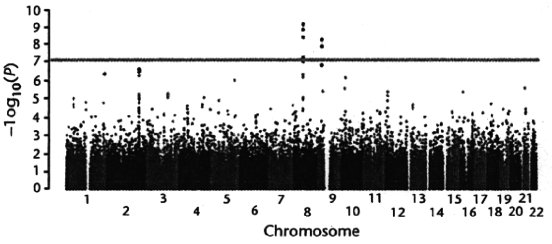
Genotyping
300 000~1 000 000 SNPs selected to surrogate the other SNPs across the whole genome

Statistical analysis: Case-control study
Allele frequency of each SNP is compared between cases and controls

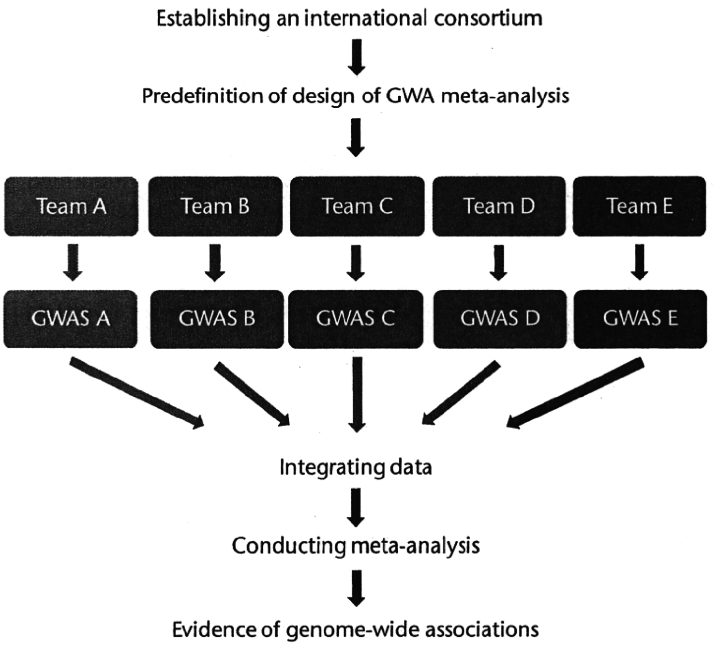


Detection of GWA signals
Select a subset of SNPs (e.g. $P < 5.0 \times 10^{-7}$)

Replication study
Associations are confirmed in independent samples



(a)



(b)

Figure 1 Flowchart of a typical genome-wide association (GWA) study (a) and a consortium-based meta-analysis of GWA studies (b).

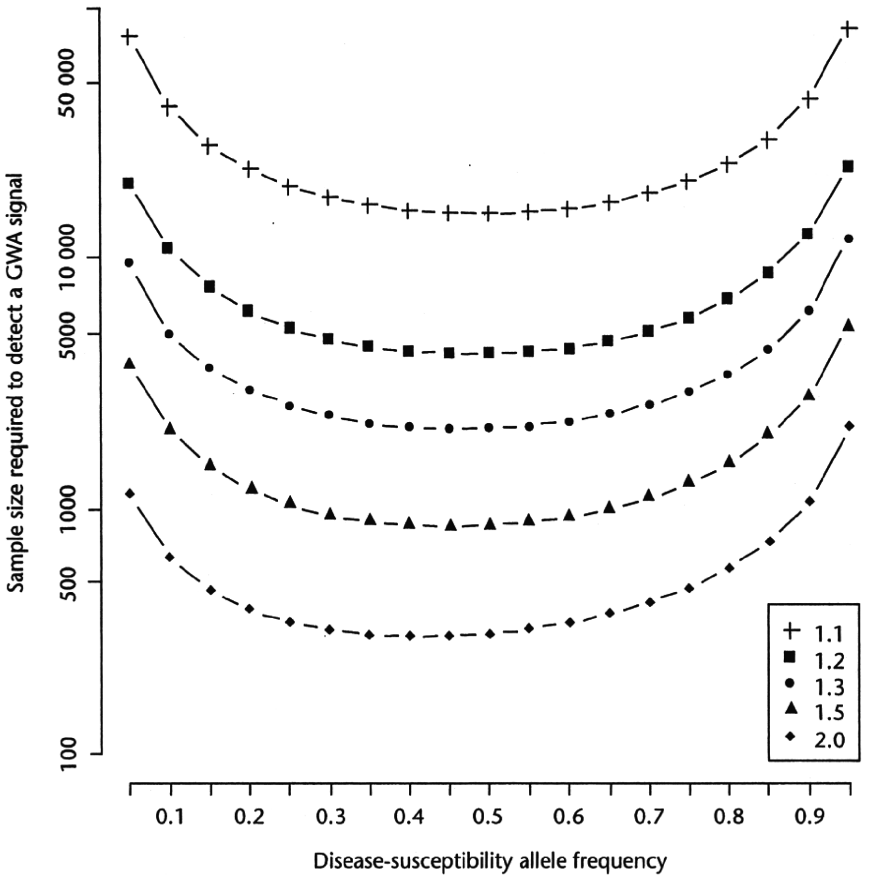


Figure 2 Sample sizes required to detect an association of per allele odds ratio (OR) of 1.1, 1.2, 1.3, 1.5 and 2.0 with power of 80% at the genome-wide significance level of 5.0×10^{-7} as a function of disease-susceptibility allele frequency ranging from 0.05 to 0.95.

Meta-analysis improves the estimation of summary OR and 95% confidence interval (CI) and increases the statistical power to detect gene–disease associations (Munafo and Flint, 2004). Consortium-based meta-analysis includes pooled analysis using prospective data collection or existing data by collaboration with a consortium of investigators in which definition of phenotype/outcome of interest, genotyping platform, quality control of genotype data and analytical method are determined by the consortium (schematic representation of meta-analysis of GWA data sets shown in **Figure 1b**; Ioannidis *et al.*, 2005; McCarthy *et al.*, 2008; Seminara *et al.*, 2007; Zeggini and Ioannidis, 2009). Furthermore, meta-analysis of GWA data sets can be performed when GWA data sets have been deposited in publicly available databases or when the authors of original GWA studies have provided ORs and 95% CIs for all genotyped SNPs. Indeed, integration of several GWA data sets has been designed and new susceptibility genes then discovered (Barrett *et al.*, 2008; Zeggini *et al.*, 2008). A meta-analysis of three GWA data sets for Crohn's disease is an astonishingly successful example of this procedure (Barrett *et al.*, 2008). The combined GWA data sets included 635 547 SNPs in 3230 cases and 4829 controls. The combined GWA data sets were used at the screening stage, and meta-analysis of the GWA data sets and additional replication data sets confirmed 11 previously reported loci and identified genome-wide significant signals for novel 21 loci.

Winner's Curse in Genetic Association Studies

Candidate gene approach era

Before the advent of GWA studies, the candidate gene approach was widely conducted. This approach examines the association between a particular allele of a gene that may be involved in a disease (i.e. a candidate gene) and disease status. Once a gene–disease association is reported, replication studies using independent populations are extensively implemented to establish the credibility of the initial positive finding. In the era of the candidate gene approach, however, most of the initial positive associations were not reproducible in the subsequent replication studies (Anonymous, 1999; Cardon and Bell, 2001; Colhoun *et al.*, 2003; Hirschhorn *et al.*, 2002; Ioannidis *et al.*, 2001; Lohmueller *et al.*, 2003).

A number of reasons have been advanced for the inconsistency of the results of various studies on the same research question (Colhoun *et al.*, 2003; Hirschhorn *et al.*, 2002; Ioannidis, 2007; Salanti *et al.*, 2005). False-positive results in the initial studies and false-negative results in the small replication studies are implicated as the most likely reasons for the lack of replication (Colhoun *et al.*, 2003; Hirschhorn *et al.*, 2002; Ioannidis *et al.*, 2001; Ioannidis, 2007; Lohmueller *et al.*, 2003). Indeed, Hirschhorn and colleagues retrieved 166 genetic associations by at least three separate publications and showed that only six associations reached statistical significance consistently, suggesting that a large number of original findings were false-positive reports. At the same time, they also showed that 97 of the 166 significant associations were reproduced in at least 1 separate replication, suggesting another possibility that the underlying genetic effects were moderate and that most of the replication studies were underpowered to detect them. Furthermore, inconsistency and between-study heterogeneity may be due to biases (e.g. population stratification, misclassification of clinical outcome, genotyping error and overestimation of genetic effect) or genuine differences in the genetic effects across populations (e.g. differences in pattern of LD structure over chromosomal regions of interest across populations and the presence of gene–gene or gene–environment interaction). Several review articles have addressed this issue (Ioannidis, 2007; Nakaoka and Inoue, 2009; Salanti *et al.*, 2005).

In the process of probing the cause of inconsistency, the winner's curse phenomenon has been identified as being partly responsible. Ioannidis *et al.* conducted a systematic review of 36 meta-analyses including a total of 370 genetic association studies (Ioannidis *et al.*, 2001). Statistically significant between-study heterogeneity was observed in 14 meta-analyses. Restricting to meta-analyses with at least 15 studies, 7 of 9 of these showed significant heterogeneity. In 25 meta-analyses, the first study showed a more predisposing or protective OR than subsequent replication studies. Using cumulative meta-analysis plots of eight gene–disease associations, the authors were able to depict a process by which strong associations claimed in the first study regressed toward null associations as subsequent replication studies were accumulated over time. Intriguingly, the initial studies for these associations were published in prestigious journals (of the eight associations, five were in journals with an impact factor higher than 9). Additionally, it is noteworthy that these associations showed extremely impressive effect sizes at discovery (ORs: 2.1–5.7 for disease predisposing alleles and 0.03–0.22 for protective alleles). Similar findings were reported in Lohmueller *et al.* (2003). These data suggest that associations passing predetermined thresholds of statistical significance tend to cause overestimation of the size of the genetic effect on disease risk, especially when the sample size of the study is small and the threshold is stringent in multiple testing situations (Ioannidis *et al.*, 2001; Ioannidis, 2008; Kraft, 2008; Lohmueller *et al.*, 2003; Nakaoka and Inoue, 2009). Such upward bias is the winner's curse phenomenon appearing in genetic association studies.

GWA study era

Theoretical considerations on the winner's curse phenomenon in GWA studies of common disease have been constructed (Garner, 2007; Ghosh *et al.*, 2008; Nakaoka and Inoue, 2009; Pereira *et al.*, 2009; Yu *et al.*, 2007; Zollner and Pritchard, 2007). These indicate that the effect sizes of initially identified associations are expected to be upwardly biased when the study design lacks sufficient statistical power. Such conditions have become ever more frequent in GWA settings. As described earlier, the unprecedented number of comparisons being made in GWA studies imposes an extremely stringent significance threshold (e.g. $P < 5.0 \times 10^{-7}$) to control the FWER. As there is a trade-off between type I and type II errors, the stringent significance threshold can be achieved only at the expense of the statistical power. Given that the effect sizes of gene–disease associations are originally small (ORs = 1.2–1.5), most single GWA studies may be underpowered. Furthermore, there is a possibility that the winner's curse also occurs in meta-analysis of several GWA data sets that is conducted to improve the power of GWA study (Nakaoka and Inoue, 2009; Pereira *et al.*, 2009).

Theoretical considerations on the winner's curse phenomenon in GWA studies

We first describe the possible cause of the winner's curse phenomenon in GWA studies and then use mathematical modelling. We also present the results of our simulation study to illustrate the winner's curse in meta-analyses of GWA data sets (Nakaoka and Inoue, 2009).

Assuming independent k GWA studies with the power of $(1-\beta)$ to detect a gene–disease association, the number of GWA studies achieving genome-wide evidence of association follows binomial distribution $B(k, 1-\beta)$. In this case, $k(1-\beta)$ studies are expected to show significant evidence. Each study has an estimated OR for the association. The estimated ORs have 'uncertainty' and therefore distribute around the true OR. Although these estimated ORs may be accurate on average, some take large values and others small values. The independent k GWA studies with the same study design can be regarded as repeated data realisations from the same population. In repeated data realisation of large sample size, the standard error of estimated log OR ($\mathrm{se}(\hat{\theta})$) does not vary remarkably; therefore, test statistic ($Z = \hat{\theta}/\mathrm{se}(\hat{\theta})$) for each study depends largely on its effect size ($\hat{\theta}$, the natural logarithm of the OR; Garner, 2007; Ghosh *et al.*, 2008). The observed effect sizes from the studies passing the genome-wide significance threshold are then larger than those from nonsignificant studies and the average of all studies. Because the estimated ORs are expected to be accurate on average, the effect sizes from the studies showing significant associations are expected to be upwardly biased.

The preceding description can be represented through a simple mathematical model. Let θ and $\hat{\theta}$ be the true and estimated log OR, respectively, and $\mathrm{se}(\hat{\theta})$ be the standard error of $\hat{\theta}$, test statistic of test for the genetic effect is given by $Z=\hat{\theta}/\mathrm{se}(\hat{\theta})$. Under the null hypothesis, $H_0{:}\theta=0$, Z follows a standard normal distribution $N(0,1)$. By denoting $\mu=\theta/\mathrm{se}(\hat{\theta})$, we can express $Z\sim N(\mu,1)$ and then $Z-\mu$ follows $N(0,1)$. When the absolute value of Z surpasses a predetermined genome-wide significance threshold $|Z|>c$, we have a genome-wide significant signal. Here, we can consider the conditional probability density function (PDF) of Z given that the corresponding association passes the genome-wide significance threshold:

$$p(Z=z||Z|>c)=\frac{p(z)}{p(|Z|>c)}=\frac{\phi(z-\mu)}{\int_{-\infty}^{-c-\mu}\phi(x)\mathrm{d}x+\int_{c-\mu}^{\infty}\phi(x)\mathrm{d}x}$$
$$=\frac{\phi(z-\mu)}{\Phi(-c-\mu)+\Phi(-c+\mu)} \quad [1]$$

where ϕ and Φ are the standard normal and standard normal cumulative distributions, respectively, and $\Phi(-c-\mu)+\Phi(-c+\mu)$ the normalisation constant. We set the genome-wide significance threshold at $c=5.0$, corresponding to $P<5.7\times10^{-7}$. We assume $\mu=4.3$, corresponding to the power to detect the association at the significance level is 0.24. That is, we assume that the study does not have an optimal power. Under the parameter settings, the two PDFs, $\phi(z)$ and $\phi(z-\mu)$, and the conditional PDF, $p(Z=z||Z|>c)$, are displayed in **Figure 3a**. In this case, the conditional PDF is a truncated distribution as depicted by green solid curve. The expected value of truncated random variable Z following the eqn [1] is given by Ghosh *et al.* (2008):

$$E(Z||Z|>c)=\frac{\int_{-\infty}^{-c}z\phi(z-\mu)dz+\int_{c}^{\infty}z\phi(z-\mu)dz}{\Phi(-c-\mu)+\Phi(-c+\mu)}$$
$$=\mu+\frac{\phi(c-\mu)-\phi(c+\mu)}{\Phi(-c-\mu)+\Phi(-c+\mu)} \quad [2]$$

This equation indicates that $E(Z||Z|>c)$ is larger than μ for $\mu>0$ (i.e. $\theta>0$), and vice versa, $E(Z||Z|>c)$ is smaller than μ for $\mu<0$. Thus, we can say that the expected value of effect size ($\hat{\theta}$) is exaggerated given that the associations pass predetermined thresholds of statistical significance. Furthermore, the degree of upward bias in effect size due to the winner's curse is larger especially when the power to detect the association is low. This can be explained by a simple graphical representation. **Figure 3b** shows the bias in test statistic due to the restriction in range of $|Z|>c$ as a function of μ:

$$g(\mu)=E(Z||Z|>c)-\mu$$
$$=\frac{\phi(c-\mu)-\phi(c+\mu)}{\Phi(-c-\mu)+\Phi(-c+\mu)} \quad [3]$$

As shown in the eqns [2] and [3], the test statistic is unbiased and takes the value of zero for $\mu=0$. For $\mu\neq0$, the absolute value of bias in the test statistic decreases as the absolute value of μ increases. Given $\mu=\theta/\mathrm{se}(\hat{\theta})$, as the true effect size of the association or the sample size gets larger, the test statistic is not influenced by the winner's curse. At the same time, the bias becomes much smaller when the significance threshold is relaxed (blue and black solid lines in **Figure 3b** correspond to $P<5.7\times10^{-7}$ and $P<0.05$, respectively).

We present results of a simulation study to illustrate the possibility of the winner's curse phenomenon in meta-analysis of several GWA data sets (Nakaoka and Inoue, 2009). We simulate a consortium-based meta-analysis of several GWA data sets (shown in **Figure 1b**). As described in the previous section, meta-analysis of GWA data sets is a powerful approach to detect gene–disease associations. In GWA settings, the presence of between-study heterogeneity is likely to be common (e.g. population stratification, different patterns of LD structures across populations and genotyping errors). We demonstrate through a simulation study that the power of meta-analysis of GWA data sets to detect a small genetic effect that would decrease due to between-study heterogeneity, and as a result, the estimated effect sizes of the discovered gene–disease association would be inflated.

We considered five simple scenarios of meta-analyses (**Table 1**). For each scenario, we generated k independent genetic association studies, where effect sizes for the studies (θ_i) followed $\theta_i\sim N(\theta,\tau^2)$ so that the effect sizes were heterogeneous across studies. Again, effect size is taken as the natural logarithm of the OR. The τ^2 was the between-study variance, which determines the degree of variability in the effect sizes across studies. Given the number of affected and unaffected subjects in the studies, we can derive the genotypes of individual subjects given disease status. $\hat{\theta}_i$ was calculated based on the observed genotype counts within studies. $\hat{\theta}_i$s from k studies were combined to estimate the summary effect size ($\hat{\theta}$) by using the DerSimonian and Laird random effects model (REM) (DerSimonian and Laird, 1986). REM assumes that the genetic effects may vary across studies due to genuine difference or differential biases. Test of association was examined for each simulation run. The simulations were repeated 100 000 times. The power of detecting a gene–disease association was evaluated as the proportion of simulation runs reaching the significance level of 5.7×10^{-7}, corresponding to $c=5.0$. To evaluate the effect of between-study heterogeneity on the conclusion of the meta-analysis, we performed simulations under the conditions that the between-study variance (τ^2) varied from 0.0 to 0.02 with increments of 0.001. The values of τ^2 were based on the literature values reported in Moonesinghe *et al.* (2008) for the confirmed 10 loci in a meta-analysis of three GWA studies of type 2 diabetes (Scott *et al.*, 2007). Thus, the simulation should reflect the possible range of between-study variance.

When the true overall OR = 1.4 (i.e. $\exp(\theta)=1.4$) under the dominant model, the power for each scenario gradually decreases as the degree of between-study heterogeneity (τ^2)

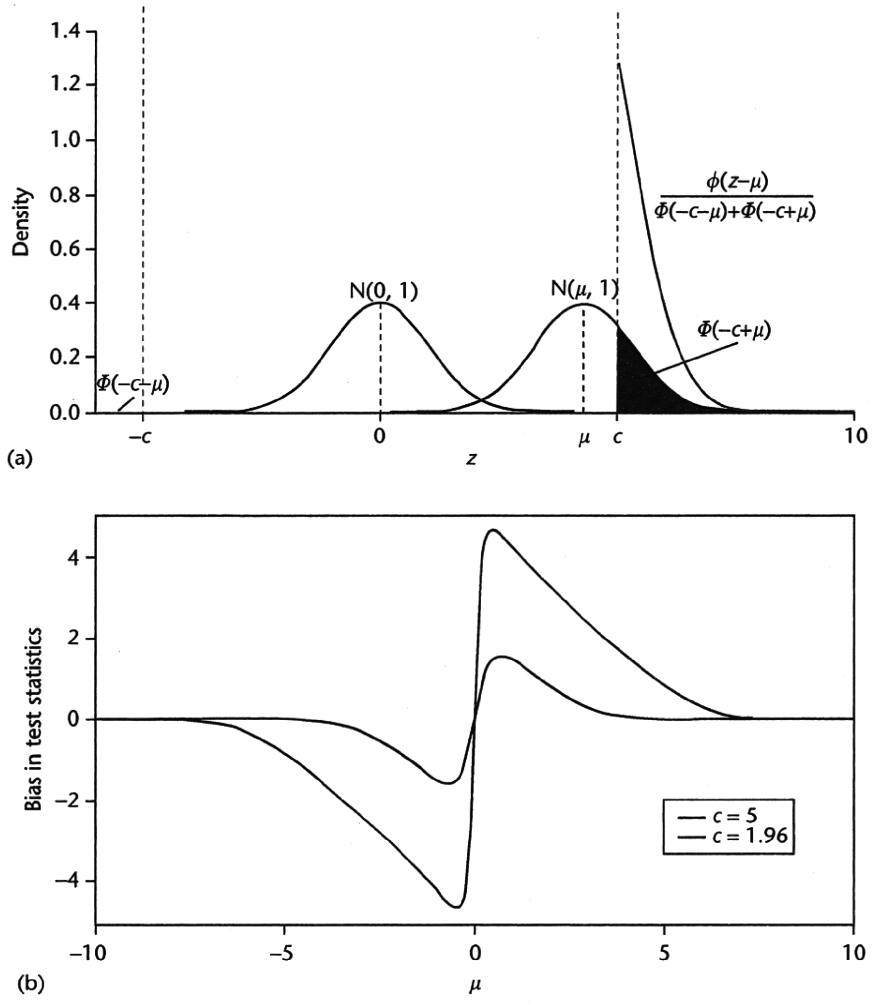


Figure 3 Graphical representation of theoretical consideration on the winner's curse using mathematical modelling. (a) Three PDFs of test statistic Z: black curve shows Z under the null hypothesis of no association; blue curve shows Z under the assumption that $\mu = 4.3$ corresponding to the power to detect the association at the significance level ($P < 5.7 \times 10^{-7}$) is 0.24 and green curve shows Z conditional on $|Z| > c$. The area highlighted in blue shows $p(|Z| > c)$. For visualisation, we set c and μ as 5.0 and 4.3, respectively. (b) Bias in test statistic due to the restriction in range of $|Z| > c$ as a function of μ.

Table 1 Description of five simulation scenarios of meta-analysis, where k denotes the number of included studies and n_{case} and $n_{control}$ are the number of cases and controls within each study, respectively

Scenario	k	n_{case}	$n_{control}$
I	5	500	500
II	10	500	500
III	20	500	500
IV	10	1000	1000
V	5	2000	2000

Source: Adapted from Nakaoka and Inoue (2009), with permission from Nature Publishing Group.

increases (**Figure 4a**). Comparison among scenarios III, IV and V shows the decreases in the power for the same τ^2 to be larger in the scenarios with large sample size per study. When the overall OR is set at 2.0, the powers did not decrease so much in the simulated range of τ^2. Additionally, we calculated the mean ORs of the simulations passing the genome-wide significance threshold (P value $< 5.7 \times 10^{-7}$). The estimates of mean ORs were upwardly biased especially in scenarios in which power of detecting gene–disease association was low (**Figure 4b**). On the contrary, if the meta-analyses were sufficiently powered (e.g. true overall OR = 2.0), upward biases were not so pronounced in the simulated range of τ^2.

Figure 5 shows that the empirical frequency distribution of the estimated summary ORs of the simulation runs passing the genome-wide significance threshold according to τ^2. The true summary OR was set to be 1.4 under a dominant model with disease allele frequency of 0.30. Results from simulation scenario I are presented. The figure depicts several characteristics of the winner's curse in meta-analysis of GWA data sets. First, the frequency distributions are truncated, indicating that evidence of association can be achieved only when a certain degree of large

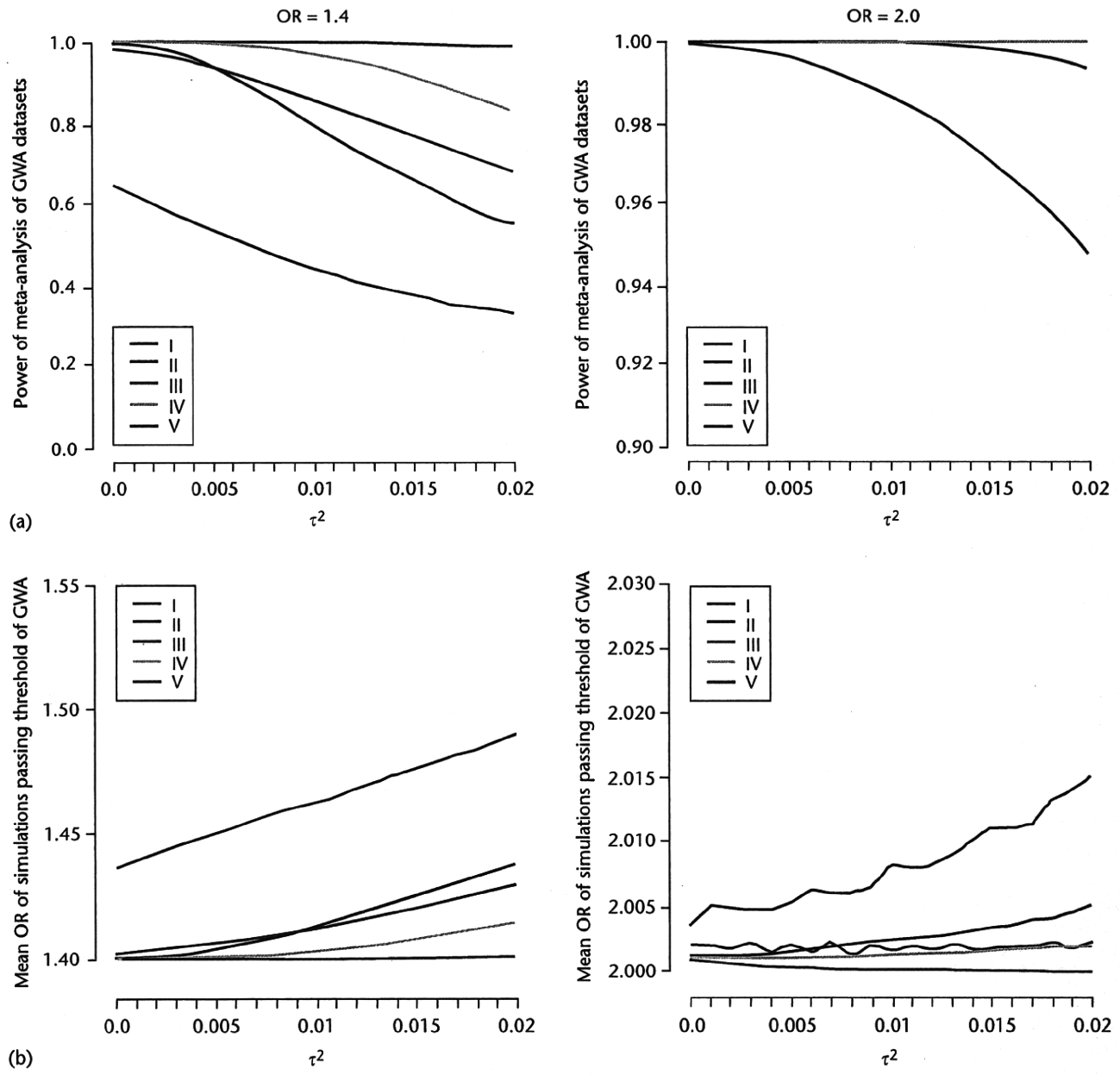


Figure 4 Simulations for (a) the powers in random effects model (REM) meta-analyses of detecting a gene–disease association at the significance level of 5.7×10^{-7} and (b) the mean OR of the simulations passing the threshold as the function of τ^2, disease allele frequency $f_A = 0.3$, the overall OR = 1.4 or 2.0. When overall OR = 2.0, the lines of the powers for scenarios II, III and IV are overlapping. The description of each simulation scenario is in Table 1. The between-study variance (τ^2) varied from 0.0 to 0.02 with increments of 0.001. Adapted from Nakaoka and Inoue (2009), with permission from Nature Publishing Group.

effect size occurs. This finding is consistent with the mathematical model described in the previous section, where the conditional PDF of Z given that the corresponding association passes the genome-wide significance threshold is truncated. Second, as the between-study heterogeneity (τ^2) increases, the standard error of the estimated summary OR increases and the numbers of simulation runs passing the threshold (or the power) decreases, and, at the same time, the modes of the distributions gradually shift upward. For example, the frequency distributions of $\tau^2 = 0.0$ and 0.02 reach their peaks at OR = 1.41 and 1.47, respectively. This is because the estimated summary OR must be enlarged to make up for the increase in the standard error attributable to between-study heterogeneity to achieve the genome-wide significance threshold.

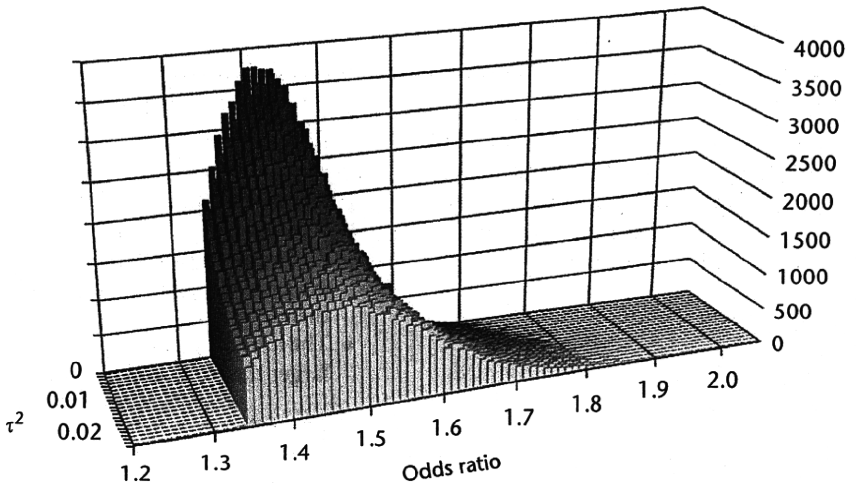


Figure 5 Frequency distribution of estimated summary odds ratios (ORs) of simulated meta-analyses passing the genome-wide significance threshold in simulation scenario I in **Table 1**. The true OR is assumed to be 1.4 under dominant model with disease susceptibility allele frequency of 0.30.

Conclusions

It is well known that the winners in an auction may be losers in that they pay too much for the value obtained. We have illustrated the characteristics of the winner's curse phenomenon in genetic association studies of common diseases. Although these two environments differ, similar characteristics manifest the winner's curse phenomenon (summarised in **Table 2**). In competitive bidding, the degree of the overestimation of the value of an item offered is larger when the number of opponents is greater or the amount of information regarding the value is limited. We should therefore bid more conservatively under such conditions. An extremely large number of explanatory variables (i.e. SNPs) in genetic association studies play a similar role to the number of opponents in auctions. In the era of GWA studies, 300 000–1 000 000 SNPs are tested for association for a disease, and a stringent statistical significance threshold must be imposed to avoid false-positive findings. As a result, the power to detect weak gene–disease associations at the threshold becomes suboptimal, and evidence of association can be achieved only when a certain degree of large effect size is observed, reminiscent of overestimations of value in competitive bidding. The sample size required for confirmatory study is underestimated when calculated based on this exaggerated genetic effect, and replication study therefore fails to corroborate the association. In addition, the lower the power of the original study, the larger the bias in the resulting estimate of the genetic effect.

To avoid the winner's curse, large-scale GWA studies with optimal power must be constructed. Recently, several theoretical approaches have been offered to reduce the bias in the estimates of the genetic effect of a discovered association (Garner, 2007; Ghosh *et al.*, 2008; Yu *et al.*, 2007; Zollner and Pritchard, 2007). These methodologies are helpful to calculate the required sample sizes to replicate the discovered associations (Yu *et al.*, 2007). In practice, replication studies on a large scale are necessary to establish the credibility of discovered findings.

Because most single GWA studies have been underpowered for the detection of small genetic effects on disease risk, consortium-based meta-analyses of GWA data sets have been designed. Although meta-analysis of GWA data sets is the most powerful approach to detect gene–disease associations, there is nevertheless a possibility that the power of a GWA meta-analysis decreases and the winner's curse reappears due to the presence of between-study heterogeneity. Simulation studies show that the presence of between-study heterogeneity can decrease power, skewing the association signals reaching genome-wide significance to be upwardly biased. Thus, a study design of consortium-based meta-analysis of GWA data sets accounting for the presence of between-study heterogeneity must be constructed and further data accumulation obtained after initial findings from GWA meta-analyses to estimate unbiased effect sizes of the discovered gene–disease associations.

GWA studies are now beginning to unravel the genetic architecture underlying complex diseases, and recent successful discoveries of genetic variations associated with diseases are valuable resources. The winner's curse in genetic association studies does not diminish the true value of genetic variations identified by GWA approaches. Indeed, only a fraction of the potential heredity risk of common disease has been accounted for by SNPs so far. Further exploration of other types of genomic variation such as rare and low-frequency single nucleotide changes and insertions and deletions of nucleotides that may explain additional disease risk is also promising.

Table 2 One-on-one relationships between competitive bidding and genetic association study with regard to the winner's curse phenomenon

	Competitive bidding	Genetic association study
Phenomenon	• Winners in competitive bidding are losers in that they frequently pay too high a price • The larger the number of opponents, the larger the degree of the overestimation of the value offered by the winner	• Effect size of a newly identified gene–disease association is overestimated when the study design lacks sufficient statistical power • The sample size required for confirmatory study is underestimated when calculated based on this exaggerated genetic effect, and replication study therefore fails to corroborate the association • The lower the statistical power of the original study, the larger the overestimation of the genetic effect
Background	• The estimated value of an item will be extremely diverse among bidders due to the difficulty of calculating its true value	• The unprecedented number of comparisons being made in GWA studies imposes an extremely stringent significance threshold to control the FWER. As a result, most single GWA studies may be underpowered • The power of meta-analysis of GWA data sets to detect a small genetic effect would decrease in the presence of between-study heterogeneity
Cause	• The winner tends to be the player who most overestimates the value of the tract	• Under the condition that the study design lacks optimal power, evidence of association can be achieved only when a large genetic effect is observed
Ways to overcome	• We should bid more conservatively if we have the limited amount of information compared to our opponents, if we feel anxiety over our estimated value, or if there are more bidders	• In the discovery phase, construction of a larger scale GWA study or a consortium-based meta-analysis of GWA studies that evaluates between-study heterogeneity is required to reduce probability of occurrence of the winner's curse • In the replication phase, methodologies for reducing bias in the estimates of genetic effect are helpful to calculate the sample sizes required to replicate the discovered associations

FWER, family-wise error rate; GWA, genome-wide association.

References

Anonymous (1999) Freely associating. *Nature Genetics* **22**: 1–2.

Barrett JC, Hansoul S, Nicolae DL *et al.* (2008) Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nature Genetics* **40**: 955–962.

Capen EC, Clapp RV and Campbell WM (1971) Competitive bidding in high-risk situations. *Journal of Petroleum Technology* **23**: 641–653.

Cardon LR and Bell JI (2001) Association study designs for complex diseases. *Nature Reviews. Genetics* **2**: 91–99.

Colhoun HM, McKeigue PM and Davey Smith G (2003) Problems of reporting genetic associations with complex outcomes. *Lancet* **361**: 865–872.

DerSimonian R and Laird N (1986) Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**: 177–188.

Garner C (2007) Upward bias in odds ratio estimates from genome-wide association studies. *Genetic Epidemiology* **31**: 288–295.

Ghosh A, Zou F and Wright FA (2008) Estimating odds ratios in genome scans: an approximate conditional likelihood approach. *American Journal of Human Genetics* **82**: 1064–1074.

Hindorff LA, Sethupathy P, Junkins HA *et al.* (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the USA* **106**: 9362–9367.

Hirschhorn JN and Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. *Nature Reviews. Genetics* **6**: 95–108.

Hirschhorn JN, Lohmueller K, Byrne E and Hirschhorn K (2002) A comprehensive review of genetic association studies. *Genetics in Medicine* **4**: 45–61.

International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* **431**: 931–945.

Ioannidis JP (2007) Non-replication and inconsistency in the genome-wide association setting. *Human Heredity* **64**: 203–213.

Ioannidis JP (2008) Why most discovered true associations are inflated. *Epidemiology* **19**: 640–648.

Ioannidis JP, Bernstein J, Boffetta P *et al.* (2005) A network of investigator networks in human genome epidemiology. *American Journal of Epidemiology* **162**: 302–304.

Ioannidis JP, Ntzani EE, Trikalinos TA and Contopoulos-Ioannidis DG (2001) Replication validity of genetic association studies. *Nature Genetics* **29**: 306–309.

Khoury MJ, Little J, Gwinn M and Ioannidis JP (2007) On the synthesis and interpretation of consistent but weak gene-disease associations in the era of genome-wide association studies. *International Journal of Epidemiology* **36**: 439–445.

Kraft P (2008) Curses–winner's and otherwise–in genetic epidemiology. *Epidemiology* **19**: 649–651, discussion 657–8.

Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genetics* **22**: 139–144.

Lander ES (1996) The new genomics: global views of biology. *Science* **274**: 536–539.

Lander ES, Linton LM, Birren B *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.

Lohmueller KE, Pearce CL, Pike M, Lander ES and Hirschhorn JN (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nature Genetics* **33**: 177–182.

Manolio TA, Brooks LD and Collins FS (2008) A HapMap harvest of insights into the genetics of common disease. *Journal of Clinical Investigation* **118**: 1590–1605.

McCarthy MI, Abecasis GR, Cardon LR *et al.* (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews. Genetics* **9**: 356–369.

Moonesinghe R, Khoury MJ, Liu T and Ioannidis JP (2008) Required sample size and nonreplicability thresholds for heterogeneous genetic associations. *Proceedings of the National Academy of Sciences of the USA* **105**: 617–622.

Munafo MR and Flint J (2004) Meta-analysis of genetic association studies. *Trends in Genetics* **20**: 439–444.

Nakaoka H and Inoue I (2009) Meta-analysis of genetic association studies: methodologies, between-study heterogeneity and winner's curse. *Journal of Human Genetics* **54**: 615–623.

Pereira TV, Patsopoulos NA, Salanti G and Ioannidis JP (2009) Discovery properties of genome-wide association signals from cumulatively combined data sets. *American Journal of Epidemiology* **170**: 1197–1206.

Reich DE and Lander ES (2001) On the allelic spectrum of human disease. *Trends in Genetics* **17**: 502–510.

Salanti G, Sanderson S and Higgins JP (2005) Obstacles and opportunities in meta-analysis of genetic association studies. *Genetics in Medicine* **7**: 13–20.

Scott LJ, Mohlke KL, Bonnycastle LL *et al.* (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**: 1341–1345.

Seminara D, Khoury MJ, O'Brien TR *et al.* (2007) The emergence of networks in human genome epidemiology: challenges and opportunities. *Epidemiology* **18**: 1–8.

Thaler R (1988) Anomalies: the winner's curse. *Journal of Economic Perspectives* **2**: 191–202.

The International HapMap Consortium (2003) The International HapMap Project. *Nature* **426**: 789–796.

Venter JC, Adams MD, Myers EW *et al.* (2001) The sequence of the human genome. *Science* **291**: 1304–1351.

Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. *Nature* **447**: 661–678.

Yu K, Chatterjee N, Wheeler W *et al.* (2007) Flexible design for following up positive findings. *American Journal of Human Genetics* **81**: 540–551.

Zeggini E and Ioannidis JP (2009) Meta-analysis in genome-wide association studies. *Pharmacogenomics* **10**: 191–201.

Zeggini E, Scott LJ, Saxena R *et al.* (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature Genetics* **40**: 638–645.

Zollner S and Pritchard JK (2007) Overcoming the winner's curse: estimating penetrance parameters from case-control data. *American Journal of Human Genetics* **80**: 605–615.

Stroke
JOURNAL OF THE AMERICAN HEART ASSOCIATION

American Stroke Association
A Division of American Heart Association

Differential Effects of Chromosome 9p21 Variation on Subphenotypes of Intracranial Aneurysm: Site Distribution

Hirofumi Nakaoka, Tomoko Takahashi, Koichi Akiyama, Tailin Cui, Atsushi Tajima, Boris Krischek, Hidetoshi Kasuya, Akira Hata and Ituro Inoue

Stroke 2010;41;1593-1598; originally published online Jul 1, 2010;
DOI: 10.1161/STROKEAHA.110.586529
Stroke is published by the American Heart Association. 7272 Greenville Avenue, Dallas, TX 72514
 Print ISSN: 0039-2499. Online ISSN: 1524-4628

Differential Effects of Chromosome 9p21 Variation on Subphenotypes of Intracranial Aneurysm

Site Distribution

Hirofumi Nakaoka, PhD; Tomoko Takahashi, PhD; Koichi Akiyama, PhD; Tailin Cui, MD; Atsushi Tajima, PhD; Boris Krischek, MD; Hidetoshi Kasuya, MD; Akira Hata, MD; Ituro Inoue, MD

Background and Purpose—Recently, a genome-wide association study identified associations between single nucleotide polymorphisms on chromosome 9p21 and risk of harboring intracranial aneurysm (IA). Aneurysm characteristics or subphenotypes of IAs, such as history of subarachnoid hemorrhage, presence of multiple IAs and location of IAs, are clinically important. We investigated whether the association between 9p21 variation and risk of IA varied among these subphenotypes.

Methods—We conducted a case-control study of 981 cases and 699 controls in Japanese. Four single nucleotide polymorphisms tagging the 9p21 risk locus were genotyped. The OR and 95% CI were estimated using logistic regression analyses.

Results—Among the 4 single nucleotide polymorphisms, rs1333040 showed the strongest evidence of association with IA ($P=1.5\times10^{-6}$; per allele OR, 1.43; 95% CI, 1.24–1.66). None of the patient characteristics (gender, age, smoking, and hypertension) was a significant confounder or effect modifier of the association. Subgroup analyses of IA subphenotypes showed that among the most common sites of IAs, the association was strongest for IAs of the posterior communicating artery (OR, 1.69; 95% CI, 1.26–2.26) and not significant for IAs in the anterior communicating artery (OR, 1.22; 95% CI, 0.96–1.57). When dichotomizing IA sites, the association was stronger for IAs of the posterior circulation–posterior communicating artery group (OR, 1.73; 95% CI, 1.32–2.26) vs the anterior circulation group (OR, 1.28; 95% CI, 1.07–1.53). Heterogeneity in these ORs was significant ($P=0.032$). The associations did not vary when stratifying by history of subarachnoid hemorrhage (OR, 1.42; 95% CI, 1.18–1.71 for ruptured IA; OR, 1.27; 95% CI, 1.00–1.62 for unruptured IA) or by multiplicity of IA (OR, 1.57; 95% CI, 1.21–2.03 for multiple IAs; OR, 1.36; 95% CI, 1.15–1.61 for single IA).

Conclusions—Our results suggest that genetic influence on formation may vary between IA subphenotypes. (***Stroke*. 2010;41:1593-1598.**)

Key Words: genetics ■ intracranial aneurysm ■ posterior circulation ■ subarachnoid hemorrhage

The rupture of an intracranial aneurysm (IA) is the most common cause of subarachnoid hemorrhage (SAH) associated with high morbidity and mortality. IA is implicated to be a multifactorial disease in which multiple genes and environmental factors influence disease risk.[1] Recently, a multistage genome-wide association study of Finnish, Dutch, and Japanese cohorts identified associations between single nucleotide polymorphisms (SNPs) on chromosomes 2q33, 8q11, and 9p21 and the susceptibility to IA.[2] Among the three loci, the 9p21 locus seems most intriguing because SNPs identified by the genome-wide association study of IA are in linkage disequilibrium with SNPs associated with several other arterial diseases (coronary artery disease, abdominal aortic aneurysm, and peripheral arterial disease),[3] and the 9p21 risk locus overlaps a newly annotated noncoding RNA, called *ANRIL* (or *CDKN2BAS*).[4]

Patient characteristics such as gender, age, smoking habit, and history of hypertension are well-established risk factors affecting formation, growth, and rupture of IA.[1] Furthermore, aneurysm characteristics such as the history of SAH, presence of multiple IAs, and location of IA (hereinafter, referred to as "subphenotypes" of IA) are clinically important because they are associated with the outcome of patients with IA.[5,6] Genetic dissection of IA subphenotypes is essential but uncharted territory.[7]

We conducted a case-control study in the Japanese population to evaluate the possibility of confounding or effect modification of the 9p21 association through the environmen-

Received April 16, 2010; accepted May 13, 2010.

From Division of Molecular Life Science (H.N., T.T., K.A., T.C., A.T., I.I.), School of Medicine, Tokai University, Kanagawa, Japan; The Japan Health Sciences Foundation (H.N.), Tokyo, Japan; Department of Neurosurgery (B.K.), University of Tübingen, Tübingen, Germany; Division of Neurosurgery (H.K.), Medical Center East, Tokyo Women's Medical University, Tokyo, Japan; Department of Public Health (A.H.), School of Medicine, Chiba University, Chiba, Japan.

Correspondence to Ituro Inoue, MD, Division of Molecular Life Science, School of Medicine, Tokai University, 143 Shimokasuya, Isehara, Kanagawa, 259-1193, Japan. E-mail ituro@is.icc.u-tokai.ac.jp

***Stroke* is available at http://stroke.ahajournals.org** **DOI: 10.1161/STROKEAHA.110.586529**

tal risk factors and to investigate whether the association between the 9p21 variation and risk of IA varied according to subphenotype.

Materials and Methods

Study Participants

All study participants were Japanese and recruited at Tokyo Women's Medical University, Chiba University, and their affiliated hospitals. The Ethics Committee of Tokyo Woman's Medical University, Chiba University, and Tokai University approved the study protocols and all participants gave written informed consent.

Nine hundred eighty-one IA patients included 322 familial IA patients (80 probands from nuclear families that had been used in our linkage study[8] and 242 patients who had a family history of IA) and 659 sporadic IA patients. The presence of IA was confirmed by digital subtraction angiography, 3-dimensional computed tomography angiography, magnetic resonance angiography, or surgical findings. Phenotyping of IAs, such as aneurysm location and number of aneurysms, was performed by certified neurosurgeons. The patients harboring saccular IA were included, but those harboring fusiform or dissecting IA were excluded. All 699 controls were screened for not harboring IA by means of neuroradiological imaging such as MRI. IA patients and controls were classified as hypertensive if they had a medical history of hypertension or were currently receiving antihypertensive therapy. The subjects were divided into 2 groups according to their self-reported smoking habit: never-smokers and ever-smokers. The dataset was updated from our previous studies.[2,8] Baseline characteristics of the study participants are summarized in Table 1.

SNP Genotyping

All study participants were genotyped for 4 9p21 SNPs (rs1333040, rs2891168, rs2383207, and rs10757278). These SNPs were selected as tag SNPs capturing SNPs previously reported to show strong association to IA, coronary artery disease, abdominal aortic aneurysm, and peripheral arterial disease[2,3,9–13] at $r^2>0.9$ in phase II HapMap JPT (Japaneses in Tokyo, Japan) samples. Linkage disequilibrium structure and selection criterion of the SNPs are shown in Supplemental Figure I available online at http://stroke.ahajournals.org. Genotyping was performed by TaqMan SNP Genotyping Assays on the ABI PRISM 7900HT Sequence Detection System (Applied Biosystems) or by direct sequencing with BigDye Terminators v3.1 Cycle Sequencing Kit on the Applied Biosystems 3730 DNA analyzer (Applied Biosystems).

Statistical Analysis

The statistical significance of departure from Hardy-Weinberg equilibrium in control samples was examined by means of the exact test using PLINK software.[14] We assessed the significance of association using the Cochran-Armitage trend test. To determine the most likely genetic model, additive and nonadditive effects were modeled in the context of logistic regression analysis by defining the SNP genotype predictor value x_1 and x_2 as $x_1 = \begin{cases} 0, aa \\ 1, Aa \\ 2, AA \end{cases}$ and $x_2 = \begin{cases} 0, aa \\ 1, Aa, \\ 0, AA \end{cases}$ respectively, where "A" was the putative risk allele. If there is no evidence of nonadditivity ($P>0.05$), then the multiplicative model (additive in logit scale) is selected; otherwise, the dominance or recessive model is selected.

The stepwise logistic regression procedure was applied to assess the relative importance of the 4 SNPs. The significance level of 0.01 was necessary for entering a SNP into the model, and the significance level of 0.01 was necessary for a SNP to stay in the model at any iteration step. When applying the stepwise logistic regression procedure, missing genotypes were imputed via Beagle software[15] with default settings. The accuracy of the imputation was assessed by means of the R^2 values (ie, the estimates of the squared correlation between the imputed allele dosage and the true allele dosage).[15] The resulting phased genotype data were used for haplotype analysis.

Table 1. Baseline Characteristics of the Study Participants Stratified by Affected Status

Variables	Cases (n=981)	Controls (n=699)
Subject characteristics		
Female, n (%)	614 (62.6)	274 (39.2)
Mean (SD) age, y	58.4 (11.3)	62.7 (9.9)
IA risk factors		
Smoking habit,* n (%)		
Ever	409 (51.4)	279 (44.3)
Never	387 (48.6)	351 (55.7)
History of hypertension,* n (%)		
Yes	487 (59.8)	306 (47.1)
No	327 (40.2)	344 (52.9)
IA subphenotypes		
Multiplicity of aneurysms, n (%)		
Multiple	192 (21.2)	...
Single	713 (78.8)	...
Missing, n	76	...
History of SAH,† n (%)		
Ruptured	513 (71.9)	...
Unruptured	200 (28.1)	...
Site distribution of aneurysms,† n (%)		...
Location		
MCA	228 (32.0)	...
AcomA	191 (26.8)	...
PcomA	148 (20.8)	...
ICA	44 (6.2)	...
ACA	32 (4.5)	...
BA	20 (2.8)	...
VA	14 (2.0)	...
PCA	1 (0.1)	...
Others	35 (4.9)	...
Grouping according to ISUIA, n (%)		
AC/MC/IC	495 (69.4)	...
Post-Pcomm	183 (25.7)	...
Others	35 (4.9)	...

*The data regarding smoking habit and history of hypertension were incomplete. Differences between total and analyzed sample size were attributable to missing data.

†The analyses concerning history of SAH and site distribution of aneurysms was restricted to 713 patients who had a single aneurysm because patients with multiple IAs may simultaneously have different types of IAs, which makes it difficult to select the correct subphenotype, eg, patients with both ruptured and unruptured IAs.

The International Study of Unruptured Intracranial Aneurysm (ISUIA)[5] classified IA sites into the following two groups: AC/MC/IC, aneurysms in ACA, AcomA, ICA or MCA; and Post-Pcomm, those in BA, VA, PCA or PcomA.

ACA indicates anterior cerebral artery; AcomA, anterior communicating artery; BA, basilar artery; IA, intracranial aneurysm; ICA, internal carotid artery (not the cavernous portion); MCA, middle cerebral artery; Others, the cavernous portion of the ICA or unknown; PCA, posterior cerebral artery; PcomA, posterior communicating artery; SAH, subarachnoid hemorrhage; VA, vertebral artery.

ORs of IA for the selected SNP and 2 risk factors (smoking habit and history of hypertension) were estimated by logistic regression analysis using a univariate analysis for each predictor, and then a multivariate model including all the predictors with adjustment for