How to Interpret a Genome-wide Association Study

Thomas A. Pearson, MD, MPH, PhD
Teri A. Manolio, MD, PhD

In the past 2 years, there has been a dramatic increase in genomic discoveries involving complex, non-Mendelian diseases, with nearly 100 loci for as many as 40 common diseases robustly identified and replicated in genome-wide association (GWA) studies (T.A.M.; unpublished data, 2008). These studies use high-throughput genotyping technologies to assay hundreds of thousands of the most common form of genetic variant, the single-nucleotide polymorphism (SNP), and relate these variants to diseases or health-related traits. Nearly 12 million unique human SNPs have been assigned a reference SNP (rs) number in the National Center for Biotechnology Information's dbSNP database and characterized as to specific alleles (alternate forms of the SNP), summary allele frequencies, and other genomic information.

The GWA approach is revolutionary because it permits interrogation of the entire human genome at levels of resolution previously unattainable, in thousands of unrelated individuals, unconstrained by prior hypotheses regarding genetic associations with disease. However, the GWA approach can also be problematic because the massive number of statistical tests performed presents an unprecedented potential for false-positive results, leading to new stringency in acceptable levels of statistical significance and requirements for replication of findings.

The genome-wide, nonhypothesis-driven nature of GWA studies represents an important step beyond candidate gene studies, in which the high cost of genotyping had limited the number of variants assayed to several hundred at most. This required careful selection of variants to be studied, often based on imperfect understanding of the biologic pathways relating genes to disease. Many such associations failed to be replicated in subsequent studies, leading to calls for all genetic association reports to include documented replication of findings as a prerequisite for publication.

For non-Mendelian conditions, GWA studies also represent a valuable advance over family-based linkage studies, in which multiply affected families are arduously assembled and inheritance patterns are related to several hundred markers throughout the genome. Family-based linkage studies, although successful in identifying genes of large effect in Mendelian diseases such as cystic fibrosis and neurofibromatosis, have had more limited success in common diseases like atherosclerosis and asthma. Major limitations of linkage studies are relatively low power for complex disorders influenced by multiple genes, and the large size of the chromosomal regions shared among family members (often comprising hundreds of genes), in whom it can be difficult to narrow the

©2008 American Medical Association. All rights reserved.

(Reprinted) JAMA, March 19, 2008—Vol 299, No. 11 1335
linkage signal sufficiently to identify a causative gene.

GWA studies build on the valuable lessons learned from candidate gene and family linkage studies, as well as the expanding knowledge of the relationships among SNP variants generated by the International HapMap Project,\(^\text{12,13}\) to capture the great majority of common genetic differences among individuals and relate them to health and disease. These studies not only represent a powerful new tool for identification of genes influencing common diseases, but also use new terminologies (Box 1), apply new models, and present new challenges in interpretation. GWA studies rely on the “common disease, common variant” hypothesis, which suggests that genetic influences on many common diseases will be at least partly attributable to a limited number of allelic variants present in more than 1% to 5% of the population.\(^\text{14}\) Many important disease-causing variants may be rarer than this and are unlikely to be detected with this approach.

Although GWA discovery studies provide important clues to genomic function and pathophysiologic mechanisms, they are as yet many steps removed from actual clinical application. Nonetheless, they have gained considerable media attention and have the potential for generating queries from patients about whether to get tested for the “new gene for disease X” based on the latest report. In this article, we describe the design, interpretation, application, and limitations of GWA studies for clinicians and scientists for whom this evolving science may have great relevance.

Overview of GWA Studies

A GWA study is defined by the National Institutes of Health as a study of common genetic variation across the entire human genome designed to identify genetic associations with observable traits.\(^\text{15}\)

Although family linkage studies and studies comprising tens of thousands of gene-based SNPs also assay genetic variation across the genome,\(^\text{16}\) the National Institutes of Health definition requires sufficient density and selection of genetic markers to capture a large proportion of the common variants in the study population, measured in enough individuals to provide sufficient power to detect variants of modest effect.

The present discussion focuses on studies attempting to assay at least 100 000 SNPs selected to serve as proxies for the largest possible number of SNPs.\(^\text{12}\) The typical GWA study has 4 parts: (1) selection of a large number of individuals with the disease or trait of interest and a suitable comparison group; (2) DNA isolation, genotyping, and data review to ensure high genotyping quality; (3) statistical tests for associations between the SNPs passing quality thresholds and the disease/trait; and (4) replication of identified associations in an independent population sample or examination of functional implications experimentally.

Most of the roughly 100 GWA studies published by the end of 2007 were designed to identify SNPs associated with common diseases. However, the technique can also be used to identify genetic variants related to quantitative traits such as height\(^\text{17}\) or electrocardiographic QT interval,\(^\text{18}\) and to rank the relative importance of previously identified susceptibility genes, such as \(APOE^\text{e4}\) in Alzheimer disease\(^\text{19}\) and \(CARD15\) and \(IL23R\) in Crohn disease.\(^\text{20}\)

GWA studies can also demonstrate gene-gene interactions, or modification of the association of one genetic variant by another, as with \(GAB2\) and \(APOE\) in Alzheimer disease,\(^\text{21}\) and can detect high-risk haplotypes or combinations of multiple SNPs within a single gene, as in exfoliation glaucoma\(^\text{22}\) and atrial fibrillation.\(^\text{23}\) These studies have also been used to identify SNPs associated with gene expression, either as confirmation of a phenotypic association, such as asthma and \(ORMDL3\) expression,\(^\text{24}\) or more globally.\(^\text{25}\) Thus, GWA studies have broader applications than those solely involving discovery of individual SNPs associated with discrete disease end points.

Study Designs Used in GWA

By far the most frequently used GWA study design to date has been the case-control design, in which allele frequencies in patients with the disease of interest are compared to those in a disease-free comparison group. These studies are often easier and less expensive to conduct than studies using other designs, especially if sufficient numbers of case and control participants can be assembled rapidly. This design also carries the most assumptions, which if not met, can lead to substantial biases and spurious associations (Table 1).

The most important of these biases involve the selected, often unrepresentative nature of the study case participants, who are typically sampled from clinical sources and thus may not include fatal, mild, or silent cases not coming to clinical attention; and the lack of comparability of case and control participants, who may differ in important ways that could be related both to genetic risk factors and to disease outcomes.\(^\text{26}\)

If well-established principles of epidemiologic design are followed, case-control studies can produce valid results that, especially for rare diseases, may not be obtainable in any other way. However, genetic association studies using case-control methodologies have often not always adhered to these principles. The often sharply abbreviated descriptions of case and control participants and lack of comparison of key characteristics in GWA reports\(^\text{27}\) can make evaluation of potential biases and replication of findings quite difficult.\(^\text{28}\)

The trio design includes the affected case participant and both of his or her parents.\(^\text{29}\) Phenotypic assessment (classification of affected status) is performed only in the offspring and only affected offspring are included, but genotyping is performed in all 3 trio members. The frequency with which an allele is transmitted to an affected offspring from heterozygous parents is then estimated.\(^\text{30}\) Under the null hy-
Box 1. Terms Frequently Used in Genome-wide Association Studies

| **Alleles** | Alternate forms of a gene or chromosomal locus that differ in DNA sequence |
| **Candidate gene** | A gene believed to influence expression of complex phenotypes due to known biological and/or physiological properties of its products, or to its location near a region of association or linkage |
| **Copy number variants** | Stretches of genomic sequence of roughly 1 kb to 3 Mb in size that are deleted or are duplicated in varying numbers |
| **False discovery rate** | Proportion of significant associations that are actually false positives |
| **False-positive report probability** | Probability that the null hypothesis is true, given a statistically significant finding |
| **Functional studies** | Investigations of the role or mechanism of a genetic variant in causation of a disease or trait |
| **Gene-environment interactions** | Modification of gene-disease associations in the presence of environmental factors |
| **Genome-wide association study** | Any study of genetic variation across the entire human genome designed to identify genetic association with observable traits or the presence or absence of a disease, usually referring to studies with marker density of 100,000 or more to represent a large proportion of variation in the human genome |
| **Genotyping call rate** | Proportion of samples or SNPs for which a specific allele SNP can be reliably identified by a genotyping method |
| **Haplotype** | A group of specific alleles at neighboring genes or markers that tend to be inherited together |
| **HapMap** | Genome-wide database of patterns of common human genetic sequence variation among multiple ancestral population samples |
| **Hardy Weinberg equilibrium** | Population distribution of 2 alleles (with frequencies p and q) such that the distribution is stable from generation to generation and genotypes occur at frequencies of p^2, 2pq, and q^2 for the major allele homozygote, heterozygote, and minor allele homozygote, respectively |
| **Linkage disequilibrium** | Association between 2 alleles located near each other on a chromosome, such that they are inherited together more frequently than expected by chance |

©2008 American Medical Association. All rights reserved.
genetic variants. Although cohort studies are typically more expensive and take longer to conduct than case-control studies, they often include study participants who are more representative than clinical series of the population from which they are drawn, and they typically include a vast array of health-related characteristics and exposures for which genetic associations can be sought. For these reasons, genome-wide genotyping has recently been added to cohort studies such as the Framingham Heart Study and the Women’s Health Study. Many GWA studies use multistage designs to reduce the number of false-positive results while minimizing the number of costly genome-wide scans performed and retaining statistical power. Genome-wide scans are typically performed on an initial group of case and control participants and then a smaller number of associated SNPs is replicated in a second or third group of case and control participants (Table 2). Some studies begin with small numbers of participants in the initial scan but carry forward large numbers of SNPs to minimize false-negative results. Other studies begin with more participants but carry forward a smaller proportion of associated SNPs. Optimal proportions of study participants and SNPs in each phase have yet to be determined, but carrying forward a small proportion (<5%) of stage 1 SNPs will often mean limiting the associations ultimately identified to those having a relatively large effect.

### Selection of Study Participants

Many genetic studies, whether GWA or otherwise, focus on case participants more likely to have a genetic basis for their disease, such as early-onset cases or those with multiple affected relatives. Misclassification of case participants can markedly reduce study power and bias study results toward no association, particularly when large numbers of unaffected individuals are misclassified as affected. For diseases that are difficult to diagnose reliably, ensuring that cases are truly affected (as by invasive testing or imaging), is probably more important than ensuring generalizability, although the limitations on diagnostic reliability and generalizability should be clearly described so that clinicians can judge the relevance to their patients.

The control participants should be drawn from the same population as the case participants and should be at risk to develop the disease and be detected in the study. Inclusion of women as controls in genetic association studies of diseases limited to men, for example, is problematic in that this approach adds individuals to the control group who had no chance of developing the disease (but might have done so had they also inherited a Y chromosome), thus mixing the controls with possible latent cases. This artificially reduces the differences in allele frequencies between cases and controls and limits the ability of the study to detect a true difference (ie, reduces study power).

If the disease is common, such as coronary heart disease or hypertension in the United States, efforts should be made to ensure that the controls are truly disease free. Some studies address this by using super-controls or persons at high risk but without even early evidence of disease, such as per-

### Table 1. Study Designs Used in Genome-wide Association Studies

<table>
<thead>
<tr>
<th>Assumptions</th>
<th>Case-Control</th>
<th>Cohort</th>
<th>Trio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants under study are more representative of the population from which they are drawn</td>
<td>Participants under study are more representative of the population from which they are drawn</td>
<td>Disease-related alleles are transmitted in excess of 50% to affected offspring from heterozygous parents</td>
<td></td>
</tr>
<tr>
<td>Genomic and epidemiologic data are collected similarly in cases and controls</td>
<td>Diseases and traits are ascertained similarly in individuals with and without the gene variant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Differences in allele frequencies relate to the outcome of interest rather than differences in background population between cases and controls</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short time frame</td>
<td>Cases are incident (developing during observation) and free of survival bias</td>
<td>Controls for population structure; immune to population stratification</td>
<td></td>
</tr>
<tr>
<td>Large numbers of case and control participants can be assembled</td>
<td>Direct measure of risk</td>
<td>Allows checks for Mendelian inheritance patterns in genotyping quality control</td>
<td></td>
</tr>
<tr>
<td>Optimal epidemiologic design for studying rare diseases</td>
<td>Fewer biases than case-control studies</td>
<td>Logistically simpler for studies of children’s conditions</td>
<td></td>
</tr>
<tr>
<td>Overestimate relative risk for common diseases</td>
<td>Continuum of health-related measures available in population samples not selected for presence of disease</td>
<td>Does not require phenotyping of parents</td>
<td></td>
</tr>
<tr>
<td>Prone to a number of biases including population stratification</td>
<td>Large sample size needed for genotyping if incidence is low</td>
<td>May be difficult to assemble both parents and offspring, especially in disorders with older ages of onset</td>
<td></td>
</tr>
<tr>
<td>Cases are usually prevalent cases, may exclude fatal or short episodes, or mild or silent cases</td>
<td>Expensive and lengthy follow-up</td>
<td>Highly sensitive to genotyping error</td>
<td></td>
</tr>
<tr>
<td>Overestimate relative risk for common diseases</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Selection of Study Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Many genetic studies, whether GWA or otherwise, focus on case participants more likely to have a genetic basis for their disease, such as early-onset cases or those with multiple affected relatives. Misclassification of case participants can markedly reduce study power and bias study results toward no association, particularly when large numbers of unaffected individuals are misclassified as affected. For diseases that are difficult to diagnose reliably, ensuring that cases are truly affected (as by invasive testing or imaging), is probably more important than ensuring generalizability, although the limitations on diagnostic reliability and generalizability should be clearly described so that clinicians can judge the relevance to their patients. The control participants should be drawn from the same population as the case participants and should be at risk to develop the disease and be detected in the study. Inclusion of women as controls in genetic association studies of diseases limited to men, for example, is problematic in that this approach adds individuals to the control group who had no chance of developing the disease (but might have done so had they also inherited a Y chromosome), thus mixing the controls with possible latent cases. This artificially reduces the differences in allele frequencies between cases and controls and limits the ability of the study to detect a true difference (ie, reduces study power). If the disease is common, such as coronary heart disease or hypertension in the United States, efforts should be made to ensure that the controls are truly disease free. Some studies address this by using super-controls or persons at high risk but without even early evidence of disease, such as per-</td>
</tr>
</tbody>
</table>

©2008 American Medical Association. All rights reserved.
sons with diabetes of long duration but without microalbuminuria in a study of diabetic nephropathy. The success of recent GWA studies using control groups of questionable representativeness due to volunteer bias, such as the blood donor cohort in the Wellcome Trust Case-Control Consortium, suggests that initial identification of SNPs associated with disease may be robust to these biases, especially given subsequent evidence of replication of these associations in studies using more traditional control groups.

Of more concern may be the risk of false-negative findings, as many biases tend to reduce the magnitude of observed associations toward the null. Use of convenience controls such as blood donors, however, may also be problematic in examining potential modification of genetic associations by environmental exposures and sociocultural factors, and in the identification of less strongly associated SNPs.

A key component in articles reporting results in the epidemiology literature of observational study is an initial table comparing relevant characteristics of those with and without disease, allowing assessment of comparability and generalizability of the 2 groups. Such comparisons are infrequent in GWA studies, but they are important because common diseases are typically influenced by multiple environmental (as well as genetic) factors. Important differences should be adjusted for in the analysis if possible, to avoid the risk of identifying genetic associations not with the disease of interest but with a confounding factor, such as smoking or obesity.

Confounding due to population stratification (also called population structure) has been cited as a major threat to the validity of genetic association studies, but its true importance is a matter of debate. When variations occur in allele frequency between population subgroups, such as those defined by ethnicity or geographic origin, that in turn differ in their risk for disease, GWA studies may then falsely identify the subgroup-associated genes as related to disease. Population structure should be assessed and reported in GWA studies, typically by examining the distribution of test statistics generated from the thousands of association tests performed (eg, the $\chi^2$ test) and assessing their deviation from the null distribution (that expected under the null hypothesis of no SNP associated with the trait) in a quantile-quantile or “Q-Q,” plot (FIGURE 1). In these plots, observed association statistics or calculated $P$ values for each SNP are ranked in order from smallest to largest and plotted against the values expected had they been sampled from a distribution of known form (such as the $\chi^2$ distribution). Deviations from the diagonal identity line suggest that either the assumed distribution is incorrect or that the sample contains values arising from different distributions.

| Table 2. Examples of Multistage Designs in Genome-wide Association Studies |
|-----------------------------|-----------------------------|
| Case Participants/Control Participants | SNPs Analyzed |
| Stage | 3-Stage Study | 4-Stage Study |
| 1 | 400/400 | 500,000 | 2,000/2,000 | 100,000 |
| 2 | 400/400 | 25,000 | 2,000/2,000 | 1,000 |
| 3 | 20,000/20,000 | 25 | 2,000/2,000 | 20 |
| 4 | 20,000/2,000 | 5 |

Abbreviation: SNP, single-nucleotide polymorphism.

A, Observed association statistics (eg, $\chi^2$ or t statistics) or $\log_2(P)$ values calculated from them, are ranked in order from smallest to largest on the $y$-axis and plotted against the distribution that would be expected under the null hypothesis of no association on the $x$-axis. Deviations from the identity line suggest either that the assumed distribution is incorrect or that the sample contains values arising in some other manner, as by a true association. B, Observed $\chi^2$ statistics of all polymorphic SNPs (dark blue) in a hypothetical genome-wide association study of a complex disease vs. the expected null distribution (black line). The sharp deviation above an expected $\chi^2$ value of approximately 8 could be due to a strong association of the disease with SNPs in a heavily genotyped region such as the major histocompatibility locus (MHC) on chromosome 6p21 in multiple sclerosis or rheumatoid arthritis. Exclusion of SNPs from such a locus may leave a residual upward deviation (light blue) identifying more associated SNPs with higher observed $\chi^2$ values (exceeding approximately 17) than expected under the null hypothesis. B, Observed (dark purple) vs expected (black line) $\chi^2$ statistics for a hypothetical genome-wide association study of a complex disease. Deviation from the expected distribution is observed above an expected $\chi^2$ of approximately 5. Inflation of observed statistics due to relatedness and potential population structure can be estimated by the method of genomic control. Correction for this inflation by simple division reduces the unadjusted $\chi^2$ statistics (dark purple) to the adjusted levels (light purple), showing deviation only above an expected $\chi^2$ of approximately 15. The region between expected $\chi^2$ of approximately 5 to approximately 15 is suggestive of broad differences in allele frequencies that are more likely due to population structure than disease susceptibility genes.
1 000 000 SNPs have been estimated to capture 67% to 89% of common SNP variation in populations of European and Asian ancestry and 46% to 66% of variation in individuals of recent African ancestry.\(^3\) Higher density platforms now also include probes for copy number variants that are not well tagged by SNPs. Copy number variants, in which stretches of genomic sequence are deleted or are duplicated in varying numbers, have gained increasing attention because of their apparent ubiquity and potential dosage effect on gene expression.\(^3\) Newer genotyping platforms are increasingly being focused on capturing copy number variants, but other structural variants such as insertions, deletions, and inversions, remain difficult to assay.\(^2\)

GWA studies frequently identify associations with multiple SNPs in a chromosomal region and display the association statistics by their genomic location on a portion of a chromosome (FIGURE 2). For ease of display, association statistics are typically shown as the \(-\log_{10}\) of the \(P\) value (the probability of the observed association arising by chance alone), so that \(P = .01\) would be plotted as “2” on the \(y\)-axis and \(P = 10^{-7}\) as “7.” Such displays also often plot a matrix of \(r^2\) values for each pair of SNPs in the region, with larger \(r^2\) values more intensely shaded. These plots can be used to identify linkage disequilibrium blocks containing SNPs associated with disease, allowing estimation of the independence of the SNP associations observed.\(^5\)

Genotyping errors, especially if occurring differentially between cases and controls, are an important cause of spurious associations and must be diligently sought and corrected.\(^5\) A number of quality control features should be applied both on a per-sample and a per-SNP basis. Checks on sample identity to avoid sample mix-ups should be described and a minimum rate of successfully genotyped SNPs per sample (usually 80%-90% of SNPs attempted) should be reported. Once samples failing these thresholds are removed, individual SNPs across the region...
maintaining samples are subjected to further checks or filters for probable genotyping errors, including: (1) the proportion of samples for which a SNP can be measured (the SNP call rate, typically >95%); (2) the minor allele frequency (often >1%, as rarer SNPs are difficult to measure reliably; (3) severe violations of Hardy-Weinberg equilibrium; (4) Mendelian inheritance errors in trio studies; and (5) concordance rates in duplicate samples (typically >99.5%).

Additional checks on genotyping quality should include careful visual inspection of genotype cluster plots, or intensity values generated by the genotyping assay to ensure that the strongest associations do not merely reflect genotyping artifact.28,39 Genotyping the most strongly associated SNPs should also be confirmed using a different method.28 Associations with any known “positive controls,” such as TCF7L2 in type 2 diabetes mellitus35 or HLA-DRB1 in rheumatoid arthritis,47 should be reported to increase confidence in the consistency of findings with prior reports.

Analysis and Presentation of GWA Results

Associations with the 2 alleles of each SNP are tested in a relatively straightforward manner by comparing the frequency of each allele in cases and controls (Table 3). Because each individual carries 2 copies of each autosomal SNP, the frequency of each of 3 possible genotypes can also be compared (Table 3). Exploratory analyses may also include testing of different genetic models (dominant, recessive, or additive), although additive models, in which each copy of the allele is assumed to increase risk by the same amount, tend to be the most common (T.A.M.; unpublished data, 2008). Odds ratios of disease associated with the risk allele or genotype(s) can then be calculated and are typically modest, often in the range of 1.2 to 1.3. Many studies also calculate population attributable risk, classically defined as the proportion of disease in the population associated with a given risk factor (in this case, a genetic variant).57

Such estimates are nearly always inflated because odds ratios overestimate relative risks (especially for common diseases58) needed for population attributable risk calculations, and because odds ratios and allele frequencies in published reports have wide confidence intervals so that those selected by exceeding a specified threshold for statistical significance tend to be biased upwards, an effect of ascertainment known as the “winner’s curse.”59

This exaggerated initial estimate of the odds ratio often leads to replication studies that lack sufficient sample size and power to replicate the association because larger samples are needed to detect smaller odds ratios.

Complexity in analysis emerges due to the multiple testing carried out in GWA studies, in that the association tests shown in Table 2 are repeated for each of the 100 000 to more than 1 million SNPs assayed (Figure 3). At the conventional P < .05 level of significance, an association study of 1 million SNPs will show 50 000 SNPs to be “associated” with disease, almost all

©2008 American Medical Association. All rights reserved.
**Box 2. Ten Basic Questions to Ask About a Genome-wide Association Study Report**

1. Are the cases defined clearly and reliably so that they can be compared with patients typically seen in clinical practice?
2. Are case and control participants demonstrated to be comparable to each other on important characteristics that might also be related to genetic variation and to the disease?
3. Was the study of sufficient size to detect modest odds ratios or relative risks (1.3-1.5)?
4. Was the genotyping platform of sufficient density to capture a large proportion of the variation in the population studied?
5. Were appropriate quality control measures applied to genotyping assays, including visual inspection of cluster plots and replication on an independent genotyping platform?
6. Did the study reliably detect associations with previously reported and replicated variants (known positives)?
7. Were stringent corrections applied for the many thousands of statistical tests performed in defining the P value for significant associations?
8. Were the results replicated in independent population samples?
9. Were the replication samples comparable in geographic origin and phenotype definition, and if not, did the differences extend the applicability of the findings?
10. Was evidence provided for a functional role for the gene polymorphism identified?

For a more detailed description of interpretation of genome-wide association studies, see NCI/NHGRI Working Group on Replication in Association Studies.28

---

falsely positive and due to chance alone. The most common manner of dealing with this problem is to reduce the false-positive rate by applying the Bonferroni correction, in which the conventional P value is divided by the number of tests performed.60 A 1 million SNP survey would thus use a threshold of $P < 0.05/10^6$, or $5 \times 10^{-8}$, to identify associations unlikely to have occurred by chance. This correction has been criticized as overly conservative because it assumes independent associations of each SNP with disease even though individual SNPs are known to be correlated to some degree due to linkage disequilibrium.

Other approaches have been proposed, including estimation of the false discovery rate or proportion of significant associations that are actually false positive associations.61,62 False-positive report probability, or probability that the null hypothesis is true given a statistically significant finding,63 and estimation of Bayes factors that incorporate the prior probability of association based on characteristics of the disease or the specific SNP.69 To date, Bonferroni correction has generally been the most commonly used correction for multiple comparisons in GWA reports (T.A.M.; unpublished data, 2008).

**Replication and Functional Studies**

Given the major challenge of separating the many false-positive associations from the few true-positive associations with disease in GWA studies, an important strategy has been replication of results in independent samples.28 This is typically included in a single GWA report as part of a multistage design34,35 or may be reported separately.39,69 Consensus criteria for replication have recently been published and include study of the same or very similar phenotype and population, and demonstration of a similar magnitude of effect and significance (in the same genetic model and same direction) for the same SNP and the same allele as the initial report.28 Replication is usually first attempted in studies as similar as possible to the initial report, but then may be extended to related phenotypes (such as fat mass in addition to obesity64), different populations (such as West Africans in addition to Icelanders65), or different study designs53 to refine and extend the initial findings and increase confidence in verity.

Lack of reproducibility of genetic associations has been frequently observed and has been varyingly attributed to population stratification, phenotype differences, selection biases, genotyping errors, and other factors.28,66 At present, the best way of resolving these inconsistencies appears to be additional replication studies with larger sample sizes, although this may not be feasible for rare conditions or for associations identified in unique populations.28

Identification of a robustly replicating SNP-disease association is a crucial first step in identifying disease-causing genetic variants and developing suitable treatments, but it is only a first step. Association studies essentially identify a genomic location related to disease but provide little information on gene function unless SNPs with predictable effects on gene expression or the transcribed product happened to be identified. Few of the associations identified to date have involved genes previously suspected of being related to the disease under study, and some have been in genomic locations harboring no known genes.27,67 Examination of known SNPs in high linkage disequilibrium with the associated SNP may identify variants with plausible biologic effects, or sequencing of a suitable surrounding interval may be undertaken to identify rarer variants with more obvious functional implications. Tissue samples or cell lines can be examined for expression of the gene variant. Other functional studies may include genetic manipulations in cell or animal models, such as knockouts or knock-ins.98
Limitations of GWA Studies

The potential for false-positive results, lack of information on gene function, insensitivity to rare variants and structural variants, requirement for large sample sizes, and possible biases due to case and control selection and genotyping errors, are important limitations of GWA studies. The often limited information available about environmental exposures and other non-genetic risk factors in GWA studies will make it difficult to identify gene-environment interactions or modification of gene-disease associations in the presence of environmental factors. Clinicians and scientists should understand the unique aspects of these studies and be able to assess and interpret GWA results for themselves and their patients. Ten basic questions to ask about GWA studies, many of which also apply generically to association studies of nongenetic risk factors, are outlined in Box 2. Most of these questions should be answered in the affirmative for a reliable report; however, many GWA reports lack sufficient detail to assess them.28

Many of the design and analysis features of GWA studies deal with minimizing the false-positive rates while maintaining power to identify true-positive associations. These same efforts to reduce false-positive results, however, may result in overlooking a true association, especially if only a small number of SNPs are carried over from the initial scan into replication studies. The most robust findings, ie, those that “survive” multiple rounds of replication, are often not the most statistically significant associations in the initial scan, and may not even be in the top few hundred associations.69,70 Another cause of false-negative results is the lack of the genetic variant of relevance on the genotyping platform, or lack of variation in that SNP in the population under study. As the number of SNPs and diversity of populations represented on genotyping platforms increase, this should become less of a problem.

An important question generated by these early GWA studies relates to the small proportion of heritability, or familial clustering explained by the genetic variants identified to date. Most of these variants have very modest effects on disease risk, increasing it by only 20% to 50%, and explaining only a small fraction of population risk or total estimated heritability for most conditions.39,71 Might the rest of the genetic influence reside in a long “tail” of common SNPs with very small odds ratios, in copy number variants or other structural variants, rarer variants of larger effect, or interactions among common variants? Or has familial clustering due to genetic factors been overestimated and important environmental influences, either acting alone or in combination with genetic variants, been overlooked? This remains to be determined, but it is important to realize that even small odds ratios or rare variants can suggest important therapeutic strategies such as the development of HMG-CoA reductase inhibitors arising from identification of LDL-receptor mutations in familial hypercholesterolemia.72

Clinical Applications of GWA Findings

Despite the considerable media attention that GWA reports frequently receive, these studies are clearly many steps removed from actual clinical application. The primary use for GWA studies for the foreseeable future is likely to be in investigation of biologic pathways of disease causation and normal health and development. This is not to suggest that some early successes may not occur in the near future, through rapid development of treatment strategies such as inhibitors of complement activation in age-related macular degeneration.73 Use of GWA findings in screening for disease risk, while beginning to be marketed commercially, is more problematic. Although obtaining the latest “gene test” may be alluring to a technology-focused society, evidence is needed that such screening adds information to known risk factors (such as age, obesity, and family history for diabetes), that effective interventions are available, that improved outcomes justify the associated costs, and that obtaining this information does not have serious adverse consequences for patients and their families. Such evidence is likely to be some ways off, but the initial burst of discovery generated by GWA scans has now mandated a concerted effort to search for these answers.

REFERENCES

INTERPRETING A GENOME-WIDE ASSOCIATION STUDY


conduct such studies are likely to assume that their work falls under this category. Conversely, public health and health services researchers might be confused to see the term applied to their work.

Furthermore, using “clinical” in both terms perpetuates the tendency of the medical profession to view health research through the clinician’s lens alone. Fiscella et al do include “organizational- and community-focused” research within their definition of applied clinical research, but labeling health interventions outside the clinic as “clinical” research may be a forced fit. Pros and cons exist with other potential terms such as knowledge translation—the term discussed by Dr Graham and Ms Tetroe—but all of them are an improvement over the ambiguity of T2.

Graham and Tetroe call attention to the excellent work of the Canadian Institutes of Health Research. Canadian investigators and institutions have played a leadership role not only in writing about the need for researchers to align their work with the information needs of end users1 but also in making real commitments in programs and funding to facilitate T2 as a nation.2 The United States would do well to follow the Canadian example.

T1 is among a group of clinical research movements that are attracting attention and resources but are ultimately unhelpful to patients without T2. Recently, politicians and industry have announced plans to channel millions of dollars per year into research on “comparative effectiveness”3 and “personalized medicine”4 while keeping funding for health services research threadbare.5 The term discussed by Dr Graham and Ms Tetroe—but all of them are an improvement over the ambiguity of T2.

Steven H. Woolf, MD, MPH
swoolf@vcu.edu
Department of Family Medicine
Virginia Commonwealth University
Richmond

Financial Disclosures: None reported.