Published in final edited form as:

Science. 2009 October 9; 326(5950): 239–240. doi:10.1126/science.1182009.

Life After GWA Studies

Emmanouil T. Dermitzakis¹ and Andrew G. Clark²

Emmanouil T. Dermitzakis: emmanouil.dermitzakis@unige.ch; Andrew G. Clark: ac347@cornell.edu

¹ Department of Genetic Medicine and Development, University of Geneva Medical School, 1 Rue Michel-Servet, Geneva 1211, Switzerland ² Department of Molecular and Genetics, Cornell University, Ithaca, NY 14853, USA

Abstract

Genome-wide association findings should be integrated into a wider scope of information including biological processes and environments.

In the Hollywood movie “*GATACCA*,” an infant’s genome sequence is produced in seconds and the probabilities of dozens of chronic disorders roll off the screen. Although real-world progress in technologies for DNA sequencing seems to be approaching the science fiction version of genomics, the ability to predict an individual’s risk of chronic disease based on DNA sequence is lagging behind. How do we bridge this gap? Or is it time to reconsider the goal of accurately predicting individual risk?

The explosion of genome-wide association (GWA) studies has expanded the set of candidate genes and genomic regions for future study (1). Examples include underscoring the role of immunity in macular degeneration (2), and shifting the emphasis in type 2 diabetes risk from genetic factors affecting insulin resistance to those that influence insulin production (3). But progress in chronic disease etiology has been slow, and GWA results have not broken any floodgates of understanding. This is because the studies only nominate candidate villains, and it takes biological insight and studies of mechanism to learn how they erode our health.

Another overarching objective for GWA studies—to facilitate prediction of the likelihood of future disease—has taken a rockier road. The challenge is highlighted by the most striking general result that pervades all GWA studies—the magnitude of genetic effects is uniformly very small. Even for a trait with strong familial clustering, the strongest associations explain little of the genetic variance for the trait. For example, the heritability of stature is 80%, yet the top 20 candidate genetic variants identified in GWA analyses explain only 3% of the variance (4,5). Possible reasons include missing low frequency alleles, genetic heterogeneity of the trait, genotype-environment interaction, and epistasis (6). And the possible roles of epigenetic variation in mediating phenotypic differences also cannot be ignored. The lesson is that we do not yet fully grasp the genetic architecture of complex disorders in humans, and we will not be able to make accurate individual predictions of risk until we do.

Predicting individual risk of complex traits is a tall challenge, in part because of the context-dependent way that the genotype manifests its effects on disease risk. Moreover, any prediction of risk must somehow integrate over possible environments. Investment analysts apply Monte Carlo simulations to project the future value of investment accounts, sampling over many possible future economic scenarios. The future of gene-based diagnostics, at least for complex disorders, might similarly have to incorporate many possible environmental and other insults into their predictions of individual risk. It might be possible to represent a projected “norm of reaction” as the distribution of possible phenotypic outcomes of a particular genotype, but it

is nearly impossible to know how useful such a prediction could be. For many complex traits, the range of predictions with and without the genetic data may differ very little. But for others, the genetic information may impose constraints that would have useful predictive value.

Opportunities for improved accuracy in predicting disease risk appear especially bleak when one considers that individual prediction of complex traits in model organisms, such as the fly *Drosophila melanogaster*, is not much better than in humans. Precise effects of specific genetic variants on bristle number in flies have been mapped to several genes, and yet those same variants within a natural population have little bearing on bristle counts (7). Prediction of individual phenotypes is practiced in plant and animal breeding, where the genetic dimensionality is radically reduced. Modeling and prediction of complex phenotypes of inbred lines in the laboratory can be quite accurate, and yet when the same genes and phenotypes are projected into a free-living population, individual prediction entails too many additional variables and accuracy suffers.

So what comes next in complex disease research? The technology for identifying genetic differences is racing forward, with the 1000 Genomes Project (8) and efforts to accurately characterize copy number variation. The notion that whole-genome sequencing will become routine medical practice no longer seems so futuristic. But the biggest opportunity for making serious progress in understanding chronic disease risk lies in developing a deeper “biological awareness” into genomic approaches to the study of complex disorders. We tend to talk about pathways and processes as if they are discrete compartments of biology. But genes and their products contribute to a network of interactions that differ radically among tissues. Even our inference of gene regulatory networks suffers from being confined to a narrow biological context. A “candidate-tissue approach” still limits attention to a subset of tissues based on incomplete information and tenuous assumptions, and like candidate gene studies, there is no guarantee that these guesses are correct.

Though we have developed sophisticated statistical tools to analyze and identify genetic variants in the context of whole-organism phenotypes and to detect associations of effects that are far apart (in terms of molecular interactions), we have been unable to bridge the gap between the genetic lesion, or the biochemical effect, and the phenotype except in a few cases (9). A major breakthrough will be to predict and interpret the effect of mutational and biochemical changes in human cells and understand how this signal is transmitted spatially (among tissues) and temporally (spanning development) (10). Manipulation of induced pluripotent stem cells (11), may allow the analysis of diverse cell types from a number of individuals to examine the effects of genotypes in different biological contexts.

To date, GWA studies have relied on a rather unlikely model for the genetics of complex traits—that common DNA sequence variations (known as single nucleotide polymorphisms) with widespread (but marginal) effects will predominate. This was a sensible place to start, but perhaps it should not be surprising that common variants provide little help in predicting risk. With the arrival of data for ever rarer variants in large, well-designed cohort samples, we should now decide how much emphasis to place on individual prediction, and ask how we can improve the currently unsatisfying record. Large cohort studies should provide information on candidate genes and on combinations of factors that influence groups that are at risk, even if accurate individual risk prediction is never achieved. An attainable public health goal that deserves emphasis is to identify, for each individual, lifestyle choices that pose particularly elevated risk of chronic disease.

References and Notes

1. <http://www.genome.gov/26525384>

2. Klein RJ, et al. *Science* 2005;308:385. [PubMed: 15761122]
3. Zeggini E, et al. *Nat Genet* 2008;40:638. [PubMed: 18372903]
4. Weedon MN, et al. *Nat Genet* 2008;40:575. [PubMed: 18391952]
5. Lettre G, et al. *Nat Genet* 2008;40:584. [PubMed: 18391950]
6. Manolio TA, et al. *Nature* 2009;XXX
7. Macdonald SJ, Pastinen T, Long AD. *Genetics* 2005;171:1741. [PubMed: 16143618]
8. <http://www.1000genomes.org>
9. Schadt EE. *Nature* 2009;461:218. [PubMed: 19741703]
10. Dermitzakis ET. *Nat Genet* 2008;40:492. [PubMed: 18443581]
11. Meissner A, Wernig M, Jaenisch R. *Nat Biotechnol* 2007;25:1177. [PubMed: 17724450]