In less than 5 years, genomewide association studies (GWASs) have completely changed the landscape of human genetic research. Our increasing knowledge of the human genome sequence and its variation (http://hapmap.ncbi.nlm.nih.gov/) and technological advances in the design of genotyping microarrays have been instrumental in this evolution; however, an essential factor for success has been the use of large international collaborations for assembling studies that encompass genomewide data for tens or hundreds of thousands of individuals. The number of firmly replicated trait-associated loci that have been identified in GWASs is impressive (http://www.genome.gov/gwastudies), but the contribution of individual single-nucleotide polymorphisms (SNPs) to the studied phenotypes is weak, with rare exceptions. In this article, this aspect of GWASs and some of its implications are discussed.

© 2011 American Association for Clinical Chemistry

Hidden Heritability

Heritability is the part of the variation of a trait in a population that is contributed by genetic differences among individuals. The heritabilities of common diseases that have been evaluated in family or twin studies are often quite high, frequently reaching 50%. GWAS results provide support for the concept of a polygenic framework underlying complex diseases in which genetic susceptibility (liability) is due to a large number of cumulative weak effects contributed by variants covering the entire range of allele frequencies. According to this model, common disorders may be considered quantitative traits (1).

For example, consider height, a trait with a high heritability (approximately 80%). A metaanalysis of GWASs (meta-GWAS) comprising >200,000 individuals revealed 180 statistically significant SNPs that together explained approximately 10% of the variance in adult height (2). In another GWAS of height involving about 4000 individuals, the investigators, instead of testing SNPs one at a time, used a model that accounts for the genetic relatedness among pairs of individuals to estimate the genetic variance explained by the entire set of 300,000 markers on the array. Interestingly, the analysis with this alternative approach was able to explain 45% of the variance of height in the population (3). This finding suggests that the genetic contribution to a trait estimated in a GWAS from statistically significant SNPs reveals only a small fraction of the heritability, not because heritability is missing (most causal variants are not appropriately tagged by measured SNPs) but because it is hidden in the random noise contributed by the numerous SNPs that are irrelevant for the trait. A straightforward consequence of this finding is that larger studies will always identify new associations. This phenomenon has been demonstrated for several traits. For example, a meta-GWAS of plasma lipids (4) that extended the sample size from 20,000 to 100,000 individuals led to the identification of 59 novel loci, for 95 loci total. Are such results an argument for conducting even larger studies? Actually, highly significant genetic effects found in GWASs are generally quite reproducible across different populations. On the other hand, the contribution of newly identified loci decreases as study size increases. The effect of these weak loci is predicted to be more heterogeneous across populations, and accounting for the complexity of the meta-GWAS design and confounding influences may be impossible, with the implication that many weak associations are devoid of any biological relevance.

Weak Effects and Disease Heterogeneity

An intriguing possibility that remains, however, is that an association may be biased downward or undetectable at some loci because the phenotype tested is heterogeneous or imprecise. In such a situation, a variant would be clinically relevant for a small set of patients but might remain undetected. For example, a meta-
GWAS of incident heart failure (HF) identified a single HF-associated locus near the USP3 gene (ubiquitin specific peptidase 3) (5). HF is a clinically and etiologically heterogeneous disease, however. In a GWAS of dilated cardiomyopathy—an uncommon form of HF (1 in 2500 adults) with a risk of familial recurrence of approximately 30%—investigators discovered and replicated the results for 2 HF-associated loci near the HSPB7 gene (heat shock 27kDa protein family, member 7 (cardiovascular)) and BAG3 (BCL2-associated athanogene 3) genes (6). Other common disorders that have been explored in GWASs and are heterogeneous include stroke and obesity, but even for diseases considered clinically homogeneous, efforts aimed at deeper phenotyping might increase the power of a GWAS to identify contributing loci.

Risk Scores

To account for the large number of trait-associated loci identified in meta-GWASs and to compensate for their weak effects, investigators have proposed risk scores that combine the contributions of several susceptibility alleles. Consider body mass index (BMI), for example. A meta-GWAS that included >250 000 participants identified 32 BMI-associated loci, but the entire set of BMI-increasing alleles explained only 1.45% of the BMI variance in an independent population-based sample (n = 8120), i.e., 3% of the genetic variance for a trait with a heritability of approximately 50% (7). The low predictive value of combinations of SNPs has also been reported for other quantitative traits and for several complex disorders, such as coronary artery disease (CAD) and late-onset Alzheimer disease (LOAD).

A meta-GWAS of CAD conducted by the CARDIoGRAM (Coronary Artery Disease Genome-wide Replication and Meta-analysis) consortium analyzed 14 GWASs of >20 000 CAD patients and 60 000 controls overall. This meta-GWAS confirmed 10 loci previously associated with CAD and replicated the results for 13 newly identified loci. Together, the 23 loci explained 10% of the heritability for CAD (i.e., approximately 5% of the variation in disease prevalence) (8). A meta-GWAS of LOAD that assembled the results for >20 000 patients and 40 000 controls identified 11 disease-associated loci. The strongest association was observed for the APOE e4 allele with LOAD was first discovered nearly 20 years ago in a study that compared 30 patients and 91 controls (10). Despite the major impact of the APOE polymorphism on LOAD, its use for diagnosis or prediction is not currently recommended. It is difficult to imagine how a risk score that would add the effect of the newly identified LOAD-associated loci to that of APOE would provide any additional benefit.

Genetic risk scores need to be evaluated in prospective cohorts. The results of preliminary prospective studies have largely been inconclusive regarding the clinical utility of such scores (11), and large consortia of prospective cohorts have currently been assembled to test risk score equations that incorporate the weak loci identified in recent meta-GWASs. Given the evidence summarized above, however, we anticipate little improvement in the prediction ability of risk scores that account for the cumulative effect of statistically significant weak loci. Whether a global genome approach to individual risk prediction would be feasible and reproducible is not obvious either (12).

Rare Variants

Disease-associated markers in GWASs are not directly responsible in most instances for the associations observed, but they are in linkage disequilibrium with one or several causal variants located in their vicinity. It has been proposed that synthetic association (i.e., the association of a genotyped SNP with a trait due to the presence of multiple rare variants on the same marker allele) might explain most of the associations observed in GWASs (13); however, the results of sequencing numerous disease-associated loci have not been very supportive of this model (14), which, in addition, predicts distributions of detected SNP risk alleles and heterogeneity across populations that are not consistent with the empirical data (15).

Most associations observed in GWASs appear to be the consequence of relatively common causal variants in linkage disequilibrium with the genotyped SNPs (16). The possibility remains, however, that rare variants with strong or moderate effects exist and contribute to common traits. For example, a study of a sample of 438 individuals with hypertriglyceridemia (HTG) revealed that 28.1% were carriers of rare missense mutations in 4 HTG-associated genes, whereas only 15.3% of controls were carriers; the rare variants identified accounted for 1.1% of the total variation in HTG diagnosis (17). This result is based on only 4 candidate genes. A better estimate of the contribution of rare variants to HTG would require a much broader exploration of the HTG-associated loci detected in GWASs; however, this and other studies have shown

---

3 Human genes: USP3, ubiquitin specific peptidase 3; HSPB7, heat shock 27kDa protein family, member 7 (cardiovascular); BAG3, BCL2-associated athanogene 3; APOE, apolipoprotein E; SORT1, sortilin 1.
that rare variants with strong or moderate effects exist and contribute to the variation of common traits.

Causal Pathways

Those who doubt the ability of GWASs to discover many key causal and physiological components of a trait should consider the example of circulating lipids. A single GWAS of circulating lipids was able to recover almost all loci laboriously identified by conventional approaches over a period of 30 years and to discover a large number of new ones (18). The most recent meta-GWAS of plasma lipids identified 95 lipid-associated loci, many of which harbor genes related to lipid metabolism (4). Functional studies are essential to carry the results of GWASs forward. One of the loci newly identified in GWASs of lipids is located in the chromosomal region 1p13.3. This locus was first discovered in a GWAS of CAD (19) and is actually the strongest genetic correlate of LDL cholesterol in the population (4). No consensus has developed yet on the mechanisms underlying this association; however, detailed functional studies have revealed a new pathway of lipoprotein assembly and secretion in the liver in which the SORT1 (sortilin 1) gene is a key player (20, 21).

What Comes Next?

Whether variants identified by GWASs or other approaches contribute to a trait’s heritability is not the main point. Major breakthroughs in human genetics with important medical implications frequently have originated via the discovery of rare variants that contribute almost nothing to the disease’s heritability in the population. Until recently, investigations of uncommon variants (a frequency of <1%) with moderate effects (e.g., an odds ratio >2) were hampered by the difficulty of identifying them by family linkage analysis and the impossibility of tagging them in GWASs. Such variants constitute a vast reservoir of genetic variations—including causal ones—that can now be explored with next-generation sequencing strategies in the context of optimized study design (22, 23). It is time for a shift in interest from heritability and prediction to causality and pathophysiology. Discovery strategies should now be optimized to identify uncommon variants with moderate effects, and new approaches to functional analysis should be developed to characterize them in a systematic way, with the hope that these approaches will reveal new perspectives for the diagnosis and treatment of common disorders.

References


Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors’ Disclosures or Potential Conflicts of Interest: No authors declared any potential conflicts of interest.