How to Interpret a Genome-wide Association Study

Thomas A. Pearson; Teri A. Manolio

JAMA. 2008;299(11):1335-1344 (doi:10.1001/jama.299.11.1335)

http://jama.ama-assn.org/cgi/content/full/299/11/1335

How to Interpret a Genome-wide Association Study

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

Genome-wide association (GWA) studies use high-throughput genotyping technologies to assay hundreds of thousands of single-nucleotide polymorphisms (SNPs) and relate them to clinical conditions and measurable traits. Since 2005, nearly 100 loci for as many as 40 common diseases and traits have been identified and replicated in GWA studies, many in genes not previously suspected of having a role in the disease under study, and some in genomic regions containing no known genes. GWA studies are an important advance in discovering genetic variants influencing disease but also have important limitations, including their potential for false-positive and false-negative results and for biases related to selection of study participants and genotyping errors. Although these studies are clearly many steps removed from actual clinical use, and specific applications of GWA findings in prevention and treatment are actively being pursued, at present these studies mainly represent a valuable discovery tool for examining genomic function and clarifying pathophysiologic mechanisms. This article describes the design, interpretation, application, and limitations of GWA studies for clinicians and scientists for whom this evolving science may have great relevance.

JAMA. 2008;299(11):1335-1344 www.jama.com

In the past 2 years, there has been a dramatic increase in genomic discoveries involvng 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for as many as 40 common diseases robustly identified and replicated in genome-wide association (GWA) studies (T.A.M.; unpublished data, 2008).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.[1] Nearly 12 million unique human SNPs have been assigned a reference SNP (rs) number in the National Center for Biotechnology Information's dbSNP database[2] and characterized as to specific alleles (alternate forms of the SNP), summary allele frequencies, and other genomic information.[3]

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.[4] 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.[5]

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.[6] Many such associations failed to be replicated in subsequent studies,[7,8] leading to calls for all genetic association reports to include documented replication of findings as a prerequisite for publication.[9,10]

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.[11] 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

Author Affiliations: Office of Population Genomics, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland (Drs Pearson and Manolio); and Clinical and Translational Science Institute, University of Rochester Medical Center, Rochester, New York (Dr Pearson).
Corresponding Author: Teri A. Manolio, MD, PhD, Office of Population Genomics, National Human Genome Research Institute, 31 Center Dr, Room 4B-09, MSC 2154, Bethesda, MD 20892-2154 (manolio@nih.gov).

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,[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.[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.[15] Although family linkage studies and studies comprising tens of thousands of gene-based SNPs also assay genetic variation across the genome,[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.[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[17] or electrocardiographic QT interval,[18] and to rank the relative importance of previously identified susceptibility genes, such as *APOE*e4* in Alzheimer disease[19] and *CARD15* and *IL23R* in Crohn disease.[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,[21] and can detect high-risk haplotypes or combinations of multiple SNPs within a single gene, as in exfoliation glaucoma[22] and atrial fibrillation.[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,[24] or more globally.[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.[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[27] can make evaluation of potential biases and replication of findings quite difficult.[28]

The trio design includes the affected case participant and both of his or her parents.[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.[29] 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[59,60]
Proportion of significant associations that are actually false positives

False-positive report probability[61]
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 genetic 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[12,13]
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

Mendelian disease
Condition caused almost entirely by a single major gene, such as cystic fibrosis or Huntington's disease, in which disease is manifested in only 1 (recessive) or 2 (dominant) of the 3 possible genotype groups

Minor allele
The allele of a biallelic polymorphism that is less frequent in the study population

Minor allele frequency
Proportion of the less common of 2 alleles in a population (with 2 alleles carried by each person at each autosomal locus) ranging from less than 1% to less than 50%

Modest effect
Association between a gene variant and disease or trait that is statistically significant but carries a small odds ratio (usually <1.5)

Non-Mendelian disease (also "common" or "complex" disease)
Condition influenced by multiple genes and environmental factors and not showing Mendelian inheritance patterns

Nonsynonymous SNP
A polymorphism that results in a change in the amino acid sequence of a protein (and therefore may affect the function of the protein)

Platform
Arrays or chips on which high-throughput genotyping is performed

Polymorphic
A gene or site with multiple allelic forms. The term *polymorphism* usually implies a minor allele frequency of at least 1%

Population attributable risk
Proportion of a disease or trait in the population that is due to a specific cause, such as a genetic variant

Population stratification (also "population structure")
A form of confounding in genetic association studies caused by genetic differences between cases and controls unrelated to disease but due to sampling them from populations of different ancestries

Power
A statistical term for the probability of identifying a difference between 2 groups in a study when a difference truly exists

Single-nucleotide polymorphism
Most common form of genetic variation in the genome, in which a single-base substitution has created 2 forms of a DNA sequence that differ by a single nucleotide

Tag SNP
A readily measured SNP that is in strong linkage disequilibrium with multiple other SNPs so that it can serve as a proxy for these SNPs on large-scale genotyping platforms

Trio
Genetic study design including an affected offspring and both parents

Abbreviation: SNP, single-nucleotide polymorphism.

pothesis of no association with disease, the transmission frequency for each allele of a given SNP will be 50%, but alleles associated with the disease will be transmitted in excess to the affected case individual. Because the trio design studies allele transmission from parents to offspring, it is not susceptible to population stratification, or genetic differences between case and control participants unrelated to disease but due to sampling them from populations of different ancestry.[30] A significant challenge of the trio design in GWA studies is its sensitivity to even small degrees of genotyping error,[4,31] which can distort transmission proportions between parents and offspring, especially for uncommon alleles. Therefore, standards for genotyping quality in trio studies may need to be more stringent than for other designs.

Cohort studies involve collecting extensive baseline information in a large number of individuals who are then observed to assess the incidence of disease in subgroups defined by

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

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

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.[17,18] For these reasons, genome-wide genotyping has recently been added to cohort studies such as the Framingham Heart Study[32] and the Women's Health Study.[33]

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.[4] 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.[34] Other studies begin with more participants but carry forward a smaller proportion of associated SNPs.[35] Optimal proportions of study participants and SNPs in each phase have yet to be determined,[36] 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.[37]

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-

sons with diabetes of long duration but without microalbuminuria in a study of diabetic nephropathy.[38] 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,[39] 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.[40-42]

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,[28] 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[43] or obesity.[44]

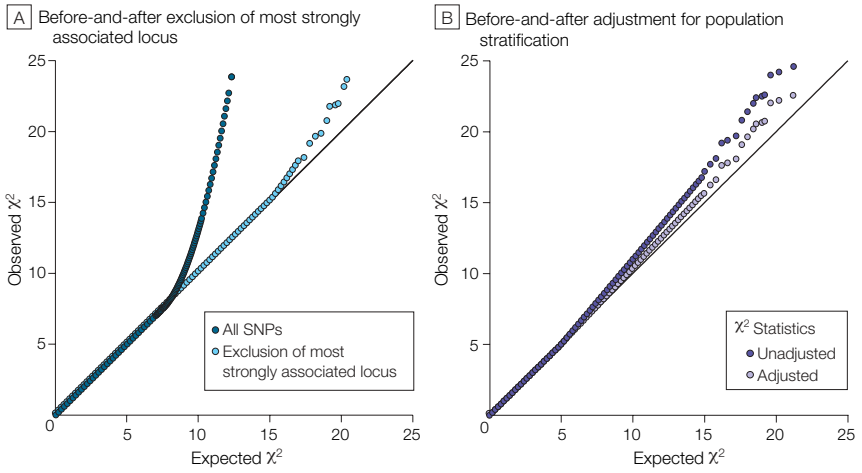
Confounding due to population stratification (also called population structure) has been cited as a major threat to the validity of genetic association studies, but its true importance is a matter of debate.[45,46] When variations occur in allele frequency between population subgroups, such as those 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.[30] 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 χ^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 χ^2 distribution).[39] Deviations from the diagonal identity line suggest that either the assumed distribution is incorrect or that the sample contains values arising

Table 2. Examples of Multistage Designs in Genome-wide Association Studies[a]

	3-Stage Study[b]		4-Stage Study[c]	
Stage	Case Participants/ Control Participants	SNPs Analyzed	Case Participants/ Control Participants	SNPs Analyzed
1	400/400	500 000	2000/2000	100 000
2	4000/4000	25 000	2000/2000	1000
3	20 000/20 000	25	2000/2000	20
4			2000/2000	5

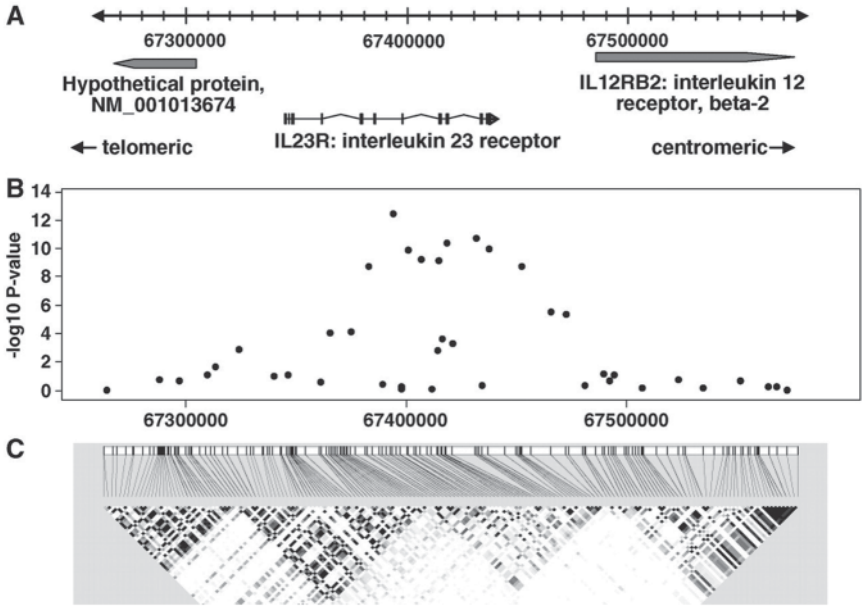
Abbreviation: SNP, single-nucleotide polymorphism.
[a] Based on hypothetical data.
[b] Five SNPs associated with disease.
[c] Two SNPs associated with disease.

Figure 1. Hypothetical Quantile-Quantile Plots in Genome-wide Association Studies



The Q-Q plot is used to assess the number and magnitude of observed associations between genotyped single-nucleotide polymorphisms (SNPs) and the disease or trait under study, compared to the association statistics expected under the null hypothesis of no association.[39] Observed association statistics (eg, χ^2 or *t* statistics) or $-\log_{10}$ *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.[39] A, Observed χ^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 χ^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.[70] Exclusion of SNPs from such a locus may leave a residual upward deviation (light blue) identifying more associated SNPs with higher observed χ^2 values (exceeding approximately 17) than expected under the null hypothesis. B, Observed (dark purple) vs expected (black line) χ^2 statistics for a hypothetical genome-wide association study of a complex disease. Deviation from the expected distribution is observed above an expected χ^2 of approximately 5. Inflation of observed statistics due to relatedness and potential population structure can be estimated by the method of genomic control.[49] Correction for this inflation by simple division reduces the unadjusted χ^2 statistics (dark purple) to the adjusted levels (light purple), showing deviation only above an expected χ^2 of approximately 15. The region between expected χ^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.

Figure 2. Associations in the *IL23R* Gene Region Identified by a Genome-wide Association Study of Inflammatory Bowel Disease



Genome-wide association studies frequently identify associations with many highly correlated single-nucleotide polymorphisms (SNPs) in a chromosomal region, due in part to linkage disequilibrium, among the SNPs. This can make it difficult to determine which SNP within a group is likely to be the causative or functional variant. A, Genomic locations of 2 genes, the interleukin 23 receptor (*IL23R*) and the interleukin 12 receptor, beta-2 (*IL12Rb2*), and a hypothetical protein, NM_001013674, between positions 62700000 and 67580000 of the short arm of chromosome 1 at region 1p31, are shown. B, The $-\log_{10} P$ values for association with inflammatory bowel disease are plotted for each SNP genotyped in the region; those reaching a prespecified value of $-\log_{10}$ of 7 or greater are presumed to show association with disease. Several strong associations, at $-\log_{10} P$ values or greater, are seen in the region just telomeric of position approximately 67400000 and extending just centromeric of position approximate 67450000. C, Pairwise linkage disequilibrium estimates between SNPs (measured as r^2) are plotted for the region. Higher r^2 values are indicated by darker shading. The region contains 4 "triangles" or "blocks" of linkage disequilibrium, 2 on either side of position 67400000 in the *IL23R* gene, another in the hypothetical protein telomeric of *IL23R,* and a fourth in the *IL12RB2* gene at the centromeric end of the region. The 2 *IL23R* linkage disequilibrium regions each contain SNPs associated with inflammatory bowel disease, while the *IL12RB2* region does not. Reproduced with permission from Duerr et al.[53]

in some other manner, as by a true association.[39]

Since the underlying assumption in GWA studies is that the vast majority of assayed SNPs are not associated with the trait, strong deviations from the null suggest either a very highly associated and heavily genotyped locus (Figure 1, A), or significant differences in population structure (Figure 1, B). Several effective statistical methods are available to correct for population structure and are a standard component of rigorous GWA analyses.[28,30]

Genotyping and Quality Control in GWA Studies

GWA studies rely on the typically strong associations among SNPs located near each other on a chromosome, which tend to be inherited together more often than expected by chance.[50] This nonrandom association is called linkage disequilibrium; alleles of SNPs in high linkage disequilibrium are almost always inherited together and can serve as proxies for each other. Their correlation with each other in the population is measured by the r^2 statistic, which is the proportion of variation of one SNP explained by the other, and ranges from 0 (no association) to 1 (perfect correlation).

Genomic coverage of GWA genotyping platforms (arrays or chips on which genotyping is performed) is often estimated by the percent of common SNPs having an r^2 of 0.8 or greater with at least 1 SNP on the platform.[13] Genotyping platforms comprising 500 000 to 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.[13] 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.[51] 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.[52]

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.[53]

Genotyping errors, especially if occurring differentially between cases and controls, are an important cause of spurious associations and must be diligently sought and corrected.[54] 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 re-

Table 3. Association of Alleles and Genotypes of rs6983267 on Chromosome 8q24 With Colorectal Cancer[a]

	Number and Frequency of rs6983267 Alleles in Colorectal Cancer					Number and Frequency of rs6983267 Genotypes in Colorectal Cancer						
	C	T	χ^2 (1*df*)	*P* Value	OR	CC	CT	TT	χ^2 (2*df*)	*P* Value	OR	OR
Cases	875 (56.5)	675 (43.5)	24.8	6.3×10^{-7}	1.35[b]	250 (32.3)	375 (48.4)	150 (19.4)	24.5	4.7×10^{-6}	1.33[c]	1.81[d]
Controls	1860 (48.9)	1940 (51.1)				460 (24.2)	940 (49.4)	500 (26.3)				

Abbreviation: OR, odds ratio.
[a] Data are hypothetical; adapted from Tomlinson et al.[56]
[b] Denotes allelic odds ratio.
[c] Denotes heterozygote odds ratio.
[d] Denotes homozygote odds ratio.

maining 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 mellitus[55] or *HLA-DRB1* in rheumatoid arthritis,[47] should be reported to increase confidence in the consistency of findings with prior reports.

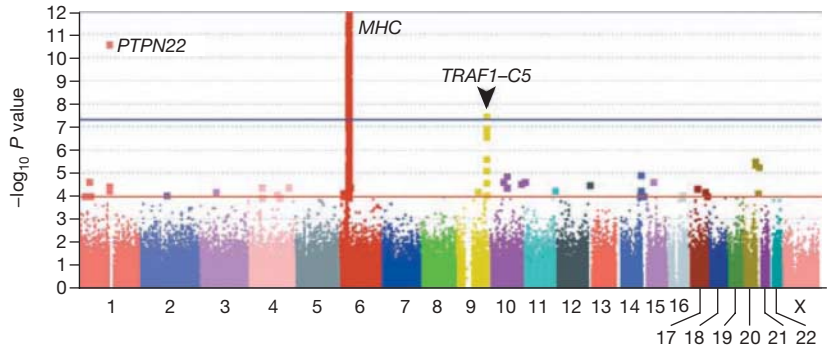
Analysis and Presentation of GWA Results

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 diseases[58]) 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

Figure 3. Genome-wide Association Findings in Rheumatoid Arthritis



Genome-wide association studies assume a priori hypotheses about candidate genes or regions that might be associated with disease; rather, they test single-nucleotide polymorphisms (SNPs) throughout the genome for possible evidence of genetic susceptibility. Associations plotted as $-\log_{10} P$ values for a genome-wide association study in 1522 cases with rheumatoid arthritis and 1850 controls, showing single data points for SNPs with $P < 10^{-4}$ (lower horizontal red line) for 22 autosomes and the X chromosome. The predefined level of significance, at 5×10^{-8} is shown with a horizontal blue line. SNPs at *PTPN22* on chromosome 1, the major histocompatibility comples (MHC) on chromosome 6, and the *TRAF1-C5* locus on chromosome 9 exceed this threshold. Reproduced with permission from Plenge et al.[47]

Box 2. Ten Basic Questions to Ask About a Genome-wide Association Study Report[a]

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?

[a]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 $P < .05/10^6$, or 5×10^{-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.[39] 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 design[34,35] or may be reported separately.[39,64] 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 obesity[44]), different populations (such as West Africans in addition to Icelanders[65]), or different study designs[53] 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.[68]

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.

Author Contributions: Dr Manolio had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Financial Disclosures: None reported.
Funding/Support: This work supported in part by a Clinical and Translational Science Award (RR024160) from the National Center for Research Resources, National Institutes of Health, to the University of Rochester.
Additional Contributions: The authors wish to acknowledge the valuable comments of Francis Collins, MD, PhD, National Human Genome Research Institute, National Institutes of Health; and Stephen Chanock, MD, National Cancer Institute, National Institutes of Health; and the contributions of Maureen Marcello, Clinical and Translational Science Institute, University of Rochester, in the preparation of this article. None of these individuals received compensation for their work in association with this article.

REFERENCES

1. Christensen K, Murray JC. What genome-wide association studies can do for medicine. *N Engl J Med*. 2007;356(11):1094-1097.
2. National Center for Biotechnology Information, National Library of Medicine. Database of Single Nucleotide Polymorphisms. http://www.ncbi.nlm.nih.gov/SNP/. Accessed February 14, 2008.
3. Wheeler DL, Barrett T, Benson DA, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2008;36(database issue):D13-D21.
4. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*. 2005;6(2):95-108.
5. Hunter DJ, Kraft P. Drinking from the fire hose–statistical issues in genomewide association studies. *N Engl J Med*. 2007;357(5):436-439.
6. Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits. *Nat Rev Genet*. 2002;3(5):391-397.
7. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. *Genet Med*. 2002;4(2):45-61.
8. Morgan TM, Krumholz HM, Lifton RP, Spertus JA. Nonvalidation of reported genetic risk factors for acute coronary syndrome in a large-scale replication study. *JAMA*. 2007;297(14):1551-1561.
9. Todd JA. Statistical false positive or true disease pathway? *Nat Genet*. 2006;38(7):731-733.
10. Patterson M, Cardon L. Replication publication. *PLoS Biol*. 2005;3(9):e327.
11. Altmüller J, Palmer LJ, Fischer G, Scherb H, Wjst M. Genome-wide scans of complex human diseases. *Am J Hum Genet*. 2001;69(5):936-950.
12. International HapMap Consortium. A haplotype map of the human genome. *Nature*. 2005;437(7063):1299-1320.
13. Frazer KA, Ballinger DG, Cox DR, et al; International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 2007;449(7164):851-861.
14. Collins FS, Guyer MS, Chakravarti A. Variations on a theme: cataloging human DNA sequence variation. *Science*. 1997;278(5343):1580-1581.

15. National Institutes of Health. Policy for sharing of data obtained in NIH supported or conducted genome-wide association studies (GWAS). Federal Regist. 2007;72(166):49290-49297. http://www.grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html. Accessed February 14, 2007.
16. Hampe J, Franke A, Rosenstiel P, et al. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet*. 2007;39(2):207-211.
17. Weedon MN, Lettre G, Freathy RM, et al. A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nat Genet*. 2007;39(10):1245-1250.
18. Arking DE, Pfeufer A, Post W, et al. A common genetic variant in the NOS1 regulator NOS1AP modulates cardiac repolarization (QT interval). *Nat Genet*. 2006;38(6):644-651.
19. Coon KD, Myers AJ, Craig DW, et al. A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease. *J Clin Psychiatry*. 2007;68(4):613-618.
20. Rioux JD, Xavier RJ, Taylor KD, et al. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet*. 2007;39(5):596-604.
21. Reiman EM, Webster JA, Myers AJ, et al. GAB2 alleles modify Alzheimer's risk in APOE ε4 carriers. *Neuron*. 2001;54(5):713-720.
22. Thorleifsson G, Magnusson KP, Sulem P, et al. Common sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma. *Science*. 2007;317(5843):1397-1400.
23. Gudbjartsson DF, Arnar DO, Helgadottir A, et al. Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature*. 2007;448(7151):353-357.
24. Moffatt MF, Kabesch M, Liang L, et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature*. 2007;448(7152):470-473.
25. Dixon AL, Liang L, Moffatt MF, et al. A genome-wide association study of global gene expression. *Nat Genet*. 2007;39(10):1202-1207.
26. Manolio TA, Bailey-Wilson JE, Collins FS. Genes, environment and the value of prospective cohort studies. *Nat Rev Genet*. 2006;7(10):812-820.
27. Libioulle C, Louis E, Hansoul S, et al. Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genet*. 2007;3(4):e58.
28. Chanock SJ, Manolio T, Boehnke M, et al; NCI-NHGRI Working Group on Replication in Association Studies. Replicating genotype-phenotype associations. *Nature*. 2007;447(7145):655-660.
29. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium. *Am J Hum Genet*. 1993;52(3):506-516.
30. Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet*. 2003;361(9357):598-604.
31. Mitchell AA, Cutler DJ, Chakravarti A. Undetected genotyping errors cause apparent overtransmission of common alleles in the transmission/disequilibrium test. *Am J Hum Genet*. 2003;72(3):598-610.
32. Cupples LA, Arruda HT, Benjamin EJ, et al. The Framingham Heart Study 100K SNP genome-wide association study resources. *BMC Med Genet*. 2007;8(suppl 1):S1.
33. Ridker PM, Chasman DI, Zee RY, et al. Rationale, design, and methodology of the Women' Genome Health Study. *Clin Chem*. 2008;54(2):249-255.
34. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*. 2007;447(7148):1087-1093.
35. Zanke BW, Greenwood CM, Rangrej J, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus or a chromosome 8q24. *Nat Genet*. 2007;39(8):989-994.
36. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Optimal designs for two-stage genome-wide association studies. *Genet Epidemiol*. 2007;31(7):776-788.
37. Hoover RN. The evolution of epidemiologic research. *Epidemiology*. 2007;18(1):13-17.
38. Mueller PW, Rogus JJ, Cleary PA, et al. Genetics of Kidneys in Diabetes (GoKinD) study. *J Am Soc Nephrol*. 2006;17(7):1782-1790.
39. Wellcome Trust Case Control Consortium. Genome-wide association study of 14 000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661-678.
40. Hakonarson H, Grant SF, Bradfield JP, et al. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature*. 2007;448(7153):591-594.
41. Samani NJ, Erdmann J, Hall AS, et al; WTCCC and the Cardiogenics Consortium. Genomewide association analysis of coronary artery disease. *N Engl J Med*. 2007;357(5):443-453.
42. Scott LJ, Mohlke KL, Bonnycastle LL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*. 2007;316(5829):1341-1345.
43. Dewan A, Liu M, Hartman S, et al. HTRA1 promoter polymorphism in wet age-related macular degeneration. *Science*. 2006;314(5801):989-992.
44. Frayling TM, Timpson NJ, Weedon MN, et al. A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*. 2007;316(5826):889-894.
45. Thomas DC, Witte JS. Point: population stratification: a problem for case-control studies of candidate-gene associations? *Cancer Epidemiol Biomarkers Prev*. 2002;11(6):505-512.
46. Wacholder S, Rothman N, Caporaso N. Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. *Cancer Epidemiol Biomarkers Prev*. 2002;11(6):513-520.
47. Plenge RM, Seielstad M, Padyukov L, et al. TRAF1–C5 as a risk locus for rheumatoid arthritis—a genome-wide study. *N Engl J Med*. 2007;357(12):1199-1209.
48. Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet*. 2007;39(7):865-869.
49. Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999;55(4):997-1004.
50. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. *Science*. 2002;296(5576):2225-2229.
51. McCarroll SA, Altshuler DM. Copy-number variation and association studies of human disease. *Nat Genet*. 2007;39(7)(suppl):S37-S42.
52. Feuk L, Marshall CR, Wintle RF, Scherer SW. Structural variants. *Hum Mol Genet*. 2006;15(spec no 1):R57-R66.
53. Duerr RH, Taylor KD, Brant SR, et al. A genome-wide association study identifies *IL23R* as an inflammatory bowel disease gene. *Science*. 2006;314(5804):1461-1463.
54. Moskvina V, Craddock N, Holmans P, et al. Effects of differential genotyping error rate on the type I error probability of case-control studies. *Hum Hered*. 2006;61(1):55-64.
55. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*. 2007;445(7130):881-885.
56. Tomlinson I, Webb E, Carvajal-Carmona L, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q2421. *Nat Genet*. 2007;39(8):984-988.
57. Last JM, ed. *Dictionary of Epidemiology*. New York, NY: Oxford University Press; 1983:7.
58. Gordis L, ed. *Epidemiology*. 2nd ed. Philadelphia, PA: WB Saunders Co; 2000:165.
59. Zollner S, Pritchard JK. Overcoming the winner's curse: estimating penetrance parameters from case-control data. *Am J Hum Genet*. 2007;80(4):605-615.
60. Yang Q, Cui J, Chazaro I, Cupples LA, Demissie S. Power and type I error rate of false discovery rate approaches in genome-wide association studies. *BMC Genet*. 2005;6(suppl 1):S134.
61. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med*. 1990;9(7):811-818.
62. Sabatti C, Service S, Freimer N. False discovery rate in linkage and association genome screens for complex disorders. *Genetics*. 2003;164(2):829-833.
63. Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false. *J Natl Cancer Inst*. 2004;96(6):434-442.
64. Todd JA, Walker NM, Cooper JD, et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet*. 2007;39(7):857-864.
65. Helgason A, Pálsson S, Thorleifsson G, et al. Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. *Nat Genet*. 2007;39(2):218-225.
66. Khoury MJ, Little J, Gwinn M, Ioannidis JPA. On the synthesis and interpretation of consistent but weak gene-disease associations in the era of genome-wide association studies. *Int J Epidemiol*. 2007;36(2):439-445.
67. Yeager M, Orr N, Hayes RB, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet*. 2007;39(5):645-649.
68. Frayling TM, McCarthy MI. Genetic studies of diabetes following the advent of the genome-wide association study: where do we go from here? *Diabetologia*. 2007;50(11):2229-2233.
69. Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet*. 2007;39(7):870-874.
70. Hafler DA, Compston A, Sawcer S, et al; The International Multiple Sclerosis Genetics Consortium. Risk alleles for multiple sclerosis identified by a genome-wide study. *N Engl J Med*. 2007;357(9):851-862.
71. Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*. 2007;316(5829):1331-1336.
72. Slater EE, MacDonald JS. Mechanism of action and biological profile of HMG CoA reductase inhibitors: a new therapeutic alternative. *Drugs*. 1988;36(suppl 3):72-82.
73. Gehrs KM, Anderson DH, Johnson LV, Hageman GS. Age-related macular degeneration–emerging pathogenetic and therapeutic concepts. *Ann Med*. 2006;38(7):450-471.

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 users[1] 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] Popular research initiatives address worthy questions: whether a treatment can be produced (T1), whether it improves health (evidence-based medicine), which treatment is best (comparative effectiveness), and which is best for an individual patient (personalized medicine). But the answers remain academic if the patient cannot obtain or use the intervention. Overcoming such obstacles so that the products of research benefit all those in need is itself a crucial research priority.

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

Financial Disclosures: None reported.

1. Ross S, Lavis J, Rodriguez C, Woodside J, Denis JL. Partnership experiences: involving decision-makers in the research process. *J Health Serv Res Policy*. 2003; 8(suppl 2):26-34.
2. Lomas J. Using 'linkage and exchange' to move research into policy at a Canadian foundation. *Health Aff (Millwood)*. 2000;19(3):236-240.
3. US House of Representatives Committee on Ways and Means Subcommittee on Health: hearing on strategies to increase information on comparative clinical effectiveness [Tuesday, June 12, 2007]. http://waysandmeans.house.gov/hearings.asp?formmode=detail&hearing=565. Accessed February 29, 2008.
4. US Department of Health and Human Services: remarks prepared for the Honorable Mike Leavitt, Secretary of Health and Human Services, on personalized medicine coalition. http://www.hhs.gov/news/speech/2007/sp20070919a.html. Accessed February 29, 2008.
5. Coalition for Health Services Research: federal funding for health services research (as of December 26, 2007). http://www.chsr.org/2008.pdf. Accessed February 29, 2008.

CORRECTIONS

Incorrect Data: In the Perspectives on Care at the Close of Life article titled "Managing an Acute Pain Crisis in a Patient With Advanced Cancer: 'This Is as Much of a Crisis as a Code' " published in the March 26, 2008, issue of *JAMA* (2008;299 (12):1457-1467), an incorrect dose ratio appeared in Table 2. The hydromorphone-to-methadone ratio for less than 330 mg/24 hours of hydromorphone that read "16:1" should have been "1.6:1."

Incorrect Legend: In the Special Communication entitled "How to Interpret a Genome-wide Association Study" published in the March 19, 2008, issue of *JAMA* (2008;299[11]:1335-1344), an integral word was omitted from the Figure 3 legend. The sentence that read, "Genome-wide association studies assume a priori hypotheses about candidate genes or regions that might be associated with disease; rather, they test single-nucleotide polymorphisms (SNPs) throughout the genome for possible evidence of genetic susceptibility" should have read, "Genome-wide association studies assume no a priori hypotheses about candidate genes or regions that might be associated with disease; rather, they test single-nucleotide polymorphisms (SNPs) throughout the genome for possible evidence of genetic susceptibility."

Unreported Research Funding: In the Research Letter titled "Exhaled Carbon Monoxide With Waterpipe Use in US Students," published in the January 2, 2008, issue of *JAMA* (2008;299[1]:36-38), the Financial Disclosures should have included the following: Dr Hammond reports that she has received research funding for studies on environmental tobacco smoke from the National Institutes of Health and from the Flight Attendants Medical Research Institute. However, none of these grants were used to support the study reported in this Research Letter.