

# Why big pharma wants to collect 2 million genomes

For the head of AstraZeneca's genomics initiative, the challenge is not just getting the sequences, but in putting them to use — wisely.

Heidi Ledford

28 September 2016



Courtesy AstraZeneca

Geneticist David Goldstein will help the company to incorporate genetic sequences into better drug-development programmes.

Five months after announcing its intentions to gather [genome sequences from 2 million people](#), pharmaceutical giant AstraZeneca has selected geneticist David Goldstein to head up the task.

In an era of ambitious genomics projects, a ten-year-long, 2-million strong initiative seems a logical next step. But the pharmaceutical industry has a spotty history with genomics. Many companies invested heavily in the approach around the turn of the century, and some analysts now blame that enthusiasm for the [spiralling research and development costs](#) — and poor return on investment — that ensued. Well-stocked databases, the industry learned, do not always translate into well-stocked drug-development pipelines.

But Goldstein, who will keep his academic position at Columbia University in New York City, believes that genomics' time has finally come. Sequencing is cheaper, allowing researchers to amass larger reams of data. Electronic medical records can tie that data to a person's clinical history. And new laboratory techniques to test the function of particular DNA sequences ([CRISPR, anyone?](#)) allow researchers to determine whether the genetic quirk they are studying has any functional relevance for a protein, a drug or a patient.

Goldstein talked to *Nature* from AstraZeneca headquarters in Cambridge, UK, about his new consulting gig and why getting his hands on 2 million genome sequences is only the beginning. This interview has been edited for length and clarity.

## Doesn't AstraZeneca already use genomics? How will this be different?

It's fair to say that AstraZeneca, like other big pharma companies, has been involved in genomics studies for years. The difference now is really, a) one of scale, and b) one of centralization and coordination. We're really going to try to bring these activities together into a centre for genomics research.

## What's the first thing you'll have to take care of in your new position?

[One thing] I would say is really building up the internal expertise to have good judgement about genomics. The literature is an

incredibly valuable source for leads for drug development, but it is a complicated source. You really have to have real expertise to follow the genomics academic literature and figure out what's in it that's relevant to your drug.

### **And to figure out what's reliable?**

A real problem with the literature is that a lot of it is unreliable. It's really striking to me how much is out there that is untrue. And that really only a relatively small proportion of the community knows that, and knows how to recognize what is likely to be true and what isn't.

### **What will your biggest challenge be?**

Probably the biggest challenge is really making sure that we can do genomics that is useful and relevant to the drug-discovery mission. Everyone now recognizes that genomics really is moving. It's not *all* hype anymore.

AstraZeneca has established a ten-year programme and it intends to keep at it and make it work. It's a challenge, we should be frank. In the pharmaceutical world, sticking to something for many years is a challenge.

### **Is ten years enough?**

A lot of the problem has been in acting like we can predict when genomic efforts will win clinically relevant insights. It's just not very predictable. That said, I am pretty comfortable with the ten-year time frame for making valuable contributions to drug development. AstraZeneca is a big company that is involved in a lot of things.

Many of the things we work on as part of the genomics effort here will not win insights that are commercially valuable to AstraZeneca. We are not banking on a single project delivering.

### **Will there be any focus on particular types of diseases?**

It's spread across anything that's of interest to AstraZeneca. Any target that you're working on — if you're trying to modulate a protein encoded by a gene, you want to know everything that genetic variation has to tell you about what that protein does.

### **Academic collaborators are going to be a big part of the initiative. Will they be publishing their work?**

This is a very outwardly facing program. My understanding is that everything that might emerge from the academic collaboration will be published in due course without exception. Collaborations do include some arrangements about timing so that you can file intellectual property materials.

I have seen a lot of companies try to use genomics in a more strictly internal way, where they don't plan to publish as they go and don't have strong academic partnerships. Invariably, the companies become almost comically bad at doing genomics. The field evolves so rapidly. If you just put up a wall and do it internally and don't publish, you get really bad at it.

### **Will we be seeing a 20-million genome project next?**

I think we'll quit calling them projects. It's clear that by one mechanism or another, a large proportion of the populations of wealthier countries will end up being sequenced. Insurers will start covering it more and more. People will be sequenced when they think about having children. There will be a lot of individuals who are simply in research studies. A higher and higher proportion of individuals will come to their pharmacist or doctor with their sequence data.

### **Some of your academic work has highlighted the [over-representation of people with European ancestry in genomic databases](#). Will the AstraZeneca initiative seek more diverse participants?**

I will personally push as hard as I can in everything I'm involved in to make sure that genomic data are generated in as inclusive a fashion as is possible.

One of the most important reasons for that is that genomics does now matter. And to the extent that it matters even at the level of the individual, [inclusivity] becomes increasingly important.

*Nature* | doi:10.1038/nature.2016.20697

1 comment

 [Subscribe to comments](#)

---

David Gurwitz · 2016-09-28 03:49 PM



Possibly David Goldstein's key point is this one: "If you just put up a wall and do it internally and don't publish, you get really bad at it." Open sharing of human genomic data collected by big pharma is essential for science and society. Incentives should be put in place for assuring it will take place. See what has been published on this topic over a decade ago: Gurwitz D, Lunshof JE, Altman RB. A call for the creation of personalized medicine databases. *Nat Rev Drug Discov.* 2006 Jan;5(1):23-6. <http://www.nature.com/nrd/journal/v5/n1/full/nrd1931.html> This call is even more timely today.

