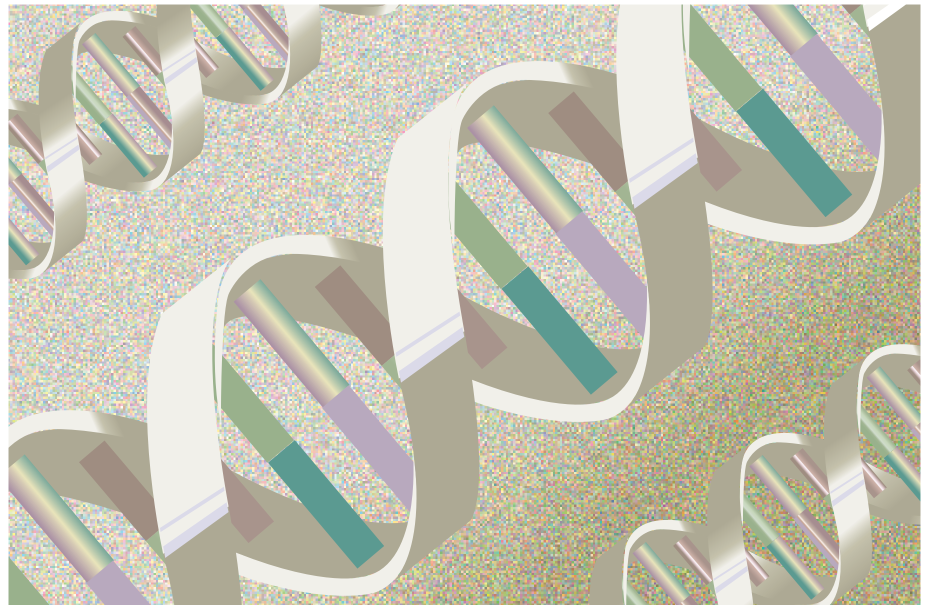
nature

FEATURE

Replicating genotype–phenotype associations

What constitutes replication of a genotype–phenotype association, and how best can it be achieved?

NCI-NHGRI Working Group on Replication in Association Studies



The study of human genetics has recently undergone a dramatic transition with the completion of both the sequencing of the human genome and the mapping of human haplotypes of the most common form of genetic variation, the single nucleotide polymorphism (SNP)[1–3]. In concert with this rapid expansion of detailed genomic information, cost-effective genotyping technologies have been developed that can assay hundreds of thousands of SNPs simultaneously. Together, these advances have allowed a systematic, even 'agnostic', approach to genome-wide interrogation, thereby relaxing the requirement for strong prior hypotheses.

So far, comprehensive reviews of the published literature, most of which reports work based on the candidate-gene approach, have demonstrated a plethora of questionable genotype–phenotype associations, replication of which has often failed in independent studies[4–7]. As the transition to genome-wide association studies occurs, the challenge will be to separate true associations from the blizzard of false positives attained through attempts to replicate positive findings in subsequent studies. The purpose of a replication study is to evaluate a positive finding from a previous study, to provide credibility that the initial finding is valid. Replication is essential for establishing the credibility of a genotype–phenotype association, whether derived from candidate-gene or genome-wide association studies. However, there is a lack of agreement about what constitutes a finding deserving of replication, what constitutes an adequate replication study and what constitutes a replication or refutation.

Investigators and journal editors have offered guidelines for how to address this problem[8–12], but these initial efforts have been hampered by limited experience and conflicting empirical data. However, as evidence has accumulated, several instructive examples have emerged of genotype–phenotype associations being reproduced reliably in follow-up studies. These include peroxisome proliferator-activated receptor-γ (*PPARG*)[13] and the transcription factor *TCF7L2* (refs 14–19), related to diabetes; nucleotide-binding oligomerization domain containing 2 (*NOD2*) and Crohn's disease[20–22]; complement factor H (*CFH*) and age-related macular degeneration[23–26]; and chromosome region 8q24 and prostate cancer risk[27–31].

Many instances have arisen in which initial findings have not been reproduced in follow-up studies because of issues in either the initial study or the attempted replication[4–6,32,33]. Small sample size is a frequent problem and can result in insufficient power to detect minor contributions of one or more alleles. Similarly, small sample sizes can provide imprecise or incorrect estimates of the magnitude of the observed effects. Poor study design — particularly a lack of comparability between cases and controls — can increase the risk of biases because there can be heterogeneity in exposure to environmental challenges and population stratification. The latter arises when investigators fail to account for case–control differences in the genetic structure of the underlying population. Heterogeneity in classification of outcomes across studies can undermine the opportunity to compare among them. Similarly, data 'dredging' can be a major problem, especially when criteria for defining phenotypes are altered to achieve statistical significance worthy of publication.

Another challenge arises when follow-up studies analyse different variants. An example is the reported association between *DTNBP1* and schizophrenia, initially identified in Irish pedigrees[34] and 'confirmed' in independent European studies[35]. Unfortunately, different risk alleles and haplotypes were reported in each study, making comparison difficult[36–39]. Although it is plausible that more than one variant could contribute to schizophrenia risk at the *DTNBP1* locus, it is difficult to draw this conclusion from the literature because follow-up studies have not consistently analysed the same markers or those in perfect linkage disequilibrium ($r^2 = 1.0$). Other recent examples for which initial reports of association have been inconsistently replicated include insulin-induced gene 2 (*INSIG2*) and obesity[40–44], and cyclic-AMP-specific phosphodiesterase (*PDE4D*) and stroke[45,46]. These have been accompanied by controversies about what actually constitutes replication.

This paper presents the conclusions of a working group on the replication of genotype–phenotype associations — whether identified in genome-wide or candidate-gene studies — convened by the National Cancer Institute and the National Human Genome Research Institute. The group was composed of experts from diverse disciplines, including biostatistics, clinical medicine, epidemiology, genetics and scientific publishing. The purpose was to review the current state of the field and propose best practices for the design, conduct and publication of replication studies that aim to follow up notable findings, particularly in genome-wide association studies. The group addressed three topics. First, assessment of the validity and limitations of any single genetic association study. Second, criteria for establishing replication in genetic association studies. Third, points to consider for publication of high-quality genotype–phenotype association reports (Box 1).

©2007 Nature Publishing Group

Box 1 | Points to consider in genotype–phenotype association reports

This checklist is intended to serve as a guide for authors, journal editors and referees to allow clear and unambiguous interpretation of the data and results of genome-wide and other genotype–phenotype association studies.

Study information
- A detailed description of the study design and its implementation
- The source of cases and controls (or cohort members, if based on cohort design), including time period and location(s) of subject recruitment
- Methods for ascertaining and validating affected or unaffected status and reproducibility of classification
- Participation rates for cases, controls or cohort members
- Presentation of case and control selection in a flow chart, including exclusion points for missing and erroneous data (possibly as supplementary tables)
- Initial table comparing relevant characteristics (such as demographics, risk factors and exposures) of cases and controls
- Success rate for DNA acquisition, including comparisons of those with and without collection, extraction failures and exclusions due to inconsistent data

Data issues
- Statement on availability of results and data so that, as far as possible, others can analyse them independently
- Links to supplemental online resources and database accession numbers

Genotyping and quality control procedures
- Sample tracking methods, such as bar-coding, to ensure accuracy of analysis
- Description of genotyping assays and protocols, particularly when new or applied in a non-standard method
- Description of genotyping calling algorithm
- Genotype quality control design for samples, including numbers, plating locations, selection criteria for:
  - External control samples from standard accepted sets (such as HapMap)
  - Internal control samples (duplicate samples; it should be specified whether these are from the same or different DNA collection, extraction or aliquot)
- Assay and DNA quality metrics by locus, sample, plate or 'batch'
- Assay call rates
- Average error rates estimated by internal duplicates or external samples
- Assay reproducibility: concordance for performance of extraction, aliquoting (internal control samples) and assay reproducibility
- Concordance with published or previously generated genotypes
- Mendelian consistency checks if related individuals are present
- Detection of inconsistent or cryptic relatedness in study subjects
- Evaluation of deviations from Hardy–Weinberg proportions to detect failed assays or large-scale stratification (for example, testing Hardy–Weinberg equilibrium 'violations') separately in cases and controls
- Assessment of population heterogeneity, including
  - Average or median value of chi-square and full distribution
  - Q–Q plots of chi-square analysis and *P*-values (with specific description of type of test used to generate the values)
- Validation of most critical results on an independent genotyping platform

Results
- Analysis methods in sufficient detail to reconstruct the analytical approach and reproduce all reported results
- Description of any pre-analysis weighting scheme for selecting variants for replication
- Simple single-locus and multi-marker (haplotype) association analyses
- Genetic models tested (unconstrained genotype effects — dominant, additiv e, multiplicative or trend)
- Graphical display of genotype clustering for assays of high interest
- Verification of results at highly correlated loci
- Discussion of choice of threshold for significance and the statistical basis for any adjustment for multiple testing and the relationship to overall study power
- Significance of any known 'positive controls' (that is, loci established in previous genetic associations)
- Consistency of results before and after application of quality control filters

Replication studies
- Description of replication samples, including source, ascertainment and comparability to initial sample
- Discussion of choice of threshold for significance and the statistical basis for any adjustment for multiple testing and the relationship to overall study power
- Summary of replication and analysis attempts by authors
- Summary of all known replication attempts by others, including non-replications

Genotyping data and specifications for deposition in standard databases
- Availability of 'raw' genotype data in the technology and vendor format, consistent with the requirements or restrictions imposed by funding agencies or informed consent
- Data extraction and processing protocols
- Normalization, transformation and data selection procedures and parameters

Points for reviewers and authors to consider regarding priority for publication
- Strength of observation
- Suitably large sample size
- Sufficiently stringent criteria for significance (small *P*-values)
- High quality of study design, including selection of study population, reliability of phenotypes, measurement and adjustment for potential confounders
- Discussion and conclusions commensurate with sample size, power, *P*-value and epidemiological quality of study design
- Quality control standards used, including assessment of genotype quality and completeness
- Usefulness of observations to others for subsequent research
- Value of initial hypothesis described
- Brief presentation of implications, especially as they relate to further follow-up both of genetic markers and for corroborative studies to investigate plausibility
- Explanations of notable findings
- Appropriate alternative explanations proposed and briefly discussed
- Biological or functional explanations based firmly on available data

Initial association studies

The initial study of any association represents an important discovery tool. In the near future, it is unlikely that a single study will unequivocally establish a valid genotype–phenotype association and not require replication. A number of points relating to the study design and reporting should be considered in determining whether a finding in an initial genome-wide or candidate-gene study merits follow-up replication studies (Box 2). Attempts to replicate a reported association are often complicated by lack of methodological detail in the initial report or lack of methodological rigour in the original study.

Because of the enormous number of genotype–phenotype associations tested in each genome-wide study, spurious associations will substantially outnumber true ones unless rigorous statistical thresholds are applied. Although no universal threshold can be specified for statistical significance in all circumstances, smaller *P*-values generally provide greater support for a true association. Extremely small *P*-values should be interpreted carefully, however, until completion of replication studies, because many can be due to inappropriate reliance on asymptotic distributions of test statistics, or to technical artefact or genotype errors that are distributed differently between cases and controls. Cluster plots for highly significant markers should be examined carefully. It may be desirable to include confirmatory data from a second genotyping technology in the initial report to verify genotype accuracy. Cases and controls should be drawn from populations that are generally comparable both in terms of genetic background and environmental exposures[47], and should be analysed for

©2007 Nature Publishing Group

confounding population stratification. This may require genotyping of ancestry informative markers (AIMs), which should be strongly encouraged as genotype costs fall and AIMs become increasingly well-characterized within marker sets. Family-based studies are affected by population stratification, so researchers should opt for methods robust to this, such as transmission disequilibrium methods[48]. They may be particularly valuable in the initial study if there is evidence for ethnic differences in the genetic effect of a trait, although at the cost of increased genotyping. Cautious interpretation is required either if significance is observed only for unusual or highly specific phenotypes (especially if they represent a small proportion of the study sample) or if significance depends on a particular analytical method that is not publicly available for confirmation.

Approaches for dealing with multiple comparisons are beyond the scope of this report, but more robust methods are clearly needed[49]. Permutation testing is an effective strategy to address the problem of multiple comparisons, especially if a large number of phenotypes are being analysed. Many methods for addressing the problem of multiple comparisons invoke a conservative approach, namely a standard Bonferroni correction, which assumes the independence of all tests performed. In many association studies, markers are not independent because they are in linkage disequilibrium, and so a standard Bonferroni correction is overly conservative. Lowering the threshold for calling a finding of particular variants — such as non-synonymous coding SNPs — positive in the analysis scheme (weighting) has merit but must be declared before initiation of the analysis and not once the analysis has begun[49,50]. The number of variants for which there is either credible laboratory evidence or a validated *in silico* prediction *a priori* is quite small. However, the temptation to create a credible biological hypothesis *post hoc* can be quite strong.

At present, many studies are barely powered to identify, much less to establish, associations of common alleles of weak effect in complex diseases[51,52]. Recently, appreciation of this crucial issue has led to larger, more definitive studies, such as the Cancer Genetic Markers of Susceptibility (CGEMS) project and the Wellcome Trust Case Control Consortium, (WTCCC). An estimated large effect (that is, with an odds ratio greater than 2) in a well-powered study can lend credence to an association, because unknown confounding factors are less likely to produce large effects[53]. Unfortunately, many risk variants contribute less than this. Small studies are prone to large variation in risk estimates, of which only selected strong positives are initially detected and reported. Furthermore, the estimate of the effect declines as replication studies are pursued, a phenomenon known as 'winner's curse'[54,55].

Consortial studies comprised of multiple independent studies combined into a pooled analysis can be viewed as a practical approach that overcomes many of the disadvantages of a disconnected set of underpowered studies. In addition, consortia may meet the need for rapid replication by achieving sufficiently large sample size[40,56]. Collaborations among multiple independent studies can offer important advantages over a single large study, particularly regarding the generalizability of findings observed in multiple studies that typically have greater diversity of populations and/or exposures.

Box 2 | Suggested criteria for establishing the soundness of an initial association report

These criteria are intended for studies of genotype–phenotype associations assessed by genome-wide or candidate-gene approaches.

- Statistical analyses demonstrating the level of statistical significance of a finding should be published or at least available so that others can attempt to reproduce the reported results
- Explicit information should be provided about the study's power to detect a range of effects
- The study should be epidemiologically sound, with careful accounting for potential biases in selection of subjects, characterization of phenotypes, comparability of environmental exposures (when possible) and underlying population structure in cases and controls
- Phenotypes should be assessed according to standard definitions provided in the report
- Associations should be consistent (within the range of expected statistical fluctuation) and reported for the same phenotypes across study subgroups or across similar phenotypes in the entire study group
- Significance should not depend on altering the quality control methods beyond standard approaches that could change inclusion or exclusion of large numbers of samples or loci
- Measures to assess the quality of genotype data should include results of known study sample duplicates or publicly available samples
- The results for concordance between duplicate samples (if applicable) as well as completion and call rates per SNP and per subject should be disclosed, along with rates of missing data
- A subset of notable SNPs should be evaluated with a second technology that verifies the same result with excellent concordance, because no technology is error-free
- Associations with nearby SNPs in strong linkage disequilibrium with the putatively associated SNP should be reported (and should be similar)
- The results of replication studies of previous findings should be reported even if the results are not significant
- Testing for differences in underlying population structure in case and control groups should be performed and reported
- Appropriate correction for multiple comparisons across all statistical tests examined should be reported. Comparison to genome-wide thresholds should be described. Similarly, for bayesian approaches, the choice of prior probabilities should be described

As far as possible, similarly rigorous criteria should be considered for evaluation of genotype–phenotype association studies with limited or no availability of subjects for replication, such as studies of rare diseases or severe toxicity due to therapy or environmental exposures. In these circumstances, additional information gathered from laboratory techniques, bioinformatic tools and *a priori* biological insight should be used to provide plausibility for interpreting genetic association findings. The expectation for demonstrated replication might be relaxed if it is unethical to attempt replication — such as in studies that link genetic variation with adverse effects of therapy or environmental exposure (for example, benzene or cigarette smoke). Similarly, the public health impact of a finding may lessen the stringency of expectation for replication before initial publication — for example, in an urgent situation in which effective intervention is available and can be readily implemented.

Genotype–phenotype associations that have been replicated widely have often used clearly defined phenotypes classified by standard and widely-accepted criteria, such as diabetes and age-related macular degeneration[57,58]. Use of accepted criteria should reduce misclassification rates[59]. Some association studies have reported intermediate phenotypes (known as endophenotypes) but have provided little detail on the actual measure or its reliability[60]. In the absence of standard criteria, sufficient detail should be provided for both the definition of the phenotypes investigated and assessment of their validity and comparability across studies.

Replication of initial studies

To establish a positive replication of a genotype–phenotype association, many of the same considerations important for genome-wide association or candidate-gene studies should be fulfilled (Box 3). In replication studies, every effort should be made to analyse phenotypes comparable to those reported in the initial study. In the first attempt to replicate a finding, comparable populations should be analysed not only for the main effect but also to guard against confounding population stratification, either in the initial or replication studies[61,62]. Because many initial studies and replication studies have been reported in populations of European descent, the challenge remains to extend the studies to other populations. It has already been shown that many variants that have a significant association with disease in several studies in one population may not necessarily have the same association in another (such as *TCF7L2* in West Africa and

©2007 Nature Publishing Group

Box 3 | Suggested criteria for establishing positive replication

These criteria are intended for follow-up studies of initial reports of genotype–phenotype associations assessed by genome-wide or candidate-gene approaches.

- Replication studies should be of sufficient sample size to convincingly distinguish the proposed effect from no effect
- Replication studies should preferably be conducted in independent data sets, to avoid the tendency to split one well-powered study into two less conclusive ones
- The same or a very similar phenotype should be analysed
- A similar population should be studied, and notable differences between the populations studied in the initial and attempted replication studies should be described
- Similar magnitude of effect and significance should be demonstrated, in the same direction, with the same SNP or a SNP in perfect or very high linkage disequilibrium with the prior SNP (r^2 close to 1.0)
- Statistical significance should first be obtained using the genetic model reported in the initial study
- When possible, a joint or combined analysis should lead to a smaller P-value than that seen in the initial report[75]
- A strong rationale should be provided for selecting SNPs to be replicated from the initial study, including linkage-disequilibrium structure, putative functional data or published literature
- Replication reports should include the same level of detail for study design and analysis plan as reported for the initial study (Box 1)

East Asia[18,63,64]; in this case, it has provided an opportunity to refine the signal to a restricted region). In some circumstances, it might be impossible to conduct follow-up studies because of the uniqueness of a study population or the lack of availability of additional subjects for replication. If replication is not an option, interpretation of association findings could be supplemented by biological insights derived from the laboratory.

Evaluation of an association in populations of different ancestry from that of the initial report would generally be expected, because genomic variation is greater when compared across populations, and should increase confidence in the finding. By contrast, failure to replicate in a population different from that of the initial report does not necessarily invalidate the original finding. In some cases, the differences in linkage disequilibrium relationships across populations can be used to narrow the region of interest for later genetic and possible functional analysis. Owing to their robustness to population stratification, as noted above, family-based studies can also serve as valuable replication studies for notable findings[48].

Reports of attempts at replication should distinguish between tests of the same SNP as in the original study, SNPs in strong linkage disequilibrium with the reported SNP, and other SNPs that were genotyped to search for additional variants associated with disease in the region (Fig. 1). In some circumstances, the initial study might have identified a marker that is not in strong linkage disequilibrium with the causal variant, which could lead to a false refutation in a different population, whereas testing additional SNPs in the region might reveal another association worthy of follow-up. For clarity, if new, previously untested SNPs are included, they should be clearly identified and the rationale for their inclusion explicitly stated. If differences in linkage disequilibrium patterns across populations are used to invoke an association at a new marker but not at the originally tested marker, the different linkage disequilibrium patterns should be empirically demonstrated in the appropriate populations and shown to be a plausible and consistent explanation for both the new and original results. Otherwise, the new association cannot be considered a replication.

Publication of associations

The evaluation of a publication addressing one or more genotype–phenotype associations is a daunting task in the age of large, dense datasets. To this end, published genome-wide association reports should include detailed descriptions of design, genotyping and statistical methods, and results, even if available only through online supplements, or perhaps in a separate journal. A checklist of key possible issues is provided in Box 1 — this could be used as a guide for authors, editors, reviewers and the general readership.

It is a challenge to make the case for the importance of the replication finding(s) without exaggerating the significance of the observation. Remarks about possible follow-up of genetic markers and corroborative studies to investigate plausibility should be brief and well referenced. Authors should practise sound judgement and temper enthusiasm based on prior publications (especially from the same investigative group), particularly if the replication study results differ from those of the initial study. Disclosure of known previous attempts to replicate the reported findings, whether positive or negative, by the authors or others is important for interpreting the replication study.

Although it is desirable for the initial report of a genotype–phenotype association to include adequately powered replication studies, requiring replication with every initial study may not be necessary, as long as the preliminary nature of a study without replication is emphasized. Such studies can still provide valuable information if the entire set of results is made available, and releasing such results before replication would be of value to the field. However, there is substantial added value in presenting robust findings based on an initial scan together with follow-up replication, and an appropriate balance is needed that facilitates rapid publication of valid findings and encourages collaboration[19,65]. If replication studies are included, each should be described or referenced in the same detail as the initial study and should include the results for all SNPs tested at each stage. As noted above, replication studies should preferably investigate the same or a very similar phenotype.

In many cases, the follow-up study will fail to replicate the initial results. Such findings are valuable for distinguishing false-positives from the true-positive signals that should be pursued for putative causal variants. The preference for publishing positive findings, even if derived from suboptimal studies, presents a formidable barrier to the dissemination of well-conducted negative studies. Failure to disseminate results from well-conducted negative studies withholds essential pieces of evidence for investigators who may be deciding whether to launch a follow-up study to replicate or to extend the original study. Thus, high-quality instances of 'meaningful negativity' are useful and should be reported succinctly in the literature. Criteria for a meaningful negative

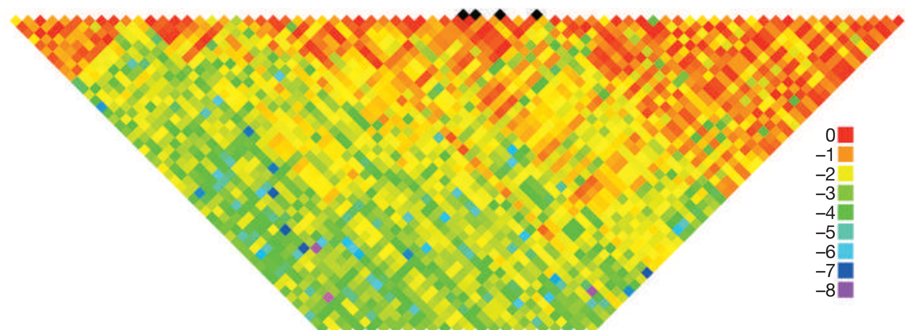


Figure 1 | Linkage disequilbrium across the region containing SNPs associated with breast cancer in *FGFR2*. Black diamonds represent four single nucleotide polymorphisms (SNPs; rs11200014, rs2981579, rs1219648 and rs2420946) for which associations with breast cancer were replicated in multiple studies[73,74]. Estimates of the square of the correlation coefficient (r^2) were calculated for each pairwise comparison of SNPs in the initial genome-wide association study across the *FGFR2* region[73]. The log(10) r^2 values are colour-coded.

©2007 Nature Publishing Group

replication study are the same as those for a positive study (Box 3), with the added requirements that the same trait should be studied in a population of comparable underlying structure with sufficient power to measure the appropriate effect size and yield a negative result.

Negative studies are difficult to publish but they are crucial for separating true-positive from false-positive findings. Journals are strongly encouraged to publish high-quality negative studies refuting earlier positive reports of genotype–phenotype associations. The journal in which the initial scan is published is encouraged to solicit and publish well-conducted follow-up studies within a specified time frame, perhaps between 3 and 9 months of the initial report. A case in point is the recent collection of reports published by *The American Journal of Human Genetics*[66–71] that failed to replicate the initial findings of a genome-wide association study on Parkinson's disease. A handful of journals — such as *Cancer Epidemiology, Biomarkers and Prevention* and the new *PLoS* series[72] — currently feature well-conducted negative reports, and such efforts are to be lauded. The value of a well-executed negative study cannot be overemphasized; more venues are needed to capture these valuable results.

Although there are challenges to making data on individual research participants available to other investigators, every effort should be made to provide researchers with an opportunity to reproduce the reported results and to investigate new hypotheses and methods. To facilitate this research in genome-wide association studies, a public data archive known as the Database of Genotypes and Phenotypes, or dbGaP (http://view.ncbi.nlm.nih.gov/dbgap) has been established at the National Library of Medicine's National Center for Biotechnology Information and will be used by many National Institutes of Health (NIH)-supported studies. dbGaP will provide study documentation and aggregated genotype and phenotype data through its website with no account or authorization required. Access to individual, de-identified genotype and phenotype data will require an authorization and approval process that is currently under development. Whether through dbGaP or other venues, genotype summaries of computed analyses should be published online unless there are strong reasons not to do so, such as data derived from special populations (that is, isolated populations or minority communities) or other groups that will not permit such sharing. There are substantial informatic challenges for data presentation and data archiving, especially on public and journal websites. Best practices for retrieval and analysis of such data continue to evolve.

Conclusion

The history of genotype–phenotype association studies has focused on initial discoveries as opposed to careful replication. Earlier attention to the appropriate design of subsequent replication studies might have helped limit the plethora of false-positive results. Determination of valid genotype–phenotype associations presents a series of challenges that will require a logical strategy for conducting well-designed studies, based on excellent quality control practices interwoven with sound analytical methods and judicious interpretation. Other than the obvious differences in the drawbacks involved in multiple comparisons, standards for assessing the validity of the initial findings of a genotype–phenotype association should not differ substantially between the candidate-gene approach and genome-wide association studies. As experience accumulates, we can look forward to methodological advances that will facilitate our interpretation of studies, such as continued improvement of proposed methods for lowering the threshold for positive findings, adjustments for population structure, and exploitation of linkage disequilibrium structure in a candidate region.

The best practices suggested here for reporting initial and replication studies are based on sufficient disclosure of study methods to permit independent confirmation of study findings. Often a sequence of studies will be required to establish a valid genotype–phenotype association, perhaps involving several rounds of replication studies. And, of course, the conclusive demonstration of a replicated association represents only the beginning of the process towards finding the causal genetic variant(s). Labour-intensive and costly investigation will subsequently be required to sequence the candidate interval in depth, genotype all the common and perhaps uncommon variants that are markers for the outcomes of interest in multiple population samples, understand their functional consequences, examine their potential interactions with other genes or environmental factors, and devise strategies for preventative or therapeutic interventions. None of these steps should proceed far, however, without conclusive replication of findings from an initial genotype–phenotype association study. ■

Note added in proof: Recently, a series of papers have also shown replication across as well as within genome-wide association studies in common complex diseases such as breast cancer, type 2 diabetes, and coronary disease[73,74,76–81].

1. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431,** 931–945 (2004).
2. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437,** 1299–1320 (2005).
3. Hinds, D. A. *et al.* Whole-genome patterns of common DNA variation in three human populations. *Science* **307,** 1072–1079 (2005).
4. Hirschhorn, J. N., Lohmueller, K., Byrne, E. & Hirschhorn, K. A comprehensive review of genetic association studies. *Genet. Med.* **4,** 45–61 (2002).
5. Ioannidis, J. P., Ntzani, E. E., Trikalinos, T. A. & Contopoulos-Ioannidis, D. G. Replication validity of genetic association studies. *Nature Genet.* **29,** 306–309 (2001).
6. 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. *Nature Genet.* **33,** 177–182 (2003).
7. Ioannidis, J. P. Common genetic variants for breast cancer: 32 largely refuted candidates and larger prospects. *J. Natl Cancer Inst.* **98,** 1350–1353 (2006).
8. Freely associating. *Nature Genet.* **22,** 1–2 (1999).
9. Todd, J. A. Statistical false positive or true disease pathway? *Nature Genet.* **38,** 731–733 (2006).
10. Neale, B. M. & Sham, P. C. The future of association studies: gene-based analysis and replication. *Am. J. Hum. Genet.* **75,** 353–362 (2004).
11. Clark, A. G., Boerwinkle, E., Hixson, J. & Sing, C. F. Determinants of the success of whole-genome association testing. *Genome Res.* **15,** 1463–1467 (2005).
12. Freimer, N. B. & Sabatti, C. Human genetics: variants in common diseases. *Nature* **445,** 828–830 (2007).
13. Altshuler, D. *et al.* The common PPARγ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nature Genet.* **26,** 76–80 (2000).
14. Field, S. F. *et al.* Analysis of the type 2 diabetes gene, *TCF7L2*, in 13,795 type 1 diabetes cases and control subjects. *Diabetologia* **50,** 212–213 (2007).
15. Grant, S. F. *et al.* Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nature Genet.* **38,** 320–323 (2006).
16. Groves, C. J. *et al.* Association analysis of 6,736 U.K. subjects provides replication and confirms *TCF7L2* as a type 2 diabetes susceptibility gene with a substantial effect on individual risk. *Diabetes* **55,** 2640–2644 (2006).
17. Saxena, R. *et al.* Common single nucleotide polymorphisms in *TCF7L2* are reproducibly associated with type 2 diabetes and reduce the insulin response to glucose in nondiabetic individuals. *Diabetes* **55,** 2890–2895 (2006).
18. Helgason, A. *et al.* Refining the impact of *TCF7L2* gene variants on type 2 diabetes and adaptive evolution. *Nature Genet.* **39,** 218–225 (2007).
19. Sladek, R. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445,** 881–885 (2007).
20. Hugot, J. P. *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411,** 599–603 (2001).
21. Ogura, Y. *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411,** 603–606 (2001).
22. Economou, M., Trikalinos, T. A., Loizou, K. T., Tsianos, E. V. & Ioannidis, J. P. Differential effects of NOD2 variants on Crohn's disease risk and phenotype in diverse populations: a metaanalysis. *Am. J. Gastroenterol.* **99,** 2393–2404 (2004).
23. Edwards, A. O. *et al.* Complement factor H polymorphism and age-related macular degeneration. *Science* **308,** 421–424 (2005).
24. Hageman, G. S. *et al.* A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc. Natl Acad. Sci. USA* **102,** 7227–7232 (2005).
25. Klein, R. J. *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* **308,** 385–389 (2005).
26. Haines, J. L. *et al.* Complement factor H variant increases the risk of age-related macular degeneration. *Science* **308,** 419–421 (2005).
27. Freedman, M. L. *et al.* Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc. Natl Acad. Sci. USA* **103,** 14068–14073 (2006).
28. Amundadottir, L. T. *et al.* A common variant associated with prostate cancer in European and African populations. *Nature Genet.* **38,** 652–658 (2006).
29. Gudmundsson, J. *et al.* Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nature Genet.* **39,** 631–637 (2007).
30. Haiman, C. A. *et al.* Multiple regions within 8q24 independently affect risk for prostate cancer. *Nature Genet.* **39,** 638–644 (2007).
31. Yeager, M. *et al.* Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nature Genet.* **39,** 645–649 (2007).
32. Colhoun, H. M., McKeigue, P. M. & Davey, S. G. Problems of reporting genetic associations with complex outcomes. *Lancet* **361,** 865–872 (2003).
33. Ioannidis, J. P., Trikalinos, T. A. & Khoury, M. J. Implications of small effect sizes of individual genetic variants on the design and interpretation of genetic association studies of complex diseases. *Am. J. Epidemiol.* **164,** 609–614 (2006).

©2007 Nature Publishing Group

34. Straub, R. E. *et al.* Genetic variation in the 6p22.3 gene *DTNBP1*, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia. *Am. J. Hum. Genet.* **71,** 337–348 (2002).
35. Van Den, B. A. *et al.* The *DTNBP1* (dysbindin) gene contributes to schizophrenia, depending on family history of the disease. *Am. J. Hum. Genet.* **73,** 1438–1443 (2003).
36. Bray, N. J. *et al.* Haplotypes at the dystrobrevin binding protein 1 (*DTNBP1*) gene locus mediate risk for schizophrenia through reduced *DTNBP1* expression. *Hum. Mol. Genet.* **14,** 1947–1954 (2005).
37. Funke, B. *et al.* Association of the *DTNBP1* locus with schizophrenia in a U.S. population. *Am. J. Hum. Genet.* **75,** 891–898 (2004).
38. Kirov, G. *et al.* Strong evidence for association between the dystrobrevin binding protein 1 gene (*DTNBP1*) and schizophrenia in 488 parent-offspring trios from Bulgaria. *Biol. Psychiatry* **55,** 971–975 (2004).
39. Mutsuddi, M. *et al.* Analysis of high-resolution HapMap of *DTNBP1* (Dysbindin) suggests no consistency between reported common variant associations and schizophrenia. *Am. J. Hum. Genet.* **79,** 903–909 (2006).
40. Herbert, A. *et al.* A common genetic variant is associated with adult and childhood obesity. *Science* **312,** 279–283 (2006).
41. Hall, D. H., Rahman, T., Avery, P. J. & Keavney, B. INSIG-2 promoter polymorphism and obesity related phenotypes: association study in 1428 members of 248 families. *BMC Med. Genet.* **7,** 83 (2006).
42. Rosskopf, D. *et al.* Comment on 'A common genetic variant is associated with adult and childhood obesity'. *Science* **315,** 187 (2007).
43. Dina, C. *et al.* Comment on 'A common genetic variant is associated with adult and childhood obesity'. *Science* **315,** 187 (2007).
44. Loos, R. J., Barroso, I., O'rahilly, S. & Wareham, N. J. Comment on 'A common genetic variant is associated with adult and childhood obesity'. *Science* **315,** 187 (2007).
45. Gretarsdottir, S., Gulcher, J., Thorleifsson, G., Kong, A. & Stefansson, K. Comment on the phosphodiesterase 4D replication study by Bevan *et al. Stroke* **36,** 1824 (2005).
46. Rosand, J., Bayley, N., Rost, N. & de Bakker, P. I. Many hypotheses but no replication for the association between *PDE4D* and stroke. *Nature Genet.* **38,** 1091–1092 (2006).
47. Manolio, T. A., Bailey-Wilson, J. E. & Collins, F. S. Genes, environment and the value of prospective cohort studies. *Nature Rev. Genet.* **7,** 812–820 (2006).
48. Spielman, R. S. & Ewens, W. J. The TDT and other family-based tests for linkage disequilibrium and association. *Am. J. Hum. Genet.* **59,** 983–989 (1996).
49. Roeder, K., Bacanu, S. A., Wasserman, L. & Devlin, B. Using linkage genome scans to improve power of association in genome scans. *Am. J. Hum. Genet.* **78,** 243–252 (2006).
50. Wacholder, S., Chanock, S., Garcia-Closas, M., El Ghormli, L. & Rothman, N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J. Natl Cancer Inst.* **96,** 434–442 (2004).
51. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273,** 1516–1517 (1996).
52. Risch, N. J. Searching for genetic determinants in the new millennium. *Nature* **405,** 847–856 (2000).
53. Angell, M. The interpretation of epidemiologic studies. *N. Engl. J. Med.* **323,** 823–825 (1990).
54. Clarke, R. *et al.* Lymphotoxin-α gene and risk of myocardial infarction in 6,928 cases and 2,712 controls in the ISIS case-control study. *PLoS Genet.* **2,** e107 (2006).
55. Zollner, S. & Pritchard, J. Overcoming the winner's curse: estimating penetrance parameters from case-control. *Am. J. Hum. Genet.* **80,** 605–615 (2007).
56. Rothman, N. *et al.* Genetic variation in *TNF* and *IL10* and risk of non-Hodgkin lymphoma: a report from the InterLymph Consortium. *Lancet Oncol.* **7,** 27–38 (2006).
57. Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* **20,** 1183–1197 (1997).
58. A randomized, placebo-controlled, clinical trial of high-dose supplementation with vitamins C and E and beta carotene for age-related cataract and vision loss. AREDS report no. 9. *Arch. Ophthalmol.* **119,** 1439–1452 (2001).
59. Freimer, N. & Sabatti, C. The human phenome project. *Nature Genet.* **34,** 15–21 (2003).
60. Flint, J. & Munafo, M. R. The endophenotype concept in psychiatric genetics. *Psychol. Med.* **37,** 163–180 (2007).
61. 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).
62. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genet.* **38,** 904–909 (2006).
63. Chandak, G. R. *et al.* Common variants in the *TCF7L2* gene are strongly associated with type 2 diabetes mellitus in the Indian population. *Diabetologia* **50,** 63–67 (2007).
64. Horikoshi, M. *et al.* A genetic variant of the transcription factor 7-like 2 gene is associated with risk of type 2 diabetes in the Japanese population. *Diabetologica* **50,** 747–751 (2007).
65. Duerr, R. H. *et al.* A genome-wide association study identifies *IL23R* as an inflammatory bowel disease gene. *Science* **314,** 1461–1463 (2006).
66. Maraganore, D. M. *et al.* High-resolution whole-genome association study of Parkinson disease. *Am. J. Hum. Genet.* **77,** 685–693 (2005).
67. Myers, D. R. Considerations for genomewide association studies in Parkinson disease. *Am. J. Hum. Genet.* **78,** 1081–1082 (2006).
68. Clarimon, J. *et al.* Conflicting results regarding the semaphorin gene (*SEMA5A*) and the risk for Parkinson disease. *Am. J. Hum. Genet.* **78,** 1082–1084 (2006).
69. Farrer, M. J. *et al.* Genomewide association, Parkinson disease, and *PARK10*. *Am. J. Hum. Genet.* **78,** 1084–1088 (2006).
70. Goris, A. *et al.* No evidence for association with Parkinson disease for 13 single-nucleotide polymorphisms identified by whole-genome association screening. *Am. J. Hum. Genet.* **78,** 1088–1090 (2006).
71. Li, Y. *et al.* A case-control association study of the 12 single-nucleotide polymorphisms implicated in Parkinson disease by a recent genome scan. *Am. J. Hum. Genet.* **78,** 1090–1092 (2006).
72. Patterson, M. & Cardon, L. Replication publication. *PLoS. Biol.* **3,** e327 (2005).
73. Hunter, D. J. *et al.* A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer. *Nature Genet.* Advance online publication, 27 May 2007 (doi:10.1038/ng2075).
74. Easton, D. F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* Advance online publication, 27 May 2007 (doi:10.1038/nature05887).
75. Skol, A. D., Scott, L. J., Abecasis, G. R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nature Genet.* **38,** 209–213 (2006).
76. Stacey, S. N. *et al.* Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nature Genet.* Advance online publication, 27 May 2007 (doi:10.1038/ng2064)
77. Saxena, R. *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316,** 1331–1336 (2007).
78. Scott, L. J. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316,** 1341–1345 (2007).
79. Zeggini, E. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316,** 1336–1341 (2007).
80. Helgadottir, A. *et al.* A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* Advance online publication, 3 May 2007 (doi:10.1126/science.1142842).
81. McPherson, R. *et al.* A common allele on chromosome 9 associated with coronary heart disease. *Science* Advance online publication, 3 May 2007 (doi:10.1126/science.1142447).

Author Contributions S.J.C. and T.M. contributed equally to this manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.J.C. or T.M. (e-mails: chanocks@mail.nih.gov; manoliot@nhgri.nih.gov).

Authorship of National Cancer Institute–National Human Genome Research Institute (NCI-NHGRI) Working Group on Replication in Association Studies

Stephen J. Chanock[1,2], Teri Manolio[3], Michael Boehnke[4], Eric Boerwinkle[5], David J. Hunter[6], Gilles Thomas[1], Joel N. Hirschhorn[7], Goncalo Abecasis[4], David Altshuler[8], Joan E. Bailey-Wilson[3], Lisa D. Brooks[3], Lon R. Cardon[9], Mark Daly[8], Peter Donnelly[10], Joseph F. Fraumeni Jr[1], Nelson B. Freimer[11], Daniela S. Gerhard[12], Chris Gunter[13], Alan E. Guttmacher[3], Mark S. Guyer[3], Emily L. Harris[3], Josephine Hoh[14], Robert Hoover[1], C. Augustine Kong[15], Kathleen R. Merikangas[16], Cynthia C. Morton[17], Lyle J. Palmer[18], Elizabeth G. Phimister[19], John P. Rice[20], Jerry Roberts[3], Charles Rotimi[21], Margaret A. Tucker[1], Kyle J. Vogan[22], Sholom Wacholder[1], Ellen M. Wijsman[23], Deborah M. Winn[24], Francis S. Collins[3]

[1]Division of Cancer Epidemiology and Genetics, and [2]Center for Cancer Research, National Cancer Institute, Bethesda, Maryland 20892-4605, USA. [3]National Human Genome Research Institute, National Institutes of Health, 31 Center Drive, Bethesda, Maryland 20892-2154, USA. [4]Department of Biostatistics, University of Michigan, 1420 Washington Heights, Ann Arbor, Michigan 48109-2029, USA. [5]Human Genetics Center, University of Texas Health Science Center, 1200 Herman Pressler, Houston, Texas 77030, USA. [6]Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, Channing Laboratory, 181 Longwood Avenue, Boston, Massachusetts 02115, USA. [7]Children's Hospital Boston, Harvard Medical School, Broad Institute of MIT and Harvard, Seven Cambridge Center, Cambridge, Massachusetts 02114, USA. [8]Massachusetts General Hospital, Broad Institute of MIT and Harvard, 185 Cambridge Street, Cambridge, Massachusetts 02114, USA. [9]Human Biology Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, P.O. Box 19024, Seattle, Washington 98109-1024, USA. [10]University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK. [11]Center for Neurobehavioral Genetics, University of California, Los Angeles, 695 Charles E Young Drive South, Box 708822, Los Angeles, California 90095-7088, USA. [12] Office of Cancer Genomics, National Cancer Institute, National Institutes of Health, 31 Center Drive, Bethesda, Maryland 20892-2580, USA. [13]*Nature*, Ninth Floor, 75 Varick Street, New York, New York 10013, USA. [14]Epidemiology and Public Health, Yale University School of Medicine, 60 College Street, P.O. Box 208034, New Haven, Connecticut 06510, USA. [15]DeCode Genetics, 8 IS-101 Sturlugata, Reykjavik, Iceland. [16]National Institutes of Mental Health, National Institutes of Health, 35 Convent Drive, Bethesda, Maryland 20892-3720, USA. [17]Brigham and Women's Hospital, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. [18]Western Australian Institute for Medical Research and University of Western Australia, Queen Elizabeth II Medical Centre, Hospital Avenue, Nedlands WA6009, Australia. [19]*New England Journal of Medicine*, 10 Shattuck Street, Boston, Massachusetts 02115-6094, USA. [20]Washington University School of Medicine, Box 8134, 660 South Euclid Avenue, St Louis, Missouri 63110, USA. [21]Genetic Epidemiology Unit, Howard University, National Human Genome Center, 2041 Georgia Avenue, NW Washington, DC 20060, USA. [22]*Nature Genetics*, Suite 104, 25 First Street, Cambridge, Massachusetts 02141, USA. [23]Division of Medical Genetics and Department of Biostatistics, University of Washington, Box 357720, Seattle, Washington 98195-7720, USA. [24]Epidemiology and Genetics Research Program, National Cancer Institute, National Institutes of Health, 6130 Executive Boulevard, Bethesda, Maryland 20892-7393, USA.

©2007 Nature Publishing Group