REVIEW ARTICLE

GENOMIC MEDICINE

W. Gregory Feero, M.D., Ph.D., and Alan E. Guttmacher, M.D., *Editors*

Genomewide Association Studies and Assessment of the Risk of Disease

Teri A. Manolio, M.D., Ph.D.

From the Office of Population Genomics, National Human Genome Research Institute, Bethesda, MD. Address reprint requests to Dr. Manolio at the Office of Population Genomics, National Human Genome Research Institute, Bldg. 31, Rm. 4B-09, 31 Center Dr., MSC 2152, Bethesda, MD 20892, or at manolio@nih.gov.

N Engl J Med 2010;363:166-76.

Genomewide association studies — in which hundreds of thousands of single-nucleotide polymorphisms (SNPs) are tested for association with a disease in hundreds or thousands of persons (Fig. 1) — have revolutionized the search for genetic influences on complex traits.[1,2] Such conditions, in contrast with single-gene disorders, are caused by many genetic and environmental factors working together, each having a relatively small effect and few if any being absolutely required for disease to occur. Although complex conditions have been referred to as the geneticist's nightmare,[3] in the past 5 years genomewide association studies have identified SNPs implicating hundreds of robustly replicated loci (i.e., specific genomic locations) for common traits.[4]

These studies raise many questions, such as why the identified variants have low associated risks and account for so little heritability.[5] Explanations for this apparent gap are being sought. Perhaps the answer will reside in rare variants (see the Glossary for this and other key terms), which are not captured by current genomewide association studies; structural variants, which are poorly captured by current studies; other forms of genomic variation; or interactions between genes or between genes and environmental factors.[6] Despite their value in locating the vicinity of genomic variants that may be causing disease, few of the SNPs identified in genomewide association studies have clear functional implications that are relevant to mechanisms of disease.[7] Narrowing an implicated locus to a single variant that directly causes susceptibility to disease by disrupting the expression or function of a protein has proved elusive to date. This will be a key step in improving our understanding of the mechanisms of disease and in designing effective strategies for risk assessment and treatment.

There are also clinical research questions that must be answered before data from genomewide association studies can be routinely incorporated into health care delivery. These questions include how to use the data obtained in these studies to screen for and predict disease and to improve the processes of drug selection and dosing. Another, more immediate question is how to respond in the rare case of a patient who has already purchased a genomewide association scan.

TECHNICAL ASPECTS OF THE GENOMEWIDE ASSOCIATION STUDY

Genomewide association studies build directly on recent efforts to map the patterns of inheritance for the most common form of genomic variation, the SNP.[8,9] An estimated 10 million common SNPs — those with a minor-allele frequency of at least 5% — are transmitted across generations in blocks, allowing a few particular, or tag,

SNPs to capture the great majority of SNP variation within each block.[10] Rapid advances in technology and quality control now permit affordable, reliable genotyping of up to 1 million SNPs in a single scan of a person's DNA.[11]

Scanning can be used in various study designs, including case–control studies, cohort studies, and clinical trials, as long as it is recognized that the known strengths and weaknesses of these designs are pertinent to the use of scanning.[12,13] A complication of genomewide association studies is the enormous number of tests of association required (at least one per SNP); thresholds of statistical significance are stringent, making it necessary to work with very large samples.[14] One frequently used approach to managing size is the tiered design, in which a subset of SNPs found to be significant in the genomewide association study (sometimes called the discovery set) is genotyped in a second tier (a replication set), yielding a smaller subset of significantly associated SNPs that are then tested in a third tier (a second replication set), and so on.[15,16] This process helps to identify false positive associations. Carrying forward a large number of SNPs identified through a genomewide association study into a test of replication also minimizes false negative results[17] while raising the bar for the establishment of true positive results. The pooling of results obtained in genomewide association studies (Fig. 2) under the auspices of large consortia is often required for the detection of variants with small effects on the risk of disease. Such pooled studies, like all genetic association studies, must be examined and controlled for differences in allele frequency between groups that can lead to spurious (false positive) associations.[12] The most reliable evidence of a true genetic association, short of defining the causal variant functionally, is replication of the association, especially if it appears in multiple populations.[18,19]

SURVEY OF FINDINGS

Nearly 600 genomewide association studies covering 150 distinct diseases and traits have been published, with nearly 800 SNP–trait associations reported as significant ($P<5\times10^{-8}$) (Fig. 3) (an interactive version of Fig. 3 is available with the full text of this article at NEJM.org).[4] Such associations are consistent with the common disease–common variant hypothesis, which posits that genetic influences on susceptibility to common diseases are attributable to a limited number of variants present in more than 1% to 5% of the population.[20,21] The common disease–common variant hypothesis is exemplified by susceptibility to age-related macular degeneration. Five major variants are associated with age-related macular degeneration, and each is associated with a risk of disease that is two to three times the risk for a person without one of the variants.[22] Two of these variants, found in the complement factor H (*CFH*) gene, are common in the populations studied (allele frequencies of 36% and 57% among unaffected persons), and the other three variants have allele frequencies of 5 to 19% in the populations studied.[23] Taken together, these five variants more than double the risk of age-related macular degeneration in the siblings of affected persons, accounting for roughly half the estimated total risk for siblings, and suggest that the complement-mediated inflammation pathway is central to pathogenesis.[23,24] The discovery that inflammation plays a role in age-related macular degeneration and is proving to be a suitable target for therapeutic intervention in animal models[25,26] demonstrates the power of the genomewide association study to implicate previously unsuspected pathways in the cause and pathogenesis of disease, leading to the development of new therapies.

The genomewide association study has also yielded more than 30 variants related to Crohn's disease.[27] Three of these variants, found in the genes *NOD2*, *IL23R*, and *LRRK2*, are common (all but one have risk-allele frequencies of more than 9% in the populations studied) and are associated with an increase in risk by a factor of 1.5 to 4. However, the remainder confer very small risk elevations (odds ratios, 1.08 to 1.35) and require extremely large studies for detection. A similar pattern of a few variants having large effects but most having small effects has emerged for type 1 diabetes, with more than 40 variants identified to date.[28,29]

Other common conditions have not been as amenable to investigation of genomewide associations. An early example was schizophrenia. Five genomewide association studies failed to find any variants reaching genomewide significance.[4] A sixth study implicated rare structural variants that disrupt neurodevelopmental pathways,[30] raising questions about the role of structural variants in neuropsychiatric disorders.[31] Subsequent,

An interactive graphic showing genomewide associations is available at NEJM.org

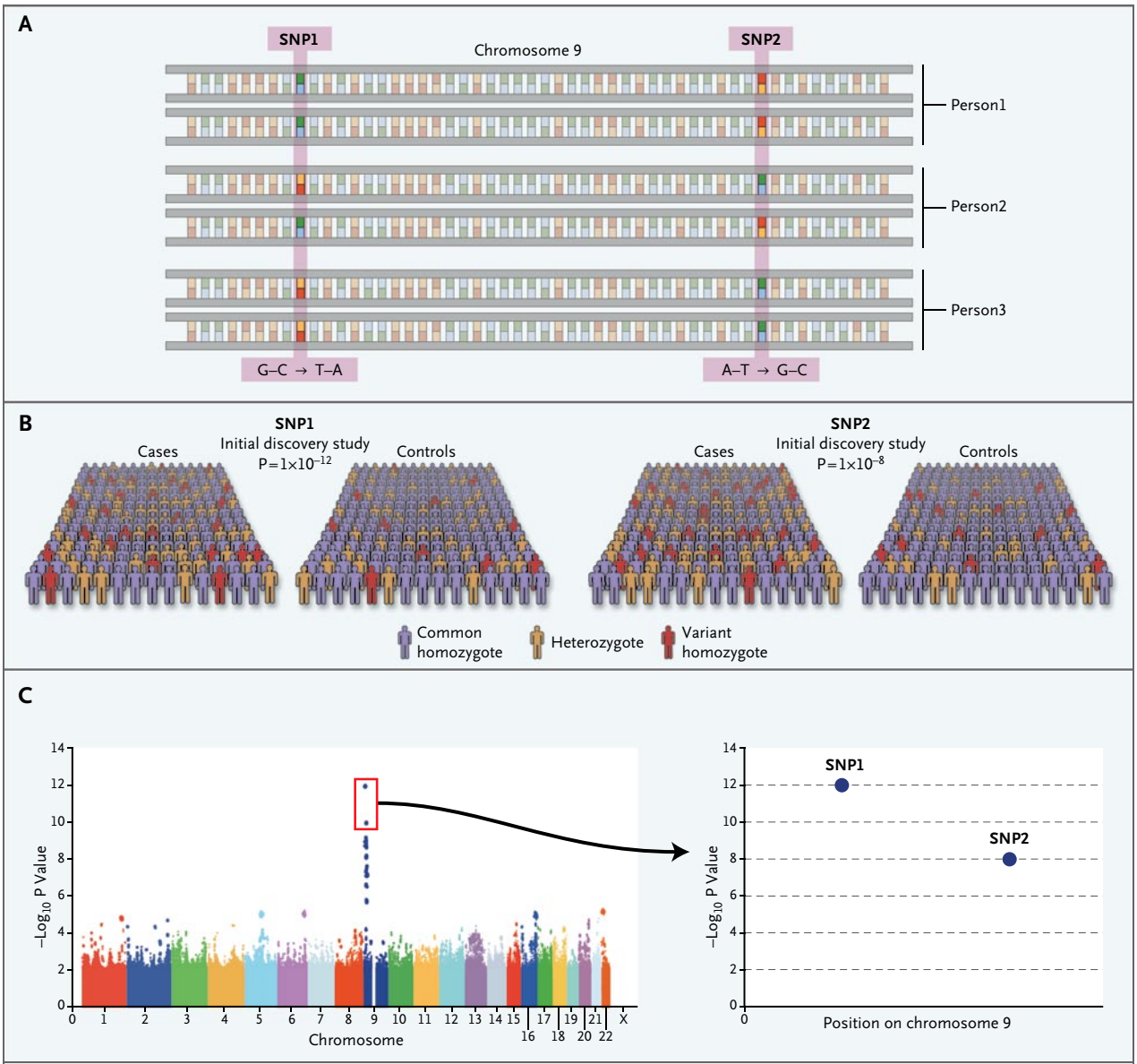

Figure 1. The Genomewide Association Study.

The genomewide association study is typically based on a case–control design in which single-nucleotide polymorphisms (SNPs) across the human genome are genotyped. Panel A depicts a small locus on chromosome 9, and thus a very small fragment of the genome. In Panel B, the strength of association between each SNP and disease is calculated on the basis of the prevalence of each SNP in cases and controls. In this example, SNPs 1 and 2 on chromosome 9 are associated with disease, with P values of 10^{-12} and 10^{-8}, respectively. The plot in Panel C shows the P values for all genotyped SNPs that have survived a quality-control screen, with each chromosome shown in a different color. The results implicate a locus on chromosome 9, marked by SNPs 1 and 2, which are adjacent to each other (graph at right), and other neighboring SNPs.

larger studies investigating the risk of schizophrenia have implicated several variants — both structural variants and SNPs — in the region of the major histocompatibility complex (*MHC*) and at other loci, associations that have been replicated in independent samples.[32-34]

Generally, associations between SNPs and traits tend to be of modest effect size, with a median odds ratio per copy of the risk allele of 1.33.[7] Several variants carry odds ratios above 3.00, including some exceeding 12.00. These are of particular interest, since it seems likely that there would have been evolutionary pressure against their selection unless they provided some survival ben-

Glossary

Annotation catalog: A map denoting the function of specific genomic regions, such as sites to which noncoding RNA or transcription factors bind.
Common disease–common variant hypothesis: The hypothesis that genetic influences on susceptibility to common diseases are attributable to a limited number of variants present in more than 1% to 5% of the population.
Complex condition: A condition caused by the interaction of multiple genes and environmental factors. Examples of complex conditions, which are also called multifactorial diseases, are cancer and heart disease.
Copy-number variation: Variation from one person to the next in the number of copies of a particular gene or DNA sequence. The full extent to which copy-number variation contributes to human disease is not yet known.
Fine mapping: An experimental approach to narrowing a genomewide association signal by typing all known SNPs in the haplotype block containing the tag SNP. If successful, this approach results in the identification of a subsegment of the block that has a stronger association than the surrounding areas.
Gene deserts: Large intergenic regions.
Haplotype: A set of DNA variations, or polymorphisms, that tend to be inherited together. A haplotype can refer to a combination of alleles or to a set of single-nucleotide polymorphisms found on the same chromosome.
Heritability: The proportion of interindividual differences (variance) in a trait that is the result of genetic factors; often estimated on the basis of parent–offspring correlations for continuous traits or the ratio of the incidence in first-degree relatives of affected persons to the incidence in first-degree relatives of unaffected persons.
Intergenic regions: Segments of DNA that do not contain or overlap genes.
Introns: The portions of a gene that are removed (spliced out) before translation to a protein. Introns may contain regulatory information that is critical to appropriate gene expression.
Inversion: A chromosomal segment that has been broken off and reinserted in the same place, but with the genetic sequence in reverse order.
Linkage disequilibrium: An association between two alleles located near each other on a chromosome, such that they are inherited together more frequently than would be expected by chance.
Low-depth coverage: A preliminary strategy in DNA sequencing whereby each base pair is sequenced a minimum of 2 to 4 times rather than the 20 to 30 times that is characteristic of complete (high-depth) sequencing.
Minor-allele frequency: The proportion of the less common of two alleles in a population (with two alleles carried by each person at each autosomal locus), ranging from <1% to <50%.
Noncoding RNAs: Segments of RNA that are not translated into amino acid sequences but may be involved in the regulation of gene expression.
Nonsynonymous single-nucleotide polymorphism: A polymorphism that results in a change in the amino acid sequence of a protein (and may therefore affect the function of the protein).
Rare variant: A genetic variant with a minor-allele frequency of less than 1%. Rare variants are typically single-nucleotide substitutions but can also be structural variants.
RNA interference: The inhibition of gene expression by noncoding RNA molecules.
Single-nucleotide polymorphism (SNP): A single-nucleotide variation in a genetic sequence; a common form of variation in the human genome.
Structural variant: A genetic variant involving the insertion, deletion, duplication, translocation, or inversion of segments of DNA up to millions of bases in length.
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.
1000 Genomes Project: An international collaboration formed to produce an extensive public catalog of human genetic variation, including SNPs and structural variants and the haplotypes on which they occur.
Transcription factor: A protein that binds to gene regulatory regions in DNA and helps to control gene expression.
Translocation: A chromosomal segment that has been broken off and reinserted in a different place in the genome.

efits in earlier periods or different environments. This is not to imply that smaller odds ratios are unimportant. The genes *PPARG* and *KCNJ11*, associated with type 2 diabetes, and *IL12B*, associated with psoriasis, encode proteins that are targets for thiazolidinediones, sulfonylureas, and anti-p40 antibodies, respectively,[2,35] yet all have odds ratios less than 1.45. Such variants may shed light on the pathophysiology of their associated traits and reveal new therapeutic targets.[7]

Only 12% of SNPs associated with traits are located in, or occur in tight linkage disequilibrium with, protein-coding regions of genes, although SNPs in protein-coding regions are heavily overrepresented on genotyping arrays (Fig. 4).[7] Approximately 40% of trait-associated SNPs fall in intergenic regions, and another 40% are located in noncoding introns. These two findings have sharpened the focus on the potential roles of intronic, and particularly intergenic, regions in regulating gene expression.[1]

Other surprising findings include the association of SNPs with genes originally not thought to have a role in a given disease (Table 1). The potential roles of the complement system in age-related macular degeneration, a disease previously thought to be primarily degenerative in origin,[39] or of autophagy in inflammatory bowel disease,[40] for example, were not widely suspected until these systems were implicated through genomewide as-

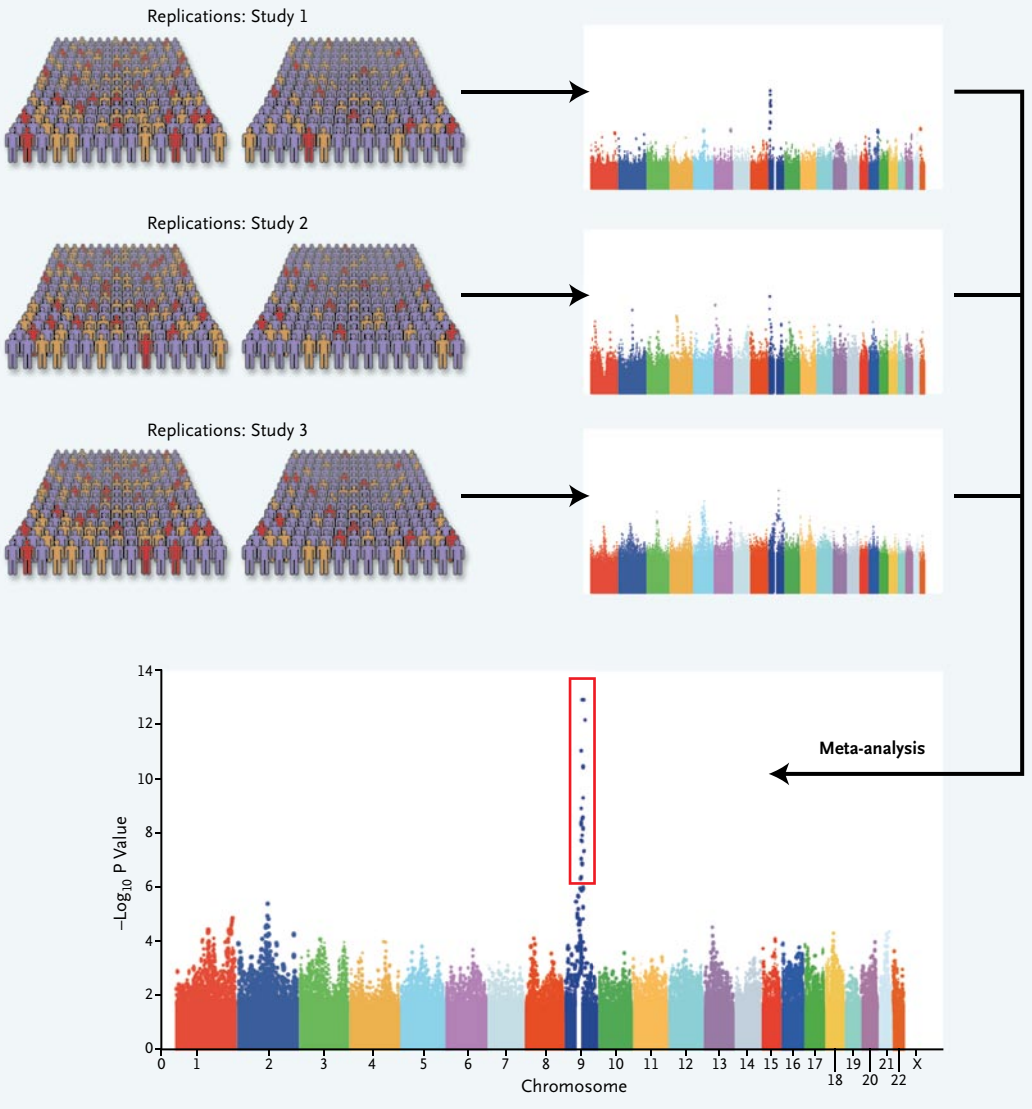

Figure 2. Meta-Analysis of Genomewide Association Studies.

The results of genomewide association studies can be evaluated in a meta-analysis, which combines the results of multiple studies to improve the power for detecting associations. In this example, the results of three studies, none of which may show genomewide significance individually, are combined in a meta-analysis to reveal a strong, significant signal on chromosome 9.

sociation studies. Signals falling in large so-called gene deserts, such as the 8q24.22 locus (which includes markers associated with prostate cancer)[41] and the 5p13.1 region (which includes markers associated with Crohn's disease),[42] raised concern initially that they were false positive, spurious associations. However, the repeated replication of these associations has established that the regions clearly exert influences — though as yet unknown — on the diseases.

Similarly, genomewide association studies have identified loci that are shared by conditions previously thought to be unrelated (Table 2). The possibility of common etiologic pathways in such disparate conditions or traits as type 2 diabetes and invasive melanoma, Crohn's disease and

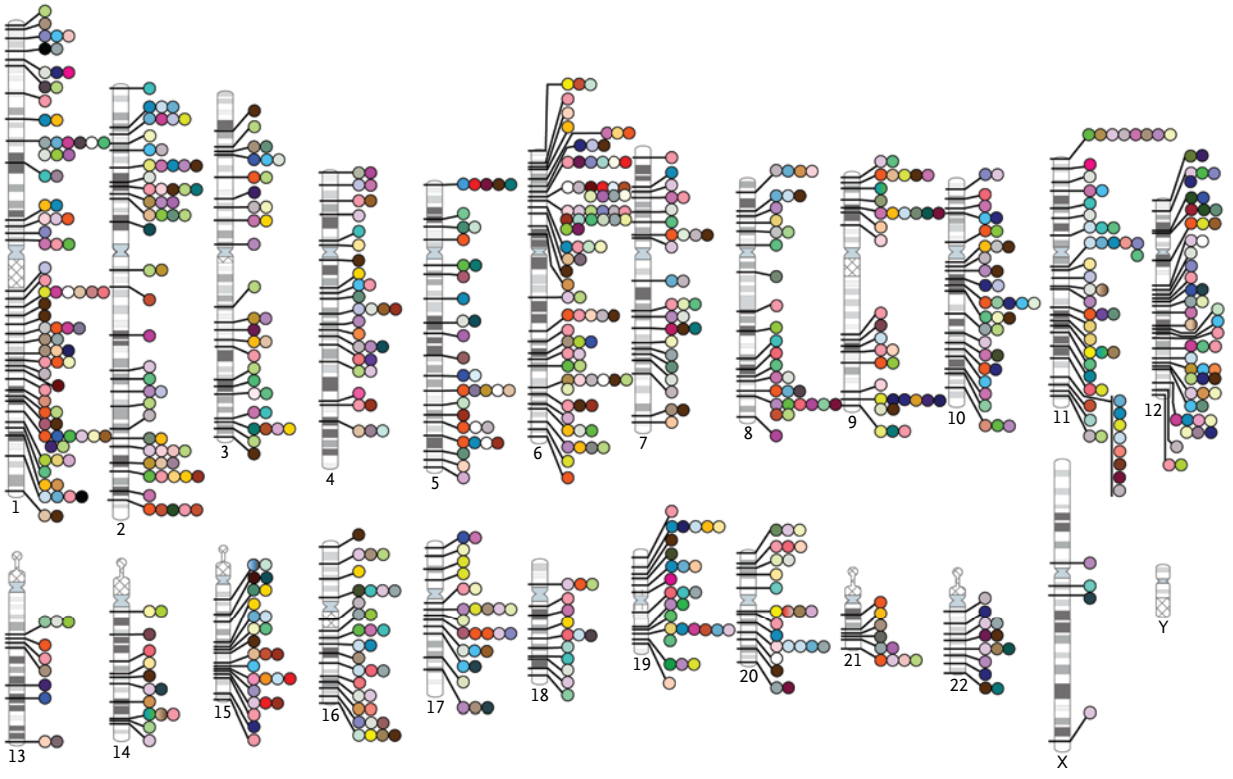

Figure 3. Genomewide Associations Reported through March 2010.

Circles indicate the chromosomal location of nearly 800 single-nucleotide polymorphisms (SNPs) significantly associated ($P<5\times10^{-8}$) with a disease or trait and reported in the literature (545 studies published through March 2010 yielded the associations depicted). Each disease type or trait is coded by color. Adapted from the National Human Genome Research Institute.[4]

Parkinson's disease, or prostate cancer and height raises intriguing questions about the pathophysiology of these seemingly unrelated conditions and about the potential for using drugs that are effective in the treatment of one condition for the treatment of the other.[51]

CHALLENGES

Trait-associated SNPs may point the way toward functional genetic variants but are unlikely themselves to be the causative variants, at least given our current understanding of genomic function and regulation. A first step in narrowing a genomewide association signal to potentially causative variants is to type all the known SNPs in the haplotype block represented by the tag SNP (a process known as fine mapping) to determine whether one of these SNPs has a stronger association (than that tag SNP) or an established functional effect. Although this approach has shown promise in identifying causal variants,[52] its yield has been limited.[53] Extensive sequencing of an associated region may identify additional, previously unknown, rare variants (frequency, <1%) with a possible biologic role. The use of this approach has suggested that variants of *IFIH1* confer susceptibility to type 1 diabetes,[54] a finding that is consistent with this gene's established role in antiviral responses and the known association between type 1 diabetes and viral infections.

Given the lack of good representation of SNPs with a prevalence of less than 5% in current genomewide association arrays, a comprehensive catalogue of SNPs with a prevalence of 1 to 5% is being generated by the 1000 Genomes Project[55] for potential inclusion in fine-mapping efforts and expanded genomewide association arrays. In the project's pilot effort, more than 11 million novel SNPs have been identified in what was ini-

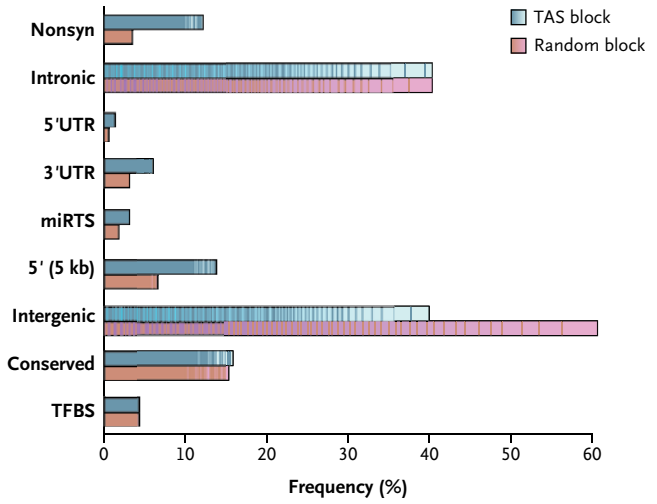

Figure 4. Functional Classifications of 465 Trait-Associated SNPs and the SNPs in Linkage Disequilibrium with Them.

The frequency of a specific functional classification among trait-associated SNPs (TAS) and their linkage disequilibrium partners is shown in blue. The frequency of functional classifications among SNPs randomly drawn from genotyping arrays is shown in pink ($r^2 \geq 0.8$). The abbreviation miRTS denotes microRNA target site. Nonsynonymous SNPs (Nonsyn) are associated with one or more traits nearly three times as often as randomly selected SNPs, and 5′ promoter SNPs nearly twice as often. Although intronic and intergenic SNPs are not overrepresented in associations as compared with randomly selected SNPs, they account for the great majority — more than 80% — of associated SNPs. TFBS denotes transcription-factor–binding site and UTR untranslated region.[7]

tially low-depth coverage of 172 persons.[56] Gene-expression data may also implicate a particular gene as underlying an association signal, as suggested by expression data implicating the gene *PTGER4* in a genomewide association study of Crohn's disease.[42] Annotation catalogues (maps of functions of variants), such as those related to transcription-factor binding (promoting gene expression) or to RNA interference (silencing genes), are currently in development and should facilitate the identification of functional variants underlying genomewide association signals.[57]

The small proportion of heritability and risk of disease typically explained by genomewide association findings presents a challenge: how to identify the variants that confer the outstanding risk — the risk that has not been accounted for.[58] Larger genomewide association studies that identify more variants are likely to identify variants with even smaller effect sizes. The importance of structural variation, including copy-number variants, inversions, and translocations, is an active area of investigation; several structural variants underlie genomewide association signals for autism, schizophrenia, Crohn's disease, and obesity.[31,59] Also needed are studies of population samples with diverse geographic ancestries, particularly recent African ancestry. These older populations, which have undergone more mutations and a greater number of recombination events, have greater degrees of genetic variation and shorter stretches of linkage disequilibrium, allowing for better localization of genomewide association signals.[6,8]

RISK ASSESSMENT

The potential for variants identified in genomewide association studies to predict the risk of complex diseases has been anticipated since the publication of the first reports, but this application is problematic.[22,60] The question of how best to assess the usefulness of genetic variants in disease prediction is the subject of lively debate, and optimal metrics for assessing the clinical effect have yet to be identified. Most would agree, however, that appropriate considerations extend beyond odds ratios or population attributable risks to more complex measures such as the area under the receiver-operating-characteristic curve (AUC) and risk-reclassification statistics.[61,62]

For the prediction of complex diseases, genotypes at multiple SNPs are often combined into scores calculated according to the number of risk alleles carried, which is the approach that Kathiresan and colleagues used in predicting the risk of cardiovascular disease on the basis of nine SNPs associated with cholesterol levels.[63] This score was strongly associated with the risk of cardiovascular disease even after adjustment for standard risk factors, including family history, but the AUC was unchanged after inclusion of the genotype score.[63] Among the subjects initially considered to be at intermediate risk for cardiovascular disease (9% of the total cohort), 26% were reclassified in the low-risk or high-risk category, and reclassification statistics showed significant improvement in risk classification. The reclassifications had implications for clinical care as recommended in standard clinical guidelines. On closer analysis, however, the reclassifications were based on only minor increments in the risk score, which shifted subjects with borderline scores from one category to the next[60] (Fig. 5). Indeed, collective odds ratios of 200 or more may be necessary if

Table 1. Examples of Previously Unsuspected Associations between Certain Conditions and Genes and the Related Metabolic Function or Pathway, According to Genomewide Association Studies.

Condition	Gene	Function or Pathway	Source of Data
Age-related macular degeneration	*CFH*	Complement-mediated inflammation	Klein et al.[25]
Coronary disease	*CDKN2A, CDKN2B*	Cell-cycle regulator	Helgadottir et al.[36]
Childhood asthma	*ORMDL3*	Unknown	Moffatt et al.[37]
Type 2 diabetes	*CDKAL1*	Cell-cycle regulator	Scott et al.[3]
Crohn's disease	*ATG16L1*	Autophagy	Rioux et al.[38]

there is to be meaningful reclassification of subjects on the basis of risk.[64] Similar attempts to use multiple SNPs to predict the risk of prostate cancer have also been of limited value, with minimal improvements in the AUC, as compared with the use of standard clinical risk factors, and identification of only a small proportion of subjects (<2%) at the highest levels of risk.[61,65] Evidence that genotype scores may be of particular value in predicting risk among persons with a family history of a particular condition is intriguing and should be explored in studies of conditions other than heart disease and prostate cancer.[61,66]

What is becoming clear from these early attempts at genetically based risk assessment is that currently known variants explain too little about the risk of disease occurrence to be of clinically useful predictive value. One can anticipate that as sample sizes increase and more risk variants are identified, the predictive value of cumulative genotypic scores will increase.[22,67,68] It has also been argued that the use of dense genotyping information, from tens of thousands of SNPs with only nominal associations with disease, may improve the accuracy of phenotypic prediction.[34] Care is needed in evaluating genetic predictive models, since they are often specific to the population in which they were developed, and their value can vary with genotypic frequencies, effect sizes, and disease incidence.[68] Possible clinical uses of predictive scores — for example, in deciding which patients should be screened more intensively for breast cancer with the use of mammography[69] or for statin-induced myopathy with the use of muscle enzyme assays[70] — will require rigorous, preferably prospective, evaluation before being accepted into clinical practice.

Genomewide scans permit screening for many conditions at once. If binomial probabilities were applied to 40 independent diseases, for example, roughly 90% of the population would be placed in the top 5% of those at genetic risk for at least one of the diseases, 33% would be in the top 1%, and 4% would be in the top 0.1%.[71] Expanding such screening to 120 diseases would nearly triple the proportion in the top 0.001% at risk and identify 1.2% at the top 0.01%, levels that could justify population-based screening if appropriate interventions were available. The ability to assess risk for 120 conditions at the same time also raises the concern that predictive models will yield conflicting recommendations; if implemented, they could reduce a person's risk for development of one condition and exacerbate the risk for development of another.

Such considerations are timely and important, since several commercial ventures are marketing genomewide association–based screening directly to consumers.[72] This testing can often be obtained without a physician's intercession and has been promoted for medical, genealogic, and even recreational purposes. The information provided to the customer is often founded on scant evidence and based on average risks that are difficult to apply to an individual person.[73] Few factors associated with differences in risk across a population will separate affected and unaffected groups widely enough to be useful for individual prediction.[64] Adequate communication of disease risk is a topic that has challenged generations of physicians and patients, and the perception of risk is more often influenced by emotion than by science. Genome-based risk information may not improve communication of risk, but its uniquely individual nature may be personally motivating and could be explored with respect to the promotion of salutary behaviors.

Patients inquiring about genomewide association testing should be advised that at present the results of such testing have no value in predicting risk and are not clinically directive. Clinicians would do well to use the discussion as an oppor-

Table 2. Examples of Loci Shared by Conditions or Traits Previously Thought to Be Unrelated, According to Genomewide Association Studies.

Gene	Conditions Sharing Associations	Source of Data
CDKN2A, CDKN2B	Coronary disease	Helgadottir et al.[36]
	Type 2 diabetes	Scott et al.[3]
	Invasive melanoma	Kamb et al.[43]
ORMDL3	Childhood asthma	Moffatt et al.[37]
	Crohn's disease	Barrett et al.[27]
CDKAL1	Type 2 diabetes	Scott et al.[3]
	Prostate cancer	Steinthorsdottir et al.[44]
LRRK2	Parkinson's disease	Paisán-Ruíz et al.[45]
	Crohn's disease	Barrett et al.[27]
KITLG	Testicular carcinoma	Rapley et al.[46]
	Blond or brown hair	Sulem et al.[47]
C10orf67	Sarcoidosis	Franke et al.[48]
	Celiac disease	Franke et al.[48]
JAZF1	Height	Johansson et al.[49]
	Type 2 diabetes	Zeggini et al.[50]
	Prostate cancer	Thomas et al.[17]

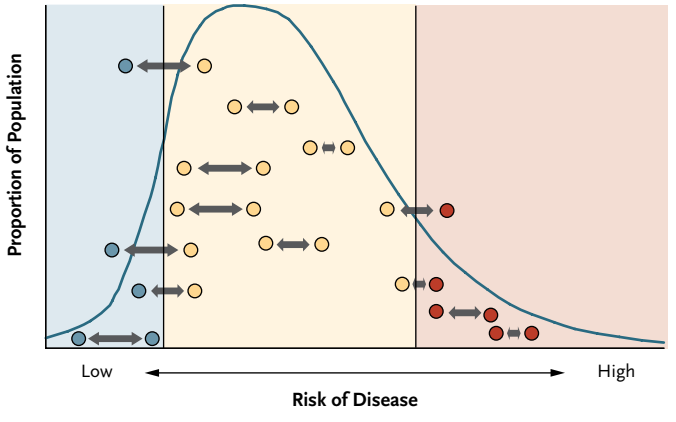

Figure 5. Reclassification of Persons at Various Levels of Risk, According to Risk Thresholds.

The majority of a population, depicted as the area under the curve, is at moderate, or average, risk of disease (yellow shading), with small proportions at low risk (blue shading) and high risk (pink shading), sometimes with a skewed distribution as a result of persons at very high risk (blue line). Additional information may produce small, incremental shifts in risk estimates (arrows), which may suffice to move persons at the margin of one risk category into another risk category.

tunity to point out other identifiable, modifiable risk factors that motivated patients can control.[12,73] Whether to heed such advice or instead undergo testing and present the physician with the test results as a fait accompli is the choice of the individual patient. A decision to undergo genomewide association testing may result in the diversion of scarce time and resources to counseling or follow-up investigation of findings.[74]

CONCLUSIONS

Genomewide association studies have proved successful in identifying genetic associations with complex traits. This reasonably unbiased approach to surveying the genome has opened doors to potential treatments by revealing the unexpected involvement of certain functional and mechanistic pathways in a variety of disease processes.[2] Although the approach has proved powerful in identifying robust associations between many SNPs and traits, much additional work is needed to determine the functional basis for the observed associations so that appropriate interventions can be developed. Much more remains to be learned about how variations in intronic and intergenic regions (where the vast majority of SNP–trait associations reside) influence gene expression, protein coding, and disease phenotypes.[1]

Despite the limitations of using data obtained from genomewide association studies to assess the individual patient's level of risk for a particular condition, genomewide scans may be useful in initiating counseling about nongenetic risk factors or perhaps in screening for a very high risk of many conditions at once. Continued efforts to identify genetic variants that influence the response to drugs may yield new associations that could be used to tailor drug selection and dosing to the profile of the individual patient, particularly if it becomes possible to query these data through a user-friendly interface when a medication is ordered. The substantial challenges of incorporating such research into clinical care must be pursued if the potential of genomic medicine is to be realized.

No potential conflict of interest relevant to this article was reported.

Disclosure forms provided by the author are available with the full text of this article at NEJM.org.

REFERENCES

1. Hardy J, Singleton A. Genomewide association studies and human disease. N Engl J Med 2009;360:1759-68.
2. Manolio TA, Brooks LD, Collins FSA. A HapMap harvest of insights into the genetics of common disease. J Clin Invest 2008;118:1590-605.
3. 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;316:1341-5.
4. Hindorff LA, Junkins HA, Manolio TA. NHGRI Catalog of published genome-wide association studies. (Accessed June 7, 2010, at http://www.genome.gov/gwastudies.)
5. Goldstein DB. Common genetic variation and human traits. N Engl J Med 2009;360:1696-8.
6. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. Nature 2009;461:747-53.
7. Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A 2009;106:9362-7.
8. International HapMap Consortium. A haplotype map of the human genome. Nature 2005;437:1299-320.
9. International HapMap Consortium, Frazer KA, Ballinger DG, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007;449:851-61.
10. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science 2002;296:2225-9.
11. Spencer CC, Su Z, Donnelly P, Marchini J. Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. PLoS Genet 2009;5(5):e1000477.
12. Pearson TA, Manolio TA. How to interpret a genome-wide association study. JAMA 2008;299:1335-44. [Erratum, JAMA 2008;299:2150.]
13. Clayton DG, Walker NM, Smyth DJ, et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. Nat Genet 2005;37:1243-6.
14. Hunter DJ, Kraft P. Drinking from the fire hose — statistical issues in genome-wide association studies. N Engl J Med 2007;357:436-9.
15. Hoover RN. The evolution of epidemiologic research: from cottage industry to "big" science. Epidemiology 2007;18:13-7.
16. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447:1087-93.
17. Thomas G, Jacobs KB, Yeager M, et al. Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet 2008;40:310-5.
18. Todd JA. Statistical false positive or true disease pathway? Nat Genet 2006;38:731-3.
19. Chanock SJ, Manolio T, Boehnke M, et al. Replicating genotype-phenotype associations. Nature 2007;447:655-60.
20. Reich DE, Lander ES. On the allelic spectrum of human disease. Trends Genet 2001;17:502-10.
21. Collins FS, Guyer MS, Chakravarti A. Variations on a theme: cataloging human DNA sequence variation. Science 1997;278:1580-1.
22. Jakobsdottir J, Gorin MB, Conley YP, Ferrell RE, Weeks DE. Interpretation of genetic association studies: markers with replicated highly significant odds ratios may be poor classifiers. PLoS Genet 2009;5(2):e1000337.
23. Maller J, George S, Purcell S, et al. Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. Nat Genet 2006;38:1055-9.
24. Bora NS, Jha P, Bora PS. The role of complement in ocular pathology. Semin Immunopathol 2008;30:85-95.
25. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. Science 2005;308:385-9.
26. Rohrer B, Long Q, Coughlin B, et al. A targeted inhibitor of the alternative complement pathway reduces angiogenesis in a mouse model of age-related macular degeneration. Invest Ophthalmol Vis Sci 2009;50:3056-64.
27. Barrett JC, Hansoul S, Nicolae DL, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet 2008;40:955-62.
28. Barrett JC, Clayton DG, Concannon P, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet 2009 May 10 (Epub ahead of print).
29. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661-78.
30. Walsh T, McClellan JM, McCarthy SE, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 2008;320:539-43.
31. McCarroll SA. Extending genome-wide association studies to copy-number variation. Hum Mol Genet 2008;17:R135-R142.
32. Shi J, Levinson DF, Duan J, et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature 2009;460:753-7.
33. Stefansson H, Ophoff RA, Steinberg S, et al. Common variants conferring risk of schizophrenia. Nature 2009;460:744-7.
34. International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 2009;460:748-52.
35. Krueger GG, Langley RG, Leonardi C, et al. A human interleukin-12/23 monoclonal antibody for the treatment of psoriasis. N Engl J Med 2007;356:580-92.
36. Helgadottir A, Thorleifsson G, Manolescu A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. Science 2007;316:1491-3.
37. Moffatt MF, Kabesch M, Liang L, et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. Nature 2007;448:470-3.
38. 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;39:596-604.
39. Ambati J, Ambati BK, Yoo SH, Ianchulev S, Adamis AP. Age-related macular degeneration: etiology, pathogenesis, and therapeutic strategies. Surv Ophthalmol 2003;48:257-93.
40. Budarf ML, Labbé C, David G, Rioux JD. GWA studies: rewriting the story of IBD. Trends Genet 2009;25:137-46.
41. Amundadottir LT, Sulem P, Gudmundsson J, et al. A common variant associated with prostate cancer in European and African populations. Nat Genet 2006;38:652-8.
42. 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.
43. Kamb A, Shattuck-Eidens D, Eeles R, et al. Analysis of the p16 gene (CDKN2) as a candidate for the chromosome 9p melanoma susceptibility locus. Nat Genet 1994;8:23-6.
44. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, et al. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. Nat Genet 2007;39:770-5.
45. Paisán-Ruíz C, Jain S, Evans EW, et al. Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. Neuron 2004;44:595-600.
46. Rapley EA, Turnbull C, Al Olama AA, et al. A genome-wide association study of testicular germ cell tumor. Nat Genet 2009;41:807-10.
47. Sulem P, Gudbjartsson DF, Stacey SN, et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat Genet 2007;39:1443-52.
48. Franke A, Fischer A, Nothnagel M, et al. Genome-wide association analysis in sarcoidosis and Crohn's disease unravels a common susceptibility locus on 10p12.2. Gastroenterology 2008;135:1207-15.
49. Johansson A, Marroni F, Hayward C, et al. Common variants in the JAZF1 gene

associated with height identified by linkage and genome-wide association analysis. Hum Mol Genet 2009;18:373-80.
50. Zeggini E, Scott LJ, Saxena R, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet 2008;40:638-45.
51. Kingsmore SF, Lindquist IE, Mudge J, Gessler DD, Beavis WD. Genome-wide association studies: progress and potential for drug discovery and development. Nat Rev Drug Discov 2008;7:221-30.
52. Jallow M, Teo YY, Small KS, et al. Genome-wide and fine-resolution association analysis of malaria in West Africa. Nat Genet 2009 May 24 (Epub ahead of print).
53. Ioannidis JP, Thomas G, Daly MJ. Validating, augmenting and refining genome-wide association signals. Nat Rev Genet 2009;10:318-29.
54. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. Science 2009;324:387-9.
55. 1000 Genomes: a deep catalog of human genetic variation. (Accessed June 14, 2010, at http://www.1000genomes.org/page.php.)
56. Abecasis GR. The 1000 Genomes Project: analysis of pilot datasets: biology of genomes. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 2009:246.
57. ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 2007;447:799-816.
58. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era — concepts and misconceptions. Nat Rev Genet 2008;9:255-66.
59. Willer CJ, Speliotes EK, Loos RJ, et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. Nat Genet 2009;41:25-34.
60. Janssens AC, van Duijn CM. Genome-based prediction of common diseases: advances and prospects. Hum Mol Genet 2008;17:R166-R173.
61. Kraft P, Wacholder S, Cornelis MC, et al. Beyond odds ratios — communicating disease risk based on genetic profiles. Nat Rev Genet 2009;10:264-9.
62. Cook NR, Ridker PM. Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures. Ann Intern Med 2009;150:795-802.
63. Kathiresan S, Melander O, Anevski D, et al. Polymorphisms associated with cholesterol and risk of cardiovascular events. N Engl J Med 2008;358:1240-9.
64. Ware JH. The limitations of risk factors as prognostic tools. N Engl J Med 2006;355:2615-7.
65. Zheng SL, Sun J, Wiklund F, et al. Cumulative association of five genetic variants with prostate cancer. N Engl J Med 2008;358:910-9.
66. Xu J, Sun J, Kader AK, et al. Estimation of absolute risk for prostate cancer using genetic markers and family history. Prostate 2009;69:1565-72.
67. van der Net JB, Janssens AC, Sijbrands EJ, Steyerberg EW. Value of genetic profiling for the prediction of coronary heart disease. Am Heart J 2009;158:105-10.
68. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk of complex disease. Curr Opin Genet Dev 2008;18:257-63.
69. Pharoah PDP, Antoniou AC, Easton DF, Ponder BAJ. Polygenes, risk prediction, and targeted prevention of breast cancer. N Engl J Med 2008;358:2796-803.
70. The SEARCH Collaborative Group. *SLCO1B1* variants and statin-induced myopathy — a genomewide study. N Engl J Med 2008;359:789-99.
71. Select Committee on Science and Technology, House of Lords. Second report — genomic medicine. HL paper 107-I. London: Stationery Office, 2009:104.
72. Kaye J. The regulation of direct-to-consumer genetic tests. Hum Mol Genet 2008;17:R180-R183.
73. Hunter DJ, Khoury MJ, Drazen JM. Letting the genome out of the bottle — will we get our wish? N Engl J Med 2008;358:105-7.
74. McGuire AL, Burke W. An unwelcome side effect of direct-to-consumer personal genome testing: raiding the medical commons. JAMA 2008;300:2669-71.