2011 Curt Stern Award Address[1]

David Altshuler[2,3,4,*]

I first want to thank the nominators and the Society for this wonderful award. It is an honor and more than a bit humbling to be in the company of the previous winners.

I'd also like to thank Aravinda for his kind and generous introduction. I have learned much from Aravinda over the years, and it is a personal pleasure to share the podium with him.

Given the time available, I won't attempt to review the science we've done over the last ten years. Rather, I would like to briefly touch on three themes.

First is the unique role of human genetics as a tool for biomedical research. Second is the critical importance of statistical rigor and reproducibility. Third is the changing nature of our field and how we rise to the grand challenge of our collective success.

First is genetic mapping and causal inference.

I have come to the view that two aspects of human genetics are particularly unique in the armamentarium of biomedicine: the unbiased nature of genetic mapping and the ability to draw causal inference in the human population.

Most of science is hypothesis driven and is thus hypothesis limited. Genetic mapping reveals new biological mechanisms without regard to our preconceptions about whether they involve known or novel genes, alter protein coding or noncoding regions, and are few in number or many.

Ed Lewis once wrote, "The laws of genetics have never depended upon knowing what the genes are chemically and would hold true even if they were made of green cheese."

Similarly, it seems to me that the virtue of genetic mapping is that it doesn't depend on our guesses about whether the variants should reside in the genes that biologists have previously studied or if they are to be found hundreds of kilobases from the nearest recognized functional element.

If unbiased search is the hallmark of our approach, then it is incumbent on us to see with clear eyes what the data are telling us even when the answers don't support our prior assumptions.

Of course, human geneticists are not unique in trying to find novel causes for disease. However, our approach is special because of its ability to support causal inference in the human population. Here, I don't mean "the" cause (there is never just one cause), but rather the key difference between a reactive process and one that participates in a causal chain.

The two most commonly taken approaches in biomedicine—experimental studies in the laboratory and observational studies in human populations—are powerful, but each has fundamental limitations. Model systems and cell cultures are highly tractable, but the findings are of uncertain relevance to patients. Observational studies in the human population can reveal correlations, but as a matter of logic, they do not support causal inference.

Why is it that observational studies do not support causal inference? At the core, the reasons are two: confounding and the arrow of time. The exposures are not assigned at random nor are they independent. For this reason, it is always possible that a measured variable reflects the influence of some other unmeasured causal factor. Moreover, expression levels, metabolites, and behaviors can and do change in response to disease and thus might follow, rather than precede, the root cause of disease.

In contrast, geneticists can safely assume the random assignment of gametes at meiosis and the independent assortment of unlinked segments due to recombination. (I'll note that as a result of linkage and linkage disequilibrium, this argument falls apart at close distances.)

[1]This article is based on the address given by the author at the meeting of the International Congress of Human Genetics on October 13, 2011 in Montreal, Quebec, Canada. The audio of the original address can be found at the web site of the American Society of Human Genetics

[2]Department of Molecular Biology and Medicine, Massachusetts General Hospital, Boston, MA 02114, USA; [3]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; [4]Departments of Genetics and Medicine, Harvard Medical School, Boston, MA, 02115, USA
*Correspondence: altshuler@molbio.mgh.harvard.edu

DOI 10.1016/j.ajhg.2012.02.017.

Moreover, the arrow of time is fixed at conception, and genes don't change in response to disease.

These two features are special to our approach. For these reasons, I believe that the unique role of human genetics is to expand the scope of knowledge of human biology and to focus our attention on those mechanisms that are causal rather than reactive.

But there is a fly in the ointment. The genome is very big, and each copy of the genome differs at millions of DNA variants. Few of these variants influence any disease, and those that do typically act in manner that is probabilistic, not deterministic.

If our goal is to find a small number of needles in a very large haystack—and it is—then our work demands a statistical approach, carried out with great rigor and skeptically interpreted under an appropriate null distribution. This brings me to my second point.

Over the last ten years, many of us have worked to enable the systematic and comprehensive testing of genetic variation for association to human diseases. This involved characterizing and cataloguing human sequencing variation, developing laboratory tools to measure that variation, developing analytical methods to find associations to phenotype, and performing studies of sufficient power to identify robust relationships between genetic variations and disease.

As we know, the first generation of such studies has been successful in identifying many novel genomic regions associated with a wide variety of diseases. Now, based on whole-genome and exome sequencing, a second generation of such studies is under way.

When I look back on the last five years, since genome-wide association studies became a new tool in our armamentarium, there is one element that strikes me as most important to today.

That element is the high standards set for statistical rigor in analysis and the demand that results be reproducible.

In the early days of linkage for complex traits and of candidate-gene association studies, there were many claims that proved to be impossible for replication. Oftentimes, these were in candidate genes that someone thought made sense and involved nonsynonymous changes that someone thought might be functional. Many worried that GWAS studies would drown in a sea of false claims.

History shows that the opposite has been the case. A set of standards were developed and stringently applied, ranging from proper controls for technical artifacts and population mismatching to interpretation under a null hypothesis that considers the entire genome to a requirement for replication prior to making claims. In return, the vast majority of GWAS findings that meet these standards have proven robust and reproducible.

I mention this because in the excitement around next-generation sequencing, findings that lack sufficient statistical evidence to justify the claim of a relationship to phenotype are again being reported, and there is a disturbing lack of concern for replication.

It almost appears as though some would argue that if a variant is rare and alters a protein or creates a large structural variant, then the need for sober analysis, genome-wide significance, and replication can be waived.

Karl Popper wrote, "The criterion of the scientific status of a theory is its falsifiability or refutability or testability."

In the realm of next-generation sequencing, this means that if the theory is that a given gene harbors mutations that influence a disease, then that theory must be stated in such a way that others can come along with an appropriately sized sample and reproduce or refute the finding. If others can't replicate the finding, then the theory should be revised or set aside until it survives attempts at refutation.

Having said this, I have full confidence that as the sample sizes increase and as convincing findings emerge, the field will embrace high standards for statistical evidence and replication of findings.

In fact, it now seems highly likely that when the history books are written, they will say that from 1980 to 2020 or so, genetic mapping revealed a vast number of causal factors for human diseases, first for rare single-gene disorders and later for diseases showing complex inheritance. They will describe that some diseases are caused by mutation at only one or a few genes but that most diseases are influenced by dozens to hundreds of genes. They will describe both common and rare mutations, very few of which act in a fully penetrant manner. Probability rather than determinism will become the linchpin of our thinking.

This genetic anatomy of human diseases will be a grand accomplishment shared by our entire field. It will lead to an inventory of disease-associated genes that students learn about in school, much like they do the bones of the hand or the enzymes of the Krebs cycle.

But for this not to be a pyrrhic victory, we will need to embrace the end of one way of thinking and focus on a new grand challenge: figuring out the biological functions of these many novel genes, their roles in pathophysiology, and how to leverage this new information to develop new and more effective approaches to prevention and treatment.

At present, much attention in our field focuses on the proposition that clinical value will derive directly from the sequencing and interpretation of genomes. In some cases, it has and it will.

More generally, the value of prediction pales in comparison to the reward of more effective prevention and treatment.

Of course, this is far from a new idea. The geneticists we admire the most are, appropriately, those who have not only found genes for a disease but who have also understood them and developed new therapies. A previous recipient of this award, Hal Dietz, described the path from the discovery of genes for Marfan syndrome to an unexpected pathophysiological mechanism to a novel treatment.

Hal is one of my heroes.

Nonetheless, it is sobering to recognize that in 2011, it remains a very difficult task to figure out the function of a new gene that is found convincingly to influence

a disease but that has no previously known biochemical or cellular function. Confronted with not too few such genes but too many, it is clear that we will need new frameworks and approaches.

More importantly, doing this will require us to commit the time and effort to doggedly pursue what we have found. Hal Dietz talks about a decade in the wilderness between the cloning of the Marfan gene and the breakthroughs that have led to therapies. Twenty years after the cloning of the gene for Huntington disease, my friends Jim Gusella and Marcy MacDonald and their many colleagues around the world are still working to crack the problem.

This is not a sign of the failure of genetics, as the newspapers sometime suggest, but simply what it takes to go from a truly new finding to something useful.

When I was in graduate school, Fred Winston's lab was next door. Fred's lab had performed genetic screens in yeast and identified a number of novel genes that played a role in transcription. Among these genes were components of histones. At that time, the idea that histones played a role in transcription was heretical.

I ate lunch every day with Joel Hirschhorn, a fellow graduate student from that lab. Joel worked for years to prove that this finding from genetics—that histones play a role in transcription—was real and important. His work was criticized by many leading lights in that field and was dismissed as irrelevant.

But today, we all know that histones play a critical role in transcription.

Similarly, Bruce Wightman, another friend in graduate school, worked in Gary Ruvkun's lab. Together with Victor Ambros, he and Gary performed positional cloning of the worm genes lin-4 and lin-14 and found a then inexplicable result: One of these genes encoded a tiny noncoding RNA complementary to the other. This observation didn't fit the dogma and was rejected by many as irrelevant or as a boutique exception.

It was a decade before the general importance of their finding of a microRNA was recognized.

I was lucky enough to encounter these inspiring examples early in my career. I took away from them that if a geneticist believes what comes out of his or her screen and studies it well and long enough, he or she can learn important and novel things.

This highlights an unstated tension in our field: We extol unbiased discovery as the sine qua non of human genetics, and yet, we are understandably frustrated when we uncover genes about which nothing is known. Too often, we move on to the next gene discovery rather than study what we've found. I've been guilty of that over the years, but I'm not alone.

Another sobering fact is that many of us might not be the right people, by training or inclination, to do this next phase of work. That means either that we have to retrain and dramatically shift gears or that we have to recruit others to take up the challenge. One way or another, it must be done, or the potential value of our collective accomplishment will be limited to fortune telling and wishful thinking.

Which brings me to collaboration.

The nature of genetics and genomics has required that many types of expertise be brought to bear on a single problem, spanning from the clinical to the technological to the statistical to the biological. Given the model of a single PI working alone, it was difficult to see how such a challenge could be met.

But necessity is the mother of invention. Faced with the choice of working alone and failing, people chose to work together. Instead of doing many underpowered studies, they did one well-powered study. Instead of staying in the comfort zone of their own subdiscipline, they sought out people with complementary skills.

It has been remarkable to see teams define goals that, at the time, seemed like they couldn't be achieved, and then accomplish them together. It's been great fun to share these years with a remarkable group of friends and colleagues.

Looking out for young people has been one of the key aspects of making collaboration work. Although it is possible for young people to get lost in a big project, such projects offer a rich opportunity for those with ideas and initiative. It is the job of the mentor to ensure that each trainee develops his or her own ideas, writes papers and give talks, and moves purposefully toward his or her own personal goals.

As we shift from gene discovery to function and therapy, the nature of the collaborations might change, but the need for collaboration will increase, not decrease.

In closing, I'd like to thank those people who have been my mentors, collaborators, and friends.

Eric Lander, who gave me my start in human genetics and who taught me to believe that anything is possible.

Stacey Gabriel and Mark Daly, who have been partners in establishing the Program in Medical and Population Genetics at the Broad Institute and in too many studies to name.

Joel Hirschhorn, whom I met the first day of medical school and with whom I return each year to teach human genetics to medical students.

Leif Groop, who 15 years ago welcomed me warmly as his visitor in Malmo, Sweden. Since that day, we have shared common cause in the goal of understanding the genetics of type 2 diabetes.

More recently, Mike Boehnke and Mark McCarthy, with whom we created the DIAGRAM Consortium and a community that together is tackling diabetes genetics.

The members of our lab, who have brought their ideas, energy, and good will to the work. Nothing in my professional life has given me more pleasure than to watch with awe as former trainees have gone on to make discoveries, lead their own labs, or thrive in leadership roles in research institutes and companies.

And finally, my parents, Julie and Alan Altshuler, my wife, Jill, and my sons, Zachary and Jason. I love my work, but my family is my life.

Thank you very much.