

# scienceWATCH<sup>®</sup>.com

TRACKING TRENDS & PERFORMANCE IN BASIC RESEARCH

[Interviews](#)[Analyses](#)[Data & Rankings](#)

2010 : May 2010 - Author Commentaries : Harvard's David Altshuler on the Complex Genetics of Diabetes

## AUTHOR COMMENTARIES - 2010

May 2010



### Harvard's David Altshuler on the Complex Genetics of Diabetes

As decades go, the last was the scientific equivalent of a gold rush: "One of the most prolific periods of discovery in human genetics," as a *New England Journal of Medicine* essay recently described it (J.N. Hirschhorn, 360[17]: 1699-1701, 2009). First the human genome sequence was published in 2003, followed five years later by a catalog, known as HapMap, of common genetic variation across that genome. By then, genome-wide association studies had begun documenting the association of common genetic variations with chronic conditions such as heart disease and **diabetes**. By the end of last year, geneticists had already identified more than 250 gene variants associated with over 40 different chronic diseases and human phenotypes.

At the forefront of this tidal wave of research is the Harvard geneticist and endocrinologist David Altshuler, whose 13 Hot Papers tracked during the course of 2009 placed him near the top of this publication's annual ranking of the year's hottest scientists (*Science Watch*, March/April 2010). Altshuler is a co-author on half a dozen papers that have received over 1,000 citations each in the past decade, and he's the first author on the 2005 *Nature* article—"A haplotype map of the human genome"—that has garnered more than 2,000 citations in less than five years (see [table](#) below). Another 20 of Altshuler's articles have garnered over 100 citations each.

Altshuler, 45, earned his Bachelor's of Science from the Massachusetts Institute of Technology in 1986 and then obtained an M.D. and his Ph.D. at Harvard in 1994. After doing his internship and residency at Massachusetts General Hospital, he began a clinical fellowship there in endocrinology in 1996, coincident with a postdoctoral fellowship in genetics with [Eric Lander](#) (see also) at MIT's Whitehead Institute. In 2000, Altshuler started his own laboratory at Massachusetts General Hospital and Harvard Medical School, where he's now a professor of genetics and medicine. Altshuler is also a founding member and deputy director of the Broad Institute of MIT and Harvard.

"...  
*surprisingly enough, we don't really have good methods yet to take a gene of unknown function and figure out what it's doing.*"

***Altshuler spoke to Science Watch from his home in Brookline, Massachusetts.***

[ScienceWatch Home](#)[Inside This Month...](#)[Interviews](#)[Featured Interviews](#)[Author Commentaries](#)[Institutional Interviews](#)[Journal Interviews](#)[Podcasts](#)[Analyses](#)[Featured Analyses](#)[What's Hot In...](#)[Special Topics](#)[Data & Rankings](#)[Sci-Bytes](#)[Fast Breaking Papers](#)[New Hot Papers](#)[Emerging Research Fronts](#)[Fast Moving Fronts](#)[Corporate Research Fronts](#)[Research Front Maps](#)[Current Classics](#)[Top Topics](#)[Rising Stars](#)[New Entrants](#)[Country Profiles](#)[About Science Watch](#)[Methodology](#)[Archives](#)[Contact Us](#)[RSS Feeds](#)

**SW:** Considering the enormous progress made in the last decade in human genetics, there's been considerable controversy about whether we'll ever learn enough to predict our risk of disease from our genotypes. Does this bother you?

In my mind, the primary goal is not to predict disease, nor to personalize medicine. It's to understand the biological systems that underlie common diseases—type 2 diabetes, for example, which is my research focus. The problem that got me started is very simple: why is it that some people living in our modern society and environment gain weight and others not? Why do some people who become **obese** also get diabetes, and others not?

If we could figure this out, we'd understand the biological basis of diabetes—and if we knew the basis we could rationally design approaches to prevention and therapy.

How do we figure out the basis of a disease? One approach is to study it in cells, and in mice—to hypothesize that diabetes is a disorder of insulin resistance, or of fat cells, or muscle. This approach can clearly identify fundamental biological mechanisms, but it can't tell us which mechanisms actually apply in patients. And it might, in fact, lead us astray.

So, another approach is based on the observation that the diseases run in families, and not to prejudge whether it's insulin resistance, or whether it's the fat cell or the muscle cell. Instead, we can simply follow the disease in families and in the population, and track down the genes that influence risk. And if we can find those genes—and we're just beginning with this, we're still in the early days—and if we can figure out what those genes do biologically, that can tell us about the underlying root cause of the disease in human patients.

**SW:** So what kind of progress have we made on diabetes?

In the 1990s, six genes were found that contribute powerfully to risk in rare families with a so-called "Mendelian" form of type 2 diabetes. And by 2003 it was clear that a couple of biological "candidate genes" also played a role. In sum, however, these genes explained less than 1% of the risk of type 2 diabetes.

In the last five, years, using so-called genome-wide association studies, we've now identified the genomic locations, at least, of three dozen genes that influence risk of diabetes. I use the word "locations" intentionally; we haven't yet precisely defined the specific genes and mutations responsible in each case.

Using the methods to date, all we can say is that each variant identifies a place in the human genome where genetic variants track with disease. That tells us there must be a gene in that region that is somehow affecting susceptibility to the disease.



"I never imagined that a disease like diabetes was going to turn out to be explained by only five or ten genes," says David Altshuler of Harvard Medical School and the Broad Institute.

Photo by Len Rubenstein

What we have to do now as a field, among many other things, is nail down exactly which gene is responsible in each case—or could it be two or three genes?—and exactly what they do biologically. What tissues do they operate in? If we can figure this out, it will tell us a lot of new information about the underlying biology of the disease in humans.

**SW: Using diabetes as our example again, have we learned anything substantive yet about the underlying biology?**

Well, it depends on how you define "substantive." At some level I'd rather stress how much we have yet to learn, because our knowledge is still so incomplete. The fact is, as I've mentioned, it's early days in the elucidation of the genetic basis of common diseases. We can, however, point to some lessons.

One thing that I find very interesting, although it is somewhat controversial, has to do with the proportion of diabetes genes that were previously identified by other approaches. In 2005 we made a list of all the genes that anyone had suggested might play a role in type 2 diabetes. We did this by searching papers, and abstracts from diabetes-research meetings. We came up with some 600 candidate genes that were already known to play some role in glucose metabolism, or in animal or cellular models of diabetes.

We then did our genome-wide association studies. We examined thousands of patients who have diabetes, and thousands without diabetes. We compared their genomes to find places that systematically differ in relation to disease risk. And we found a couple of dozen such reproducible influences on diabetes in humans. When we asked which of the 600 genes mapped to those locations, the answer was, almost none.

**SW: What does this mean? That we don't understand diabetes, or that the genome-wide association studies don't work?**

Well, it led some people to say there's something wrong with our method—that it's telling us the wrong answer.

**SW: And you obviously don't think so?**

To be entirely honest, I have more faith in genetics than I do in what journals tell us about the basis of disease. What's interesting is that if you take other diseases, such as autoimmune disorders, and you do the same experiment—well, you find 100 different places in the genome that affect the risk of autoimmune disease. And, in that group of diseases, the answers make sense: you find genes that play a role in the immune system.

If you do the same experiment with cholesterol levels, and make a list of genes influencing cholesterol levels in humans, you find genes that make sense based on previous biology. You don't find all the known genes, of course, because not all genes have genetic variations in them. And you find lots of new genes, because there is a lot left to learn.

But, certainly, in autoimmune disease and cholesterol, there is a big overlap between the list of prior "candidate" genes and human genetics. In fact, when my colleagues took glucose levels as the phenotype, rather than type 2 diabetes, they found 15 regions in the genome that affect blood-glucose levels. Almost all of them were on our list of 600 genes.

**SW: So how are we supposed to interpret this divergence in the overlap between biological candidate genes for human genetics—yes, for glucose level and autoimmune disease and**

## cholesterol, no for diabetes?

As someone biased toward human genetics, I take this as a referendum on how well the model systems and prior research found the genes relevant to humans (at least, relevant based on inheritance). What turns out to be the case is that if you make a list of the genes that researchers say are involved in diabetes, you do find most of the genes identified so far that affect fasting glucose levels in healthy people. But you don't find many of the genes that seem to actually affect risk of type 2 diabetes the disease.

Well, that kind of makes sense if you think about it. Imagine that I looked for the genes that affect the control of day-to-day body temperature, and I examined mice that had altered body temperature. I'd find a lot of genes that influence the homeostasis of body temperature. But that might or might not tell me about why people get a fever when they have an infection. Normal homeostasis and biology might not be the same thing as pathophysiology—the process that actually leads to disease.

### **SW:** You're saying it speaks to our underlying assumptions about the disease?

Yes. It comes back to what people's assumptions are. There's been an assumption on many people's part that the biology of glucose must be the same as the biology of diabetes. And much of it is, of course; I'm not saying there's no overlap between those two. But there's a lot of ways you can think that getting diabetes may not be the same as how you regulate blood sugar, day to day, or may not be captured by studying the basic biology of cells in a dish.

### **SW:** Another criticism of this kind of work is that the genes found so far explain only a very small percentage of the genetic variation in diabetes. Is that disappointing?

Some people seem to think so, but I don't. To be honest, I was much more worried about failing completely than I was confident that we'd explain everything! I never imagined that a disease like diabetes was going to turn out to be explained by only five or ten genes.

Keep in mind that when we started doing these studies, about 1% of the variability in type 2 diabetes was explicable by known genetic factors. It's gone up in three years to 10%. If you go back and read the literature, you'll be hard-pressed to find anyone who wrote that the genetics of complex, common diseases was going to be due to only a few common variants—and you'll find others who argued that genome-wide association studies weren't going to work at all, that nothing would be found. Certainly, that was a reasonable point of view, as all previous efforts to understand the genetics of common disease had failed.

In light of this, it's clear that the hundreds of reproducible variations that have been found for dozens of diseases are a striking success. Having said this, it's also clear that the road from initial genetic mapping to understanding what we've

### **Selected Highly Cited Papers by David Altshuler and Colleagues, Published Since 1999**

(Ranked by citations)

Rank	Paper	Citations
1	S.B. Gabriel, <i>et al.</i> , "The structure of haplotype blocks in the human genome," <i>Science</i> , 296 (5576): 2225-9, 2002.	2,043

learned—let alone applying it to medicine—is long.

Moreover, it's also clear that the ultimate answer is going to be that diabetes is genetically very complicated—that there are many, many processes that could probably contribute to diabetes. And, in each such biological process, there are many genetic variations. Some of those will be common. Some of them will be rare. Some will directly affect diabetes risk, while some will affect obesity, say, and so only indirectly affect diabetes.

This is very fundamental to where we are in human genetics. We're trying to put together a holistic picture of the human body that says, look, things that are common and complex like diabetes are actually common because there's a lot of ways to get to them, and they are indeed complex. And in the end there will be multiple processes, and there will be many genetic variants and many environmental exposures, and each one of them alone will explain only a very small percentage of that variation.

Some people say this is too complex, and that we should only study simple cases where one gene is responsible, or mice we've engineered to have only one mutation at play. Those things are very important, but I find it hard to understand how we'll ever know we've understood diabetes if we turn our back on the evident complexity that exists in the patient. Especially now that we have methods that actually work!

**SW: Ultimately, how far do you think you'll get with genome-wide association studies, and what will you have learned when you get there?**

First of all, it depends on what you mean by "genome-wide association studies." That term has come to mean the first wave of studies that compared the set of common variants we had catalogued by 2005. However, when I think about it, I think of testing all the genetic variation that

2	Intl. HapMap Consortium (D. Altshuler, <i>et al.</i> ), <b>"A haplotype map of the human genome,"</b> <i>Nature</i> , 437 (7063): 1299-1320, 2005.	1,989
3	Intl. SNP Map Working Group (R. Sachidanandam, <i>et al.</i> ), <b>"A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms,"</b> <i>Nature</i> , 409 (6822): 928-33, 2001.	1,330
4	M. Cargill, <i>et al.</i> , <b>"Characterization of single-nucleotide polymorphisms in coding regions of human genes,"</b> <i>Nature Genetics</i> , 22(3): 231-8, 1999.	1,042
5	V.K. Mootha, <i>et al.</i> , <b>"PGC-1 alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes,"</b> <i>Nature Genetics</i> , 34(3): 267-73, 2003.	1,031
6	D. Altshuler, <i>et al.</i> , <b>"The common PPAR gamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes,"</b> <i>Nature Genetics</i> , 26(1): 76-80, 2000.	864

exists in the patient. Most of our effort for the last couple of years has been aimed at obtaining a more complete understanding—developing and applying next-generation sequencing and other methods to query all the genetic variation in each patient.

So, if you broaden the term genome-wide association study to mean testing all the DNA variation in each patient for its relationship to disease, well then, I think we're still at least five years away from knowing how far we can get with this approach. So, let's say five years from now we've sequenced the genomes of many, many diabetics, and not just of diabetics but of many other phenotypes that are relevant—obesity and high fasting glucose and so on—by then I think we'll have identified 100 or 200 genes that are affecting these traits in humans.

Well, that's 1% of the genome—200 genes out of 20,000. It makes sense to me that it will be some number like that. Not five genes. How could it be? My colleague Gary Ruvkun published a paper in *Nature* years ago showing that if you systematically knocked out each of the genes in a worm, and studied the fat cells, then 500 genes influenced that phenotype. If 500 genes influence fat biology in a worm, how many influence diabetes in a human? My guess is that if we knew those 200 genes and studied them, we'd find they are involved in five or ten main processes. That is, each process contains multiple genes, and there are many processes that contribute to disease. But, if we know which processes they are, we can study them to figure out how they lead to diabetes. I would see this as a big success.

I don't measure the success by whether we're up to 30% of the heritability or 70%. And I certainly don't measure the success based on whether or not I can predict in the clinic what's going to happen to any particular patient. Because even if I take identical twins, I can't predict in the clinic what's going to happen a lot of the time. And, as a physician, I never found prediction to be rewarding unless we had something to offer that really helped the patient. And those cures don't exist, and won't exist unless we better understand the root cause of the disease in our patients.

So, if we could understand the biology underlying diabetes, and if that was in textbooks and the next generation could base their research and medical practice on that foundation, well, that would seem to me like really a big deal.

**SW:** Another startling finding from these studies is that many of the gene variants found so far are in non-coding regions of the genome. How do you interpret that?

Some people certainly consider that a surprise, and many find it disappointing. That's because they're used to thinking in terms of Mendelian genetics, where if you have a very strong phenotype, it's often due to a coding mutation. And, certainly, most of biology has focused on the coding regions of genes.

7	R. Saxena, <i>et al.</i> , "Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels," <i>Science</i> , 316(5829): 1331-6, 2007.	728
8	E. Zeggini, <i>et al.</i> , "Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes," <i>Nature Genetics</i> , 40(5): 638-45, 2008. 317	317
SOURCE: Thomson Reuters <i>Web of Science</i> ®		

So, scientists are comfortable with this.

But if you think about it, you could have predicted that this would be the case if you compare the human genome to the genome of the mouse and the cow and the dog and the chimp, and you say, show me the parts of the genome that are conserved throughout evolution. One thing you find is that all the exons—the stretches of DNA that actually code for proteins—are conserved. That's been known for a long time, and that fits the model: things that are functional are conserved across evolution.

But now turn it around and ask how much of the human genome is conserved across species? The answer is about 6%. Exons make up about 1.5% of the genome. In other words, three-quarters of the DNA in your genome that matters, based on evolution, is non-coding. Now, that's not what matters to a textbook—that's what matters to evolution. And evolution works at the level of fitness and function. It says that most of the functional DNA in our genome is non-coding, and if that's right, then most of the variation in nature that influences phenotype must also be non-coding.

But because we as biologists don't yet know how to interpret non-coding DNA variations, there's a lot of disappointment and frustration that genome-wide association studies point to non-coding sections. It's not disappointing! It's why we did it—to learn things we didn't already know. And it says we need to learn how to interpret non-coding DNA, not because of genome-wide association studies, but because evolution has told us that that's where three-quarters of the functional DNA in our genomes resides.

**SW: What are the big issues that still have to be addressed in human genetics? What are the outstanding problems?**

Well, we're moving toward a world in which so many species have had their genomes sequenced, and many individuals in each species. And certainly in humans, mice, and flies, there will be lots of correlations between genotype variation and phenotype variation. So, once we've sorted through the data and figured out which relationships are durable, the big question is going to be, how do the changes in DNA affect function at the level of molecular biology, of cells, of tissues, and of whole organisms. That's one obvious question.

Another question is, how do we take a gene of unknown function and figure out what biology it's involved in? We are going to be increasingly able to say, well, here are the 100 genes that are involved in lupus, or heart disease, or cancer, but most of them we know nothing about. No papers exist to tell us about them. We don't have cellular models or animal models that tell us what's going on.

And, surprisingly enough, we don't really have good methods yet to take a gene of unknown function and figure out what it's doing. The biggest problem today with genetic studies is not that they haven't explained enough of the heritability—it's that when new genes are found, it's very hard to figure out what they do. And, as the bottleneck of gene discovery is overcome, the next bottleneck looms large.


The third issue, of course, is that even if we knew all the genes that influence risk of a disease, and what biology they're involved in—who cares, unless we can find some ways to use that information to help a patient? The geneticist's credo is that the patient leads you to the genes and the genes lead you to the biology. But the real goal is to develop an intervention that prevents or cures the disease.

I realize that in the last 13 years since I finished my medical training I could have done more direct good for patients by staying in clinical medicine than by doing this research. But I also believe that if in the end we can discover important new insights about the causes of diabetes, at some point the ledger will turn

positive. I'm a big believer in the long-term value of fundamental understanding, and I think genetics is one of the best tools we have to figure out disease. I'm excited to spend the next 20 years seeing how the story turns out!■

KEYWORDS: David Altshuler, diabetes, type 2 diabetes, genome-wide association studies, gene mapping, genome sequencing.

 PDF

[back to top](#) 

2010 : [May 2010 - Author Commentaries](#) : Harvard's David Altshuler on the Complex Genetics of Diabetes

[Science Home](#) | [About Thomson Reuters](#) | [Site Search](#)

[Copyright](#) | [Terms of Use](#) | [Privacy Policy](#)