Common Genetic Variation and Human Traits

David B. Goldstein, Ph.D.

The human genome has been cracked wide open in recent years and is spilling many of its secrets. More than 100 genome-wide association studies have been conducted for scores of human diseases, identifying hundreds of polymorphisms that are widely seen to influence disease risk. After many years in which the study of complex human traits was mired in false claims and methodologic inconsistencies, genomics has brought not only comprehensive representation of common variation but also welcome rigor in the interpretation of statistical evidence. Researchers now know how to properly account for most of the multiple hypothesis testing involved in mining the genome for associations, and most reported associations reflect real biologic causation. But do they matter?

Unfortunately, most common gene variants that are implicated by such studies are responsible for only a small fraction of the genetic variation that we know exists. This observation is particularly troubling because the studies are largely comprehensive in terms of common single-nucleotide polymorphisms (SNPs), the genomic markers that are genotyped and with which disease associations are tested. We're finding the biggest effects that exist for this class of genetic variant, and common variation is packing much less of a phenotypic punch than expected. Some experts emphasize that small effect sizes don't necessarily mean that a gene variant is of no interest or use. Effect size is a function of what a variant does: it may change only slightly a gene's expression or a protein's function. The gene's pathway, however, may be decisive for a particular condition, or pharmacologic action on the same protein may produce much larger effects in controlling disease. These arguments are reasonable, as far as they go, and there are supporting examples, such as a polymorphism of modest effect in *PPARG*, a gene that encodes a drug target for diabetes.

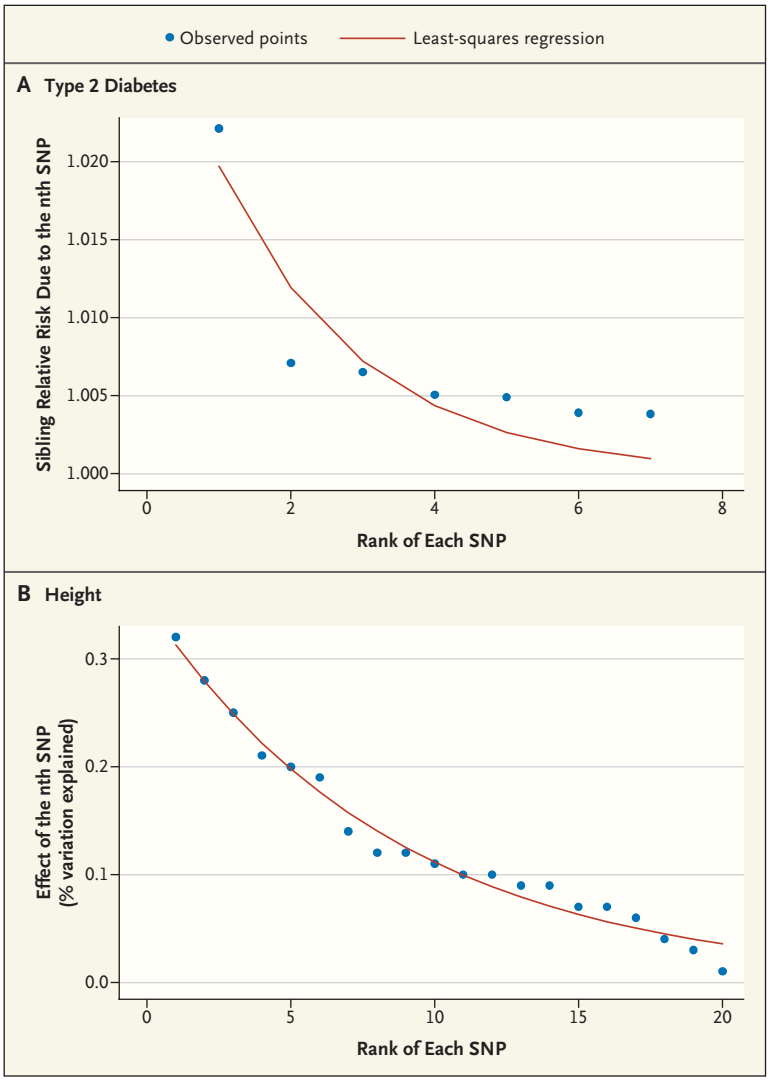
But the arguments hold only if common genetic variation implicates a manageable number of genes. If effect sizes were so small as to require a large chunk of the genome to explain the genetic component of a disorder, then no guidance would be provided: in pointing at everything, genetics would point at nothing. To assess whether effect sizes are too small in this sense, consider two examples of complex human traits — type 2 diabetes and height. In their recent review, Manolio et al.[1] described seven gene variants that influence the risk of type 2 diabetes. In addition to these variants, the one with the strongest effect on familial aggregation is in the *TCF7L2* gene.

One way to assess a variant's effect is by comparing the disease risk of the sibling of an affected person with that in the general population (sibling relative risk). The *TCF7L2* variant is associated with a sibling relative risk for type 2 diabetes of only about 1.02, whereas the overall risk of disease among siblings of affected persons is three times that in the general population. If the human genome carried scores of variants with such effects, they would collectively generate a substantial sibling relative risk. Unfortunately, we now know this is not the case: the contribution of common risk alleles to familial clustering falls off dramatically after *TCF7L2* and appears to become asymptotic at a level only marginally above 1 (see Panel A of the figure).[2] It seems likely, then, that an unreasonably large number of such variants would be required to account for the genetic component of diabetes risk, even if the sibling relative risk values overestimate the genetic component of disease.

A more quantitative evaluation is available for height, for which Weedon et al.[3] identified 20 polymorphisms. Using a replication sample set, they estimated that collectively, the variants they studied explain less than 3% of the population variation in height (see Panel B of the figure). To estimate the full distribution of effect sizes (including those of variants not yet discovered), one could assume an exponential distribution and estimate the parameters from the observed data. The predicted effect of the nth SNP is calculated as follows:

$$\text{Effect size of nth SNP} = k + a \times \text{Exp}[-bn],$$

in which $k = 0.0008$, $a = 0.35$, and $b = 0.1152$. To estimate the number of SNPs required to explain 80% of population variation in height (the most common estimate of height's heritability), this equation can be integrated and solved numerically. The answer is that approximately 93,000 SNPs are required to explain 80% of the population variation in height. In



Sibling Relative Risk for Each of 7 SNPs Associated with Type 2 Diabetes (Panel A) and Percentage of Variation Explained by Each of 20 SNPs Associated with Height (Panel B).

Panel A shows the contribution to a sibling relative risk of type 2 diabetes for each of seven SNPs, as estimated from data reported by Manolio et al.[1] with the use of formulas from Risch and Merikangas[2] and plotted against the rank order of the SNPs in terms of the magnitude of their contributions. Panel B shows the percentage of variation explained by each of 20 SNPs associated with height, as reported by Weedon et al.[3] For a quantitative trait, the natural measure of effect size is the proportion of variation in the trait that the SNP explains, which depends on both the allele frequency and the intergenotype differences. Effect sizes are shown as points as well as a fitted exponential function with the use of least-squares regression.

the fitted distribution, the constant term (0.0008) can be viewed as the predicted smallest effect size in the genome, given the 20 strongest effects already identified. The resulting integral can be considered valid only over the range of 1 to approximately 93,000, at which point all heritability would be explained.

I assume that all SNPs yet to be discovered have weaker effect sizes than the weakest so far found. Though the strongest SNP may have been found, many SNPs could remain unidentified in the range of the lower effects that have been determined. If such SNPs are accounted for, fewer SNPs will be required to explain a given proportion of variance. The sample sizes that have been studied for height, however, range from 14,000 to 34,000. At the lower sample size, the power of detection is 90% for the largest effect size; for effect sizes as small as 0.05%, the largest sample size provides a 10% chance of detection. Even if we conservatively assume that all remaining unidentified variants influencing height each explained as much as 0.05% of the variation, 1500 such variants would be required to explain the missing heritability. These calculations also assume that the effects of "height SNPs" are additive. If variants show meaningful interactions, a somewhat stronger genetic effect could emerge among variants with small individual effect sizes. But only dramatic departures from these assumptions would allow a manageable number of common SNPs to account for a sizable fraction of the heritability of height.

If common variants are responsible for most genetic components of type 2 diabetes, height, and similar traits, then genetics will provide relatively little guidance about the biology of these conditions, because most genes are "height genes" or "type 2 diabetes genes." It seems much more likely, however, that most genetic control is due to rarer variants, either single-site or structural, that are not represented in the current studies and that have considerably larger effects than common variants. Whether these "rarer" variants are only slightly below the threshold for detection on current platforms or substantially more rare remains to be

seen. If, however, rarer variants are primarily responsible for the missing heritability, we may yet identify a manageable number of genes and pathways.

Either way, it's hard to have any enthusiasm for conducting genome scans with the use of ever larger cohorts after a study of the first several thousand subjects has identified the strongest determinants among common variants. These initial studies for a given common disease are worth doing, since common variants do appear to explain a sizable fraction of the heritability of certain conditions — notably, exfoliation glaucoma, macular degeneration, and Alzheimer's disease. Beyond studies of this size, however, we enter the flat or declining part of the effect-size distributions, where there are probably either no more common variants to discover or no more that are worth discovering.

By contrast, genome scans have not yet been performed in search of variants involved in many responses to drugs or infectious agents, even though there are examples in both categories of common polymorphisms whose effects dwarf those seen for type 2 diabetes and many other diseases. For example, when exposed to the anti-HIV drug abacavir, a hypersensitivity reaction develops in more than half the carriers of the HLA-B*5701 allele, whereas such a reaction occurs in less than 5% of patients without this allele.[4] Similarly, just three common variants are sufficient to explain 14% of the population variation in HIV-1 viral load.[5]

But with traits such as height or type 2 diabetes, it seems that an inordinate number of common SNPs would be needed to account for a sizable fraction of heritability. Indeed, it's possible that the way genome scans are being interpreted actually overestimates the contributions of common variants. Most variants that have been identified to date are markers, not causal variants, and are generally assumed to reflect the effects of some other, as-yet-unidentified common variant. Another possibility, however, is that some of the associations that are credited to common variants are actually synthetic associations involving multiple rare variants that occur, by chance, more frequently in association with one allele at a common SNP than with the other. In this case, as well, genome scans will overestimate the contribution of common variants.

The apparently modest effect of common variation on most human diseases and related traits probably reflects the efficiency of natural selection in prohibiting increases in disease-associated variants in the population. I believe attention should shift from genome scans of ever larger samples to studies of rarer variants of larger effect. Effectively searching the full human genome for rare variants will require not only sequencing capacity but also thoughtful selection of the most appropriate groups of individual genomes to resequence and thoughtful evaluation and prioritization of the many rare variants identified. There's no guarantee that associations with rare variants will point directly to causation. Nevertheless, the limited role of common variation in many highly heritable diseases argues strongly that there are many rare variants to be found, and it seems reasonable to hope that some of them will suggest novel therapeutic targets or help in the design of personalized prevention or treatment regimens.

These conclusions imply no criticism of the strikingly successful efforts to represent common variation and relate it to common diseases. Indeed, I share the view Hirschhorn presents in his Perspective article (pages 1699–1701) that the early skeptics have been proved wrong about genomewide association studies in most details: patterns of linkage disequilibrium are sufficiently consistent to allow efficient representation of common variation with the use of "tagging" SNPs, and secure associations between polymorphisms and diseases were rapidly and easily identified. But even though genomewide association studies have worked better and faster than expected, they have not explained as much of the genetic component of many diseases and conditions as was anticipated. We must therefore turn more sharply toward the study of rare variants.

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

This article (10.1056/NEJMp0806284) was published at NEJM.org on April 15, 2009.

Dr. Goldstein is director of the Center for Human Genome Variation, Institute for Genome Sciences and Policy, Duke University, Durham, NC.

1. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. J Clin Invest 2008; 118:1590-605.
2. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996;273:1516-7.
3. Weedon MN, Lango H, Lindgren CM, et al. Genome-wide association analysis identifies 20 loci that influence adult height. Nat Genet 2008;40:575-83.
4. Mallal S, Phillips E, Carosi G, et al. HLA-B*5701 screening for hypersensitivity to abacavir. N Engl J Med 2008;358:568-79.
5. Fellay J, Shianna KV, Ge D, et al. A whole-genome association study of major determinants for host control of HIV-1. Science 2007; 317:944-7.

Genomewide Association Studies — Illuminating Biologic Pathways

Joel N. Hirschhorn, M.D., Ph.D.

Human geneticists seek to understand the inherited basis of human biology and disease, aiming either to gain insights that could eventually improve treatment or to produce useful diagnostic or predictive tests. As recently as 2004, few genetic variants were known to reproducibly influence common polygenic diseases (including cancer, coronary artery disease, and diabetes) or quantitative phenotypes (including lipid levels and blood pressure). This relative ignorance limited potential insights into the pathophysiology of common diseases.

The completion of the human genome sequence in 2005 and the provision of an initial catalogue of human genetic variation and a haplotype map (known as the HapMap), together with rapid improvements in genotyping technology and analysis, have permitted genomewide association studies to be undertaken in a large number of samples.[1] In the first and current implementation of this approach, the great majority of genetic variants with population frequencies of 5% or more could be tested directly or indirectly for association with disease risk or quantitative traits — thus providing a potential path to gene discovery for polygenic diseases and traits.

Before the initiation of genomewide association studies, there was considerable and healthy skepticism about their likely success. For example, in 2005, two friends and well-known geneticists, Francis Collins and Thomas Gelehrter, made a public bet: Gelehrter predicted that no more than three new common variants would be reproducibly associated with common diseases by the time the American Society of Human Genetics (ASHG) held its meeting in the autumn of 2008.

During the past 2 years, however, genomewide association studies have identified more than 250 genetic loci in which common genetic variants occur that are reproducibly associated with polygenic traits.[1-4] This explosion represents one of the most prolific periods of discovery in human genetics, with most new loci identified in genomewide association studies published during the past 18 months. The bet was settled: Collins was the clear winner, by a margin of more than 200 new associated variants.

New skeptics have now questioned the value of these recent discoveries. They cite the modest effect sizes of common variants, both individually and in combination, and argue that the small fraction of heritability that is explained by these variants precludes practical prediction or meaningful biologic insights. A second argument is articulated by Goldstein in his Perspective article in this issue of the *Journal* (pages 1696–1698); he predicts that genomewide association studies will not yield too few loci but rather too many. Extrapolating from recent discoveries, he builds a speculative mathematical model and infers that there will be tens of thousands of common variants influencing each disease and trait. Assuming that these variants will be evenly distributed across the genome, he concludes that every gene in the genome could theoretically be implicated, a scenario that would prohibit useful biologic insights.

I believe that the skeptics' arguments either misconstrue the primary goal of genomewide association studies or are contradicted by their findings. The main goal of these studies is not prediction of individual risk but rather discovery of biologic pathways underlying polygenic diseases and traits. It is already clear that the genes being identified expose relevant biology. Genomewide association studies have "rediscovered" many genes that have been shown by decades of work to be important. Of the 23 loci found to be associated with lipid levels, 11 implicate genes encoding apolipoproteins, lipases, and other key proteins in lipid metabolism.[2] Studies of other diseases and traits have highlighted equally relevant genes.[1,3,4] Nearly one fifth of the approximately 90 loci that were found to be associated with type 2 diabetes, lipid levels, obesity, or height include a gene that is mutated in a corresponding single-gene disorder.[2,4] The number of such overlaps is overwhelmingly greater than what would be expected by chance. Furthermore, genomewide association studies have highlighted genes encoding the sites of action of drugs approved by the Food and Drug Administration, including thiazolidinediones and sulfonylureas (in studies of type 2 diabetes),[2] statins (lipid levels),[2]

and estrogens (bone density).[5] Each of the associated variants at a drug-target locus explains less than 1% of phenotypic variation in the population, demonstrating that small effect sizes do not preclude biologic importance.

Critically, genomewide association studies have also highlighted pathways whose relevance to a particular disease or trait was previously unsuspected. The genetic variants that are associated with age-related macular degeneration strongly implicate components of the complement system, the loci associated with Crohn's disease[3] point unambiguously to autophagy and interleukin-23–related pathways, and the height loci[4] include genes encoding chromatin proteins and hedgehog signaling. This clustering into biologic pathways is highly nonrandom (as has been demonstrated by Raychaudhuri and Daly). Already, efforts are under way to translate the new recognition of the role of autophagy in Crohn's disease into new therapeutic leads. As more pathways are highlighted and additional hypotheses emerge, new projects can be born.

Finally, many newly identified loci do not implicate genes with known functions. It is hardly surprising that we do not yet understand the biologic import of every recently associated locus: the associations sometimes do not point unambiguously to a particular gene, and even genes that are clearly implicated are often unannotated with respect to function. For these genes, greater effort will be required before we can generate hypotheses for future work, but by charting new paths, such efforts could eventually lead to the most novel and important insights.

With regard to prediction, the common variants described by genomewide association studies almost universally have modest predictive power, and for most diseases and traits, these variants in combination explain only a small fraction of heritability. However, the success of genomewide association studies is not tied to prediction. If we identify only new pathways underlying disease, these studies will have a tremendous impact.

Nevertheless, it remains likely that for some diseases, the loci that are highlighted in the studies will provide useful predictive information. For several diseases, associated variants already explain 10 to 20% or more of heritability, a magnitude that is similar to the proportion of risk explained by nongenetic tests in widespread clinical use (such as levels of low-density lipoprotein cholesterol or prostate-specific antigen). Furthermore, current estimates are a lower bound for the eventual predictive power of recently discovered loci, which have not been thoroughly examined for additional common and rare variation. Indeed, early experience suggests that multiple independent causal variants may be found at each locus, accounting for additional increments of heritability.[1] Genomewide association studies that are performed in larger samples and that use genotyping platforms designed to test variants with a prevalence of less than 5% will increase the variation explained at these and other as-yet-undiscovered loci, as will studies taking into account interactions among genes and between genes and the environment. Ultimately, the usefulness of genetic information for prediction will depend not on the absolute fraction of heritability explained but rather on how much this additional information can shift the cost–benefit ratios of available clinical interventions. For diseases without potential therapies, even perfect prediction might not be clinically useful. By contrast, for diseases with effective preventive measures that are too costly or for which the risk–benefit balance is nearly neutral, small increments in predictive power could help effectively target preventive efforts, with substantial clinical impact.

The biologic pictures being revealed by genomewide association studies are still quite incomplete. We should strive for as complete a catalogue of validated risk variants as possible, through additional genomewide association studies and complementary approaches (such as exon-based or genomewide sequencing in sufficiently large samples) as they become available.

New biologic insights do not guarantee a rapid translation into clinical practice; the latter will require great effort by basic, translational, and clinical researchers. The difficulty in translation is not unique to genetic discoveries: nearly a century and three Nobel Prizes separate the determination of the chemical composition of cholesterol from the development of statins. Each discovery of a biologically relevant locus is a potential first step in a translational journey, and some journeys will be shorter than others. With a more complete collection of relevant genes and pathways, we can hope to shorten the interval between biologic knowledge and improved patient care.

In response to the skeptics, I offer a new bet. I predict that by the 2012 ASHG meeting, genomewide association studies will have

yielded important new biologic insights for at least four common diseases or polygenic traits — and that efforts to develop new and improved treatments and preventive measures on the basis of these insights will be well under way.

Dr. Hirschhorn reports receiving consulting fees from Correlagen and Ipsen, having an equity interest in Correlagen, receiving lecture fees from Pfizer, and receiving grant support from Novartis. No other potential conflict of interest relevant to this article was reported.

This article (10.1056/NEJMp0808934) was published at NEJM.org on April 15, 2009.

Dr. Hirschhorn is an associate professor in the Program in Genomics and the Divisions of Genetics and Endocrinology, Children's Hospital, Boston; an associate professor of genetics at Harvard Medical School, Boston; and an associate member and coordinator of the Metabolism Initiative at the Broad Institute of Harvard and MIT, Cambridge, MA.

1. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. Science 2008;322:881-8.
2. Mohlke KL, Boehnke M, Abecasis GR. Metabolic and cardiovascular traits: an abundance of recently identified common genetic variants. Hum Mol Genet 2008;17:R102-R108.
3. Lettre G, Rioux JD. Autoimmune diseases: insights from genome-wide association studies. Hum Mol Genet 2008;17:R116-R121.
4. Hirschhorn JN, Lettre G. Progress in genome-wide association studies of human height. Horm Res (in press).
5. Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, et al. Multiple genetic loci for bone mineral density and fractures. N Engl J Med 2008;358:2355-65.

Genetic Risk Prediction — Are We There Yet?

Peter Kraft, Ph.D., and David J. Hunter, M.B., B.S., Sc.D., M.P.H.

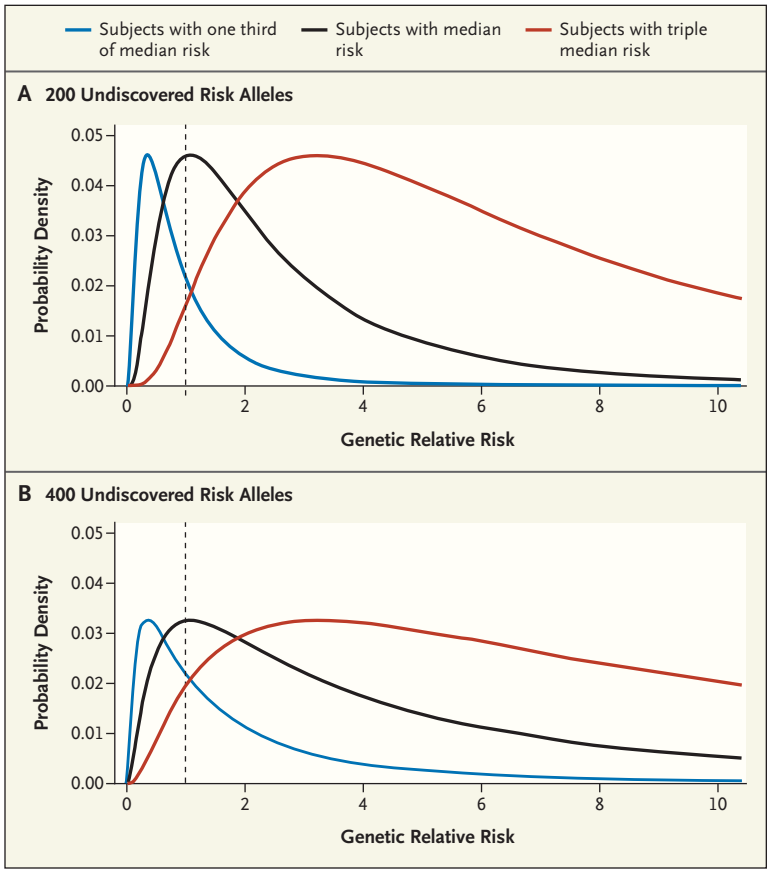
A major goal of the Human Genome Project was to facilitate the identification of inherited genetic variants that increase or decrease the risk of complex diseases. The completion of the International HapMap Project and the development of new methods for genotyping individual DNA samples at 500,000 or more loci have led to a wave of discoveries through genomewide association studies. These analyses have identified common genetic variants that are associated with the risk of more than 40 diseases and human phenotypes. Several companies have begun offering direct-to-consumer testing that uses the same single-nucleotide polymorphism chips that are used in genomewide association studies. These companies claim that such testing should be made available to consumers who are interested in their personal level of risk for the relevant diseases. Now, "risk tests" for specific diseases such as breast cancer are also being marketed to physicians and consumers.[1]

The availability of highly predictive and reasonably affordable tests of genetic predisposition to important diseases would have major clinical, social, and economic ramifications. But the great majority of the newly identified risk-marker alleles confer very small relative risks, ranging from 1.1 to 1.5,[2] even though such analyses meet stringent statistical criteria (i.e., the identification of associations with disease that have very small P values and hence are unlikely to be false positives). However, even when alleles that are associated with a modest increase in risk are combined, they generally have low discriminatory and predictive ability.[3]

One argument in favor of using the available genetic predictors is that some information must be better than no information, and we should not let the perfect be the enemy of the good by refusing to make use of our knowledge until it is more complete. Why not begin testing for common genetic variants whose associations with susceptibility to disease have been established?

The answer lies in the stability of the current risk estimates. Genetic variants conferring the highest relative risks are almost certainly overrepresented in the first wave of findings from genomewide association studies, since considerations of statistical power predict that they will be identified first. However, a striking fact about these first findings is that they collectively explain only a very small proportion of the underlying genetic contribution to most studied diseases. (Some exceptions exist — notably, age-related macular degeneration, for which a few alleles explain a substantial fraction of the genetic contribution.) Several lines of evidence support this overall conclusion.

First, the relative risks that are found to be conferred by common risk genotypes account for only a small proportion of the sibling recurrence risk (or the risk that a sibling will also have the disease of interest). Second, in multivariate analyses of large epidemiologic data sets in which a family history of a disease is a risk factor, the inclusion of data regarding which subjects carry the known associated variants only minimally reduces the risk asso-



Distribution of True Genetic Relative Risk of Disease for Three Subjects, According to Their Current Estimated Risk and the Number of Undiscovered Risk Alleles for the Disease.

Panel A shows the genetic relative risk of disease for three subjects with differing estimated risks with the assumption that there are 200 undiscovered independent risk alleles for the disease, each with a frequency of 25% and a multiplicative relative risk of 1.1. Panel B shows the same scenario with the assumption that there are 400 such undiscovered risk alleles. Risks are relative to the prevalence in the general population. The dashed line represents a relative risk of 1.0, the population median risk.

ciated with a family history of the disease. Third, in the case of diseases that have been the focus of several genomewide association studies, some alleles have been detected more than once, but each study has identified multiple alleles that were not identified in other studies, suggesting that many more alleles remain to be discovered.

These factors suggest that many, rather than few, variant risk alleles are responsible for the majority of the inherited risk of each common disease. The good news is that pooling the results of multiple genomewide association studies has led to increased statistical power and the discovery of many new loci linked to small increases in the risks of major diseases and phenotypes. For example, pooling efforts have led to the identification of more than 16 new loci associated with diabetes and more than 30 loci linked to Crohn's disease. From a scientific perspective, we would like to know roughly how many risk loci remain to be discovered. From a clinical and policy perspective, we would like to know the extent to which the available associations are useful for measuring risk, both in absolute terms and relative to the expanded set of associations that are likely to be discovered in the next few years.

The table shows the approximate number of risk loci (with allele frequencies and relative risks similar to those of most markers that have been discovered through genomewide association studies) that would be needed to reach a level of risk equivalent to the sibling recurrence risks of complex diseases such as diabetes, heart disease, and many cancers. Even a relatively large genomewide association study (one with 5000 case subjects and 5000 control subjects) has a rather low power to detect any specific marker: 0.9% at a P value of 10^{-7}, the least stringent accepted measure of genomewide significance, for an allele with a frequency of 25% and a relative risk per allele of 1.1. However, because there are so many risk loci, the probability of detecting at least 1 is good (83% if there are 200 risk loci). Although we know of many more risk loci than we did 2 years ago, there are probably many more associations that are yet to be discovered for these complex diseases. Less common variants in the prevalence range of 0.5 to 5.0% also remain to be discovered.

Estimates of risk based on established locus associations are therefore likely to change substantially in the next few years. The graphs show the distribution of true genetic relative risks under the assumption that there are either 200 undiscovered locus associations (Panel A) or 400 undiscovered locus associations (Panel

Number of Risk Alleles Needed to Produce a Sibling Relative Risk of 1.5, 2.0, or 3.0.*

Relative Risk Per Allele	Sibling Relative Risk		
	1.5	2.0	3.0
		no. of risk alleles	
1.10	203–507	347–867	550–1374
1.20	51–135	87–231	138–367

* The number of risk alleles was calculated over a range of allele frequencies (10 to 90%); the minimum and maximum numbers are presented. All alleles were assumed to have the same frequency and relative risk and to be independent.

B), each with a risk-allele frequency of 25% and a per-allele relative risk of 1.1, for three hypothetical subjects: one who is estimated (on the basis of her current genetic profile) to have one third of the median level of risk, one who is estimated to have a median level of risk, and one who is estimated to have three times the median level of risk. There is a high degree of variability around the current estimates. If there are 200 undiscovered locus associations, more than 7% of the subjects who are estimated to have triple the median risk of disease actually have less than the median risk; if there are 400 undiscovered associations, 15% of such subjects would be in that category.

As the number of known risk loci increases, the correlation between the predicted risk and the actual risk will also increase.[4] But this correlation, though important, is only one of the factors determining whether knowledge of genetic risk is beneficial. The clinical value of a genetic test also depends on its sensitivity, specificity, and positive and negative predictive values; the costs and benefits of interventions; and the availability of data linking specific variants to improved clinical outcomes.[5] In particular, although we will be better able to distinguish subtle differences in risk as we discover more risk loci, most people will still be at or near the median level of risk. As a result, for less-common diseases (with a prevalence of 1% or less), the positive predictive value of a genetic test will almost always be low.

We are still too early in the cycle of discovery for most tests that are based on newly discovered associations to provide stable estimates of genetic risk for many diseases. Although the major findings are highly unlikely to be false positives, the identified variants do not contribute more than a small fraction of the inherited predisposition. Estimates that are based on combinations of the current risk alleles (even estimates indicating substantial relative risks for a very small number of persons who carry many risk alleles) will undergo constant revision as new loci are found. Such estimates are poor predictors of risk, both in absolute terms and in relation to risk estimators that will be available when more of the remaining locus associations are discovered.

The rapid progress being made through meta-analyses suggests that many more common variants conferring a risk of disease will be identified in the next several years, leading to increasing stability of individual risk estimates. Once risk estimates are more stable, the usefulness of genetic screening will need to be considered for each disease, and recommendations about potential interventions will need to be made for persons whose predicted risk exceeds some threshold. Although testing for inherited susceptibility on the basis of common risk alleles is premature for most diseases, the situation may be very different in just 2 or 3 years. Appropriate guidelines are urgently needed to help physicians advise patients who are considering this form of genetic testing as to how to interpret, and when to act on, the results as they become more stable.

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

This article (10.1056/NEJMp0810107) was published at NEJM.org on April 15, 2009.

Dr. Kraft is an associate professor of epidemiology and biostatistics at the Harvard School of Public Health, Boston. Dr. Hunter is a professor in the Departments of Epidemiology and Nutrition at the Harvard School of Public Health and serves as a statistical consultant to the *Journal*.

1. Couzin J. Genetics: DNA test for breast cancer risk draws criticism. Science 2008; 322:357.

2. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. J Clin Invest 2008; 118:1590-605.

3. Janssens AC, van Duijn CM. Genome-based prediction of common diseases: advances and prospects. Hum Mol Genet 2008;17:R166-R173.

4. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. Genome Res 2007;17:1520-8.

5. Katsanis SH, Javitt G, Hudson K. Public health: a case study of personalized medicine. Science 2008;320:53-4. [Erratum, Science 2008;320:316.]