Genome-Wide Association Studies: Results from the First Few Years and Potential Implications for Clinical Medicine

Joel N. Hirschhorn and Zofia K. Z. Gajdos

Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115; Program in Genomics and Divisions of Genetics and Endocrinology, Children’s Hospital, Boston, Massachusetts 02115; Broad Institute, Cambridge, Massachusetts 02142; email: joelh@broadinstitute.org, zofia@broadinstitute.org

Keywords
human genetics, common disease, quantitative traits, personalized medicine, biological pathways, common variants

Abstract
Most common diseases and quantitative traits are heritable: determined in part by genetic variation within the population. The inheritance is typically polygenic in that combined effects of variants in numerous genes, plus nongenetic factors, determine outcome. The genes influencing common disease and quantitative traits remained largely unknown until the implementation in 2006 of genome-wide association (GWA) studies that comprehensively surveyed common genetic variation (frequency > 5%). By 2010, GWA studies identified > 1,000 genetic variants for polygenic traits. Typically, these variants together account for a modest fraction (10%–30%) of heritability, but they have highlighted genes in both known and new biological pathways and genes of unknown function. This initial effort prefigures new studies aimed at rarer variation and decades of functional work to decipher newly glimpsed biology. The greatest impact of GWA studies may not be in predictive medicine but rather in the development over the next decades of therapies based on novel biological insights.
DISEASE GENETICS: SINGLE-GENE DISORDERS AND POLYGENIC DISEASES

Most diseases have an inherited component to susceptibility, meaning that genetic variation influences disease risk. The influence of genetics is clearest for Mendelian, single-gene disorders, where mutations in a particular gene are necessary and sufficient to cause the disease. The transmission of these mutations from parents to children typically results in clearly recognizable familial patterns of inheritance, because the penetrance of disease-causing mutations (the probability of being affected with disease) is usually quite high, even though clinical presentation and severity can vary. These highly penetrant mutations are usually evolutionarily deleterious and hence quite rare, and many—but not all—are coding changes that significantly disrupt or alter the protein encoded by the disease gene. Because of these characteristics, genetic linkage analysis with follow-up sequencing, which is particularly well suited to finding highly penetrant mutations with clear functional consequences, was highly successful for mapping and identifying many Mendelian disease genes (1, 2).

Most common diseases (for example, cancer, coronary heart disease, diabetes, and hypertension) and even quantitative traits (such as height, weight, blood pressure, and blood chemistries) also have an inherited component. For most common diseases, heritability—the fraction of variation in risk in a population that is attributable to inherited factors—typically ranges from 30% to well over 50% (for example, see 3–5). For quantitative traits, heritability estimates are often above 50% and can approach 80%–90% (6–8). These diseases and traits are polygenic, meaning that variation in many genes typically combine to influence phenotype (as do nongenetic factors such as environmental exposures). Because no single gene carries mutations that are sufficient to determine the outcome, families do not display clearly recognizable patterns of inheritance.

Given the success of linkage analysis in Mendelian disorders, it was natural to attempt linkage analysis for polygenic disorders. However, with a few exceptions, large linkage studies were unsuccessful at identifying regions that met stringent statistical thresholds (2). These important negative results suggested that, for most polygenic diseases and quantitative traits, no individual gene contains variants that together account for a substantial fraction of heritability. Because linkage has low power to detect genetic variants with modest effects (9), especially if, as had been hypothesized, they are common in the population (10–13), a different approach was proposed to uncover the genetic basis of common disease: association studies (see sidebar “Comparison of Association and Linkage Analysis”).

GENETIC ASSOCIATION STUDIES FOR POLYGENIC DISEASES AND TRAITS

The initial association studies were candidate gene studies, in which one or a handful of sequence variants from a single gene were tested for correlation with disease. Associations with human leukocyte antigen (HLA) were among the first that could be readily reproduced, in part because these often turned out to have large effects on disease risk that could be reproducibly demonstrated even in small samples of tens or hundreds of patients. For example, nearly half of the excess risk to siblings of patients with type 1 diabetes is explained by variation at multiple genes within the HLA locus (see 14, 15 and references therein). A few early associations with type 1 diabetes and inflammatory bowel disease (IBD) reflected effect sizes that were large enough to produce signals in linkage analysis (14–16), and perhaps created expectations that many common diseases would have variants with large effect sizes. Although several common variants outside of the HLA locus turned out to have moderate or large effects on diseases and quantitative traits (17–23), most associated common variants have modest...
effects that require large association studies for reliable detection (24, 25) and do not produce readily apparent signals even in extremely large linkage studies (2, 9).

The modest effect sizes of common variants meant that early association studies, which were limited by logistical considerations to handfuls of variants tested in hundreds of samples, typically did not provide strong evidence of association. Furthermore, nominally significant ($p < 0.05$) results were often interpreted as definitive, when in fact much lower $p$ values are required to demonstrate that an association is likely valid (2, 25–27). Thus, false positive and false negative results of association studies were commonplace (24), and the problem was likely exacerbated by technical artifacts and the presence of multiple genetic ancestries within individual studies (population stratification) (28, 29). Gradually, well-powered and well-controlled studies emerged that employed large samples, methods to detect and correct for population stratification (30), appropriate statistical thresholds, and robust genotyping methods. With these new standards in place, association studies slowly began to yield more reproducible results. With the development of genome-wide association (GWA) studies, a more comprehensive approach became available that could greatly accelerate progress.

RESULTS OF THE FIRST WAVES OF GENOME-WIDE ASSOCIATION (GWA) STUDIES

The transition from candidate gene association studies to GWA studies required four key technical and scientific advances: the discovery of widespread correlation (linkage disequilibrium) between groups of nearby common variants (31); catalogues of common variants and their correlations (32, 33); the development of high-throughput genotyping methods that could assay hundreds of thousands of genetic markers, typically single-nucleotide polymorphisms (SNPs) (34); and statistical methods to use the correlations between common variants to impute genotypes at additional variants not directly assayed by the genotyping platforms (35, 36). With the dramatic increase in scale from candidate to GWA studies, increasing rigor and sample sizes were essential to avoid a flood of false positive and false negative results. Fortunately, the importance of stringent statistical thresholds, high-quality data, and careful avoidance of artifacts were already becoming widely accepted when GWA studies became feasible. Consequently, most of the results from GWA studies that surpass these stringent thresholds of significance—“genome-wide significance,”

COMPARISON OF ASSOCIATION AND LINKAGE ANALYSIS

Association studies, like linkage studies, are designed to identify regions of the genome that, when inherited, influence phenotype. In a genetic association study of disease, DNA sequence variants are genotyped in affected individuals and in controls (who can be relatives or unrelated individuals of similar genetic ancestry). The frequencies of genotypes (e.g., C/C versus C/T versus T/T) or alleles (C versus T) are compared, and if one genotype or allele is statistically significantly more frequent in cases than in controls, the variant is said to be associated with the disease. Association studies of quantitative traits are identical, except that variants are tested for a correlation between genotype and trait values. The fundamental difference between association and linkage studies is that linkage analysis tests for correlation between a genomic region and phenotype in families, whereas association analysis searches for correlation in the population (essentially treating the population as a single large family with an unknown pedigree). A major consequence of this difference is that association analysis is better suited to finding common genetic variants with more modest effects, whereas linkage analysis is better powered to identify regions that contain variants (common or rare) that in aggregate have a more substantial effect (see 2, 9). Furthermore, because close relatives share stretches of genetic material over a larger distance than do unrelated individuals, association analysis can map disease-causing variants to a smaller region, defined by the extent of correlation between variants in the population (linkage disequilibrium, or LD). Although these regions are smaller, they often contain multiple genes; thus, association studies often can highlight a small region but often cannot pinpoint a causal gene within that region.
typically a $p$ value threshold of $5 \times 10^{-8}$ (25, 27) have been replicable.

The first widely replicable result from GWA studies was the association between variation at the complement factor H ($\text{CFH}$) gene and age-related macular degeneration (AMD) (18). This discovery was the earliest of its kind in part because variation at $\text{CFH}$ has a large effect—greater than fourfold—on AMD risk. Identification of associated variants for other diseases and traits, where effect sizes are generally much more modest, awaited more robust genotyping methods that could be applied in large samples. The first such studies, published in 2006 and the first half of 2007, typically involved at most a few thousand individuals. Although some of these did not identify any clear associations, most identified one or a few common genetic variants that were associated at genome-wide levels of significance with common diseases such as IBD (26, 37–39), prostate cancer (40, 41), type 2 diabetes (26, 42–46), obesity (47), coronary heart disease (26, 48), breast cancer (49, 50), and type 1 diabetes (26, 51). The GWA studies successfully completed by the end of 2007 are too numerous to list here, but these identified many more novel associations between common variants and polygenic diseases or quantitative traits.

These initial discoveries established several themes. First, many of the associated common genetic variants did not alter any predicted proteins, and some were hundreds of kilobases away from the nearest gene, suggesting that variation in the regulation of gene expression—perhaps exerted over long distances—was likely to be important in polygenic diseases and traits. These results largely contrast with those of single-gene disorders: Variants known to underlie single-gene disorders often are in the coding region and severely alter protein function (52). Second, the individual effect sizes of the variants from GWA studies were modest, typically explaining <1% of the population variation in phenotype. Thus, the initial GWA studies did not substantially improve prediction of disease risk (53), again in contrast to the experience from single-gene disorders. Third, the GWA results were largely replicable in additional studies (25, 27), demonstrating the utility of combining careful study design with sufficient power to achieve rigorous statistical thresholds, despite recent assertions that many GWA results are artifactual (54). Finally, although some associated variants were in or near known candidate genes, most associations highlighted genomic regions that contained no genes with previously known biological relevance, suggesting that these unbiased, genome-wide genetic studies were providing insights into novel biology (55). This pattern is similar to the experience from single-gene disorders, where the identification of disease-causing genes often opened up new windows into disease biology (2).

With the recognition that modest effect sizes meant that large sample sizes were critical to achieve genome-wide statistical significance, the next two years saw the formation of multiple consortia. Multiple groups with individual GWA studies joined forces to increase sample size through meta-analysis of association results from each component study. These efforts required a different, team-based scientific culture that recognized not only the efforts of key individuals but also the valuable and substantial contributions of large numbers (typically >100) of investigators. The result of these collaborations was an explosion in new associations. More than 1,000 genome-wide significant associations between common genetic variants and polygenic traits and diseases have been cataloged by Manolio’s group at the National Human Genome Research Institute (56; http://www.genome.gov/gwastudies/), and most of these were published in 2009 and 2010.

The list of associations is too extensive and too rapidly expanding to capture comprehensively here; instead, we briefly review the status (as of August 2010) of an illustrative group of diseases and quantitative traits: IBD, type 2 diabetes, autism, lipid levels, height, and hemoglobin F concentrations (Table 1). Given the rate of discovery, the numbers of associations will no doubt continue to grow rapidly,
but the principles illustrated by these examples are likely to remain relevant.

Inflammatory Bowel Disease

Prior to GWA studies, several common variants had been associated with IBD. These were found by association studies that focused on genes within regions highlighted by linkage studies or by large-scale studies of nonsynonymous coding variants (see 16, 57 and references therein). Subsequent GWA studies identified 32 loci associated with Crohn’s disease (CD) and additional loci associated with ulcerative colitis (UC), with partial overlap between the CD- and UC-associated loci (see 16, 57 and references therein). The CD-associated loci account for ∼20% of the heritable portion of disease risk, raising the theoretical possibility of predictive and/or diagnostic genetic testing. However, the greater impact of these studies has been to generate new understanding of the pathogenesis of IBD. The associated loci define several biological pathways as causally involved in pathophysiology, including innate immunity, IL23 signaling/Th17 lymphocytes, autophagy, epithelial barrier integrity, and the secondary immune response (16, 57). Critically, several of these pathways were not appreciated as being involved in the pathogenesis of IBD until the GWA results emerged. Some of these pathways are active areas of research that may accelerate the development of new therapies.

The comparison between the UC- and CD-associated loci has defined both common and distinct pathways across these two related diseases (57), and the loci that are uniquely associated with UC or CD might eventually help distinguish these two diseases, which in some patients can be hard to distinguish clinically.

Type 2 Diabetes

Like IBD, type 2 diabetes had been associated with a few common variants prior to GWA studies (see 63 and references therein).}

Table 1 Characteristics of genome-wide association (GWA) discoveries for several exemplary common diseases and quantitative traits

<table>
<thead>
<tr>
<th>Disease/TRAIT</th>
<th>Number of loci</th>
<th>Heritability explained</th>
<th>Overlap with Mendelian disorders</th>
<th>Drug targets within loci?</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inflammatory bowel disease</td>
<td>&gt;40</td>
<td>∼20%</td>
<td>N/A</td>
<td>No</td>
<td>Several clear, novel pathways guiding new therapeutic approaches</td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>32</td>
<td>∼10%</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Autism</td>
<td>1–2</td>
<td>Low</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
</tr>
<tr>
<td>Lipid levels</td>
<td>95 across 3 traits</td>
<td>∼25–30%</td>
<td>Near complete</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Height</td>
<td>180</td>
<td>∼12%</td>
<td>Many</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Hemoglobin F levels</td>
<td>3</td>
<td>∼35%</td>
<td>Yes</td>
<td>Yes</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Abbreviations: N/A: Not applicable, as precise sites of action of therapies are not known, or therapies are available, or no known Mendelian forms of disease exist. UC: ulcerative colitis; CD: Crohn’s disease.
dozen GWA studies identified a total of 32 loci that together account for \(\sim 10\%\) of the heritable portion of disease risk (see 64 and references therein). Although the fraction of variability in disease risk explained by these variants is still low, the variants can add information to known risk factors, particularly earlier in life (65). The variants are also beginning to outline known and novel aspects of disease pathogenesis. Some of the loci highlight known genes, such as those encoding monogenic forms of diabetes (neonatal diabetes, mature-onset diabetes of youth, or Wolfram syndrome), genes involved in insulin signaling \((IRS1)\), or genes encoding the sites of action of antidiabetic medications \((KCNJ11, PPARG)\). Few pathways other than the cell cycle were unambiguously defined by the genes within associated loci, but some of the genes suggest the involvement of less well-studied biological processes such as circadian rhythms, beta cell zinc transport, and RNA binding proteins. These results suggest that many of the pathways underlying the pathophysiology of type 2 diabetes have yet to be defined or annotated. Examination of the effects of these variants in nondiabetic individuals demonstrated that many associated variants increase diabetes risk through diminished beta cell function, although some clearly affect insulin action.

Another interesting finding was that many of the type 2 diabetes loci overlap with loci associated with other diseases or traits. In some cases the associated variants are the same (and have pleiotropic effects), but in many other cases the associated variants differ, suggesting that the variants exert different regulatory effects on genes that play a role in the pathogenesis of multiple diseases (64). Similar overlaps have been seen with variants that influence the risk of different autoimmune diseases (15) or different types of cancer, most notably at several locations tens to hundreds of kilobases upstream of the \(MYC\) gene (66). Finally, many of the variants that influence diabetes risk are quite distant from any known genes, raising the possibility of long-range regulation. Indeed, at the \(MYC\) cancer-associated locus looping between elements separated by hundreds of kilobases has been documented, providing a possible mechanism for regulation over large distances (67).

**Autism**

Initial GWA studies of autism did not identify any associated common variants, and only recently have two common variant associations reached (or nearly reached) genome-wide levels of significance (68, 69). By contrast, several studies have shown that rare structural variants (deletions or duplications) or point mutations can have substantial effects on disease risk (70–72). The contrast between autism and most other common diseases suggests that the genetic architecture of autism, and perhaps other neuropsychiatric disorders such as schizophrenia (70, 71), may be somewhat different from that of type 2 diabetes, IBD, and many other common disorders.

**Lipid Levels**

Prior to the advent of GWA studies, several genetic associations had already been identified with levels of HDL cholesterol, LDL cholesterol, and triglycerides, including both common and rare variants in known candidate genes (73, 74). GWA studies have since identified nearly 100 genetic loci with common variants that together explain 25%–30% of the population variation in these lipid levels (see 73, 75 and references therein). Almost every gene known to underlie a single-gene disorder of lipid metabolism has also been identified by the GWA studies, not only demonstrating that GWA studies are recapitulating the known biology, but also suggesting that many of the other genes identified by GWA studies will be novel regulators of lipid metabolism and have rare as well as common variants that influence lipid levels. Indeed, of the many genes that are new candidates for playing a role in lipid metabolism, several have recently been validated. By using in vitro and in vivo models, four genes identified in GWA studies, \(PPPIRB, SORT1, TTC39B,\) and \(GALNT2\), have been confirmed as regulating human lipid levels (75, 76). In the case of \(SORT1\), there is
compelling evidence that a particular regulatory variant is responsible for the variation in LDL cholesterol levels (76). Two genes encoding sites of action of lipid-lowering medications were highlighted in the GWA studies, suggesting that some of the other genes with no previously suspected connections to known lipid metabolism pathways may provide suitable new drug targets. The variants discovered by GWA studies of population-based samples have also been examined for their ability to predict clinical disease, and have been shown to explain part of the risk for clinically defined hypertriglyceridemia (77). Furthermore, some of these variants contribute to the risk of myocardial infarction, and, together with other variants that are associated with coronary heart disease, they provide nearly the same discriminative power for risk of coronary heart disease as do individual traditional clinical risk factors such as blood pressure (78). (We note that when assessing predictive models, it is important that the outcomes being predicted align well with the phenotypes associated with the variants used in the model.)

**Height**

Height is the classic polygenic trait (79), is highly heritable (8), and is routinely measured in large populations. Thus, quite large GWA studies of height are feasible and may provide early general insights into the genetic architecture of complex traits. Prior to GWA studies, no common variants were known to be associated with height, but a series of increasingly large GWA studies has now been completed. The most recent involved a first stage of 129,000 individuals and a replication stage of up to \( \sim 50,000 \) individuals; it identified variants at 180 loci that together explain \( \sim 12\% \) of the heritable variation in height (80). Power considerations suggested that there are several hundred more common variants yet to be discovered that have similar effect sizes on height, and that together these common variants will explain at least \( \sim 2\% \) of heritability. A separate analysis, using data from all the SNPs in a single GWA study, suggested that about half of the heritability of height can be accounted for by additive effects of a large number of common (>5% frequency) variants (81). At least 19 height-associated loci have multiple independently associated common variants, suggesting that many of the as-yet-unidentified variants may lie in already-identified loci. An enrichment of associated common missense variants showed that, at least for some of the associated loci, common variants themselves—rather than collections of rare variants—were responsible for the signals of association. Finally, the studies of height also confirmed that, if enough loci are identified, GWA studies can highlight biological pathways. For height, these include groups of genes involved in chromatin structure, hedgehog signaling, TGF-\( \beta \) signaling, growth hormone/IGF-1 pathways, BMP/noggin signaling, and the extracellular matrix (80). Interestingly, some of these highly relevant pathways emerged only after hundreds of associations were identified, suggesting that very large numbers of loci may have a greater impact on implicating biological processes than on explaining a larger fraction of phenotypic variation.

**Hemoglobin F Levels**

During human development, globin gene expression shifts from the fetal form (hemoglobin F, HbF) to the adult form (hemoglobin A), but detectable levels of HbF can still be found in adults due to residual expression of the gamma globin gene. HbF levels are an important modifier of the severity of sickle-cell disease, thalassemia, and other hemoglobinopathies, and are under genetic regulation, partially explained by rare and common \textit{cis}-acting variation at the globin locus (21). However, despite decades of intensive efforts, the genes responsible for regulating the switch from gamma to beta globin expression were not known. The first GWA studies identified common variants that accounted for >30% of the population variability in HbF levels, including variation at the globin locus and multiple variants at two other loci, near BCL11A and MYB-HBS1L (19, 20, 22). These
variants can account for some of the variability in severity of sickle-cell disease or thalassemia (22, 82), but the major impact of these discoveries is likely to be through the new understanding of globin regulation. Specifically, the BCL11A protein was rapidly shown to be a key regulator of globin switching in vitro and in vivo (83, 84). Thus, GWA studies have opened new areas of research that could plausibly lead to new therapies targeting these proteins.

THE NEXT WAVE OF GENETIC STUDIES OF POLYGENIC TRAITS AND DISEASES

The first three years’ worth of GWA studies discovered numerous associated genetic loci for many different polygenic traits and diseases. However, the variants identified thus far generally explain less than half (and in a number of cases less than 5%) of the heritability. Several possible explanations of this “missing heritability” have been proposed (81, 85, 86). These include a role for additional common variants of small effect, a role for less common variants (including variants with frequencies between 0.5% and 5%, which have not been well surveyed by current GWA studies), the presence of even rarer variants (those that are essentially private to each unrelated individual), nonadditive interactions between variants, and epigenetic effects. Regardless, these initial GWA studies, along with the failure of linkage studies to generate reproducible signals, have provided strong evidence that many common diseases and quantitative traits are quite polygenic, and this is likely an important contributor to as-yet-unexplained variance.

Indeed, despite the large sizes of GWA studies to date, it remains possible that much of the unexplained variance can be accounted for by additional common variants that have not yet been detected. Recent work suggests that common variants of small effect may account for up to half of the heritability of phenotypes such as height and even schizophrenia (80, 81, 87), so further increases in sample size will likely yield new associated variants. Current GWA studies have largely interrogated variants with frequencies of at least 5%–10% in the population. The next versions of GWA studies could build on data from the 1000 Genomes Project (http://www.1000genomes.org) to identify additional associated variants, as exemplified by a recent candidate gene sequencing study of type 1 diabetes that found associated variants in the 1% frequency range in IFIH1 (88).

Advances in sequencing technologies will enable genome-wide sequencing studies to assess more comprehensively the role of rare variation. These will no doubt identify rare variation that influences polygenic traits and diseases, based on early successes of classic candidate gene sequencing studies of obesity, lipid levels, and blood pressure (74, 89–91). However, implementation and interpretation of sequencing studies in polygenic traits and diseases will not be straightforward. First, methods for rigorous analysis, high-quality data generation, careful interpretation, and avoidance of artifact (especially confounding by ancestry) have not been well-established for sequencing studies. Second, in sequencing studies, rare variants must be analyzed in aggregate, and it may be difficult to distinguish rare functional variants from the background of rare neutral variants, particularly for non-coding variation. Finally, the lack of strong linkage signals provides evidence that the contribution of rare variants at any one gene will be modest, so it is likely that large samples will be required, perhaps comparable to those that have been needed for GWA studies (2).

IMPACT OF GWA STUDIES ON CLINICAL MEDICINE

The identification of genetic loci can have two major impacts on clinical medicine—through prediction of future outcomes or through elucidation of underlying biology (Figure 1). Prediction has the potential for immediate impact, although, as discussed below and by others (53, 92), the clinical utility of predictive genetic tests depends both on the degree of additional prediction provided by the genetic markers and on
whether the clinical course of action would be altered by predictive information. By contrast, the insights into biology and pathophysiology are likely to develop over a longer time frame.

**Prediction Using Genetics: The Potential for “Personalized Medicine”**

Much of the immediate focus on the discoveries from GWA studies has been on the feasibility of using genetic testing to predict disease or to guide therapy. One illustration is the rapid appearance of direct-to-consumer genetic testing companies that utilize common variants to provide estimates of relative risk, but not absolute risk (93). However, it is not certain that current or even future collections of associated variants will be clinically useful for prediction. First, the level of prediction thus far is typically modest, although in some cases the discriminative power of common genotypes is comparable to clinical tests that are in widespread use [such as LDL cholesterol for the prediction of myocardial infarction (78)]. More critically, the clinical utility of predictive models depends on their ability to alter clinical management, and this in turn depends on the availability of clinical interventions that are not universally indicated for reasons of safety or cost, but that might on balance be beneficial or cost-effective for higher-risk subgroups of patients (Figure 1).

There are many examples of genetics influencing clinical management for Mendelian or near-Mendelian traits and diseases, including diagnosis, prognosis, and reproductive counseling for many single-gene disorders, as well as blood typing prior to transfusion. In a few polygenic cases, particularly those involving drug response, the predictive information is strong and is used to make management decisions; for example, patients are HLA-typed to avoid hypersensitivity reactions to abacavir (94). Other GWA results in the area of pharmacogenetics may lead to similar clinical algorithms that incorporate genetic testing, perhaps with genetically guided interventions to avoid adverse reactions (95). In other cases, predictive power is strong but there are no preventive measures that are not universally indicated, so the associations do not affect clinical management. Examples include type 1 diabetes and AMD. For yet other diseases, the level of prediction may be lower, but might still be sufficient to identify a subgroup of patients who would benefit from an intervention that is either too expensive or too risky to apply to the general population. One possible example is the use of intensive lifestyle management and/or pharmacotherapy in patients who are at high risk of developing type 2 diabetes; although these interventions are effective (96), they are too costly for...
universal adoption, but genetic test results in combination with other known risk factors might eventually identify a high-risk group for whom the interventions would be cost-effective. As predictive power continues to improve, clinical trials that test whether it is useful to take genetic factors into consideration could become more informative. Finally, we note that genetic risk scores would most appropriately supplement rather than replace traditional risk estimates, and should be evaluated in this context.

Understanding Human Biology and Pathophysiology

Looking into the future, the greatest impact of GWA studies (and genetic studies more generally) will likely flow from expanding our understanding of underlying biology in patients. GWA studies of polygenic diseases and traits recapitulate many known biological pathways relevant to those diseases and traits. Furthermore, the delineation of pathways becomes clearer as the number of loci increases, with the picture continuing to sharpen even when the number of loci reaches the hundreds (80), so the continued identification of loci may aid in the identification of new pathways. Even more important, the sites of action of many known therapies have already been highlighted by associated variants (55, 75, 80), suggesting that the other loci may provide clues to new therapies. It is clear that many GWA loci do not contain genes that had previously been suspected as biological candidates. In some cases (as described in this review), these new genes identify well-studied biological pathways that were not previously connected to a particular disease, such as complement factor H and AMD, or autophagy and Crohn’s disease. In these cases, efforts directed toward developing potential therapies could begin immediately. In others, more work is needed to understand the clues that have been provided by GWA studies.

It is important to realize that, because of the genetic correlation (linkage disequilibrium) between nearby variants (see sidebar “Comparison of Association and Linkage Analysis”), association studies do not usually pinpoint the variant responsible for the association, or even the gene that is affected by the associated variant. Instead, association studies typically highlight a small region that may contain several genes (although sometimes the relevant gene may be obvious from previous genetic or biological studies). Although many GWA studies typically report the nearest gene(s) as a way of identifying the location of an associated variant, connecting a specific gene with a trait or disease generally requires additional genetic, biological, or functional data. Thus, additional functional or genetic studies (e.g., sequencing) will often be needed in order to identify the relevant gene as a prerequisite to deciphering the relevant biology.

Even if the relevant genes and biological pathways are understood, the time frame between biological knowledge and a new therapy can be long, as illustrated by the two decades that elapsed between Goldstein & Brown’s work implicating HMGCR and cholesterol metabolism in atherosclerosis (97) and the successful use of HMGCR inhibitors (statins) to prevent coronary artery disease (98). Where the genes and/or biology are unknown, the time frame will be longer. Still, GWA studies (and the studies that will follow in their wake) hold the promise of new and better therapies that target previously unsuspected aspects of disease biology that are genetically validated as relevant in patients. The promise of revolutionizing clinical medicine is great enough to justify follow-up studies based on the findings from these genetic studies. However, the likely time frame of these follow-up studies is long enough that the true level of success of genetic studies of common diseases and polygenic traits will be best judged in the coming decades rather than in the next few months or years.
DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Contents

Role of Postmarketing Surveillance in Contemporary Medicine
Janet Woodcock, Rachel E. Bebrman, and Gerald J. Dal Pan ........................................... 1

Genome-Wide Association Studies: Results from the First Few Years and Potential Implications for Clinical Medicine
Joel N. Hirschhorn and Zofia K.Z. Gajdos ................................................................. 11

Imaging of Atherosclerosis
D.R.J. Owen, A.C. Lindsay, R.P. Choudhury, and Z.A. Fayad ........................................ 25

Novel Oral Factor Xa and Thrombin Inhibitors in the Management of Thromboembolism
Bengt I. Eriksson, Daniel J. Quinlan, and John W. Eikelboom ......................................... 41

The Fabry Cardiomyopathy: Models for the Cardiologist
Frank Weidemann, Markus Niemann, David G. Warnock, Georg Ertl, and Christoph Wanner ........................................................ 59

Kawasaki Disease: Novel Insights into Etiology and Genetic Susceptibility
Anne H. Rowley ................................................................. 69

State of the Art in Therapeutic Hypothermia
Joshua W. Lampe and Lance B. Becker ................................................................. 79

Therapeutic Potential of Lung Epithelial Progenitor Cells Derived from Embryonic and Induced Pluripotent Stem Cells
Rick A. Wetsel, Dachun Wang, and Daniel G. Calame ............................................. 95

Therapeutics Development for Cystic Fibrosis: A Successful Model for a Multisystem Genetic Disease
Melissa A. Ashlock and Eric R. Olson ................................................................. 107

Early Events in Sexual Transmission of HIV and SIV and Opportunities for Interventions
Ashley T. Haase .................................................................................. 127

HIV Infection, Inflammation, Immunosenescence, and Aging
Steven G. Deeks .................................................................................. 141
The Increasing Burden of HIV-Associated Malignancies in Resource-Limited Regions
Corey Casper ............................................................... 157

Biliary Atresia: Will Blocking Inflammation Tame the Disease?
Kazuhiko Bessho and Jorge A. Bezerra ................................ 171

Advances in Palliative Medicine and End-of-Life Care
Janet L. Abraham .......................................................... 187

Clostridium difficile and Methicillin-Resistant Staphylococcus aureus:
Emerging Concepts in Vaccine Development
David C. Kaslow and John W. Shiver .................................. 201

Antiestrogens and Their Therapeutic Applications in Breast Cancer
and Other Diseases
Simak Ali, Laki Buluwela, and R. Charles Coombes .................. 217

Mechanisms of Endocrine Resistance in Breast Cancer
C. Kent Osborne and Rachel Schiff ..................................... 233

Multiple Myeloma
Jacob Laubach, Paul Richardson, and Kenneth Anderson ............ 249

Muscle Wasting in Cancer Cachexia: Clinical Implications, Diagnosis,
and Emerging Treatment Strategies
Shontelle Dodson, Vickie E. Baracos, Aminah Jatoi, William J. Evans,
David Cella, James T. Dalton, and Mitchell S. Steiner ................ 265

Pharmacogenetics of Endocrine Therapy for Breast Cancer
Michaela J. Higgins and Vered Stearns ................................ 281

Therapeutic Approaches for Women Predisposed to Breast Cancer
Katherine L. Nathanson and Susan M. Domchek ...................... 295

New Approaches to the Treatment of Osteoporosis
Barbara C. Silva and John P. Bilezikian ............................... 307

Regulation of Bone Mass by Serotonin: Molecular Biology
and Therapeutic Implications
Gerard Karsenty and Vijay K. Yadav .................................... 323

Alpha-1-Antitrypsin Deficiency: Importance of Proteasomal
and Autophagic Degradative Pathways in Disposal of Liver
Disease–Associated Protein Aggregates
David H. Perlmutter ...................................................... 333

Hepcidin and Disorders of Iron Metabolism
Tomas Ganz and Elizabeta Nemeth ...................................... 347
Interactions Between Gut Microbiota and Host Metabolism
Predisposing to Obesity and Diabetes
Giovanni Musso, Roberto Gambino, and Maurizio Cassader

The Brain-Gut Axis in Abdominal Pain Syndromes
Emman A. Mayer and Kirsten Tillisch

Cognitive Therapy: Current Status and Future Directions
Aaron T. Beck and David J.A. Dozois

Toward Fulfilling the Promise of Molecular Medicine
in Fragile X Syndrome
Dilja D. Krueger and Mark F. Bear

Stress- and Allostasis-Induced Brain Plasticity
Bruce S. McEwen and Peter J. Gianaros

Update on Sleep and Its Disorders
Allan I. Pack and Grace W. Pien

A Brain-Based Endophenotype for Major Depressive Disorder
Bradley S. Peterson and Myrna M. Weissman

Indexes
Cumulative Index of Contributing Authors, Volumes 58–62
Cumulative Index of Chapter Titles, Volumes 58–62

Errata
An online log of corrections to *Annual Review of Medicine* articles may be found at
http://med.annualreviews.org/errata.shtml