NEWS & ANALYSIS

Nature Reviews Drug Discovery | Published online 20 May 2016; doi:10.1038/nrd.2016.95

AN AUDIENCE WITH...

Alan Shuldiner

In 2014 Regeneron launched the Regeneron Genetics Center to search for novel targets that lie buried in the human genome. Two years on they've hit their initial goal of sequencing 100,000 patients, and now plan to sequence around 100,000 more patients per year. These data have already helped the company to identify and validate dozens of potential targets, says Alan Shuldiner, co-head of the Regeneron Genetics Center. Other pharmaceutical companies have since followed suit, with AstraZeneca announcing in April that it will sequence 2 million genomes over the next 10 years. The Precision Medicine Initiative, too, could soon provide drug companies with access to genomic data on 1 million patients. Shuldiner tells **Asher Mullard** about why genomics is finally ready for pharmaceutical prime time.

Drug companies have been investing on and off in genomics since the 1980s, with little to show for it. There has been another surge of interest in recent years. What is different this time?

I've been around long enough to have experienced some of those early starts in genomics in the 1980s and 1990s, and for the most part they were indeed not all that successful. I think that the main reason for this is that the tools were just too coarse back then. We didn't know a lot about the sequence of the human genome. We couldn't do genome-wide association studies (GWAS), which are very powerful from a statistical point of view. Another factor is that the cost of sequencing has come down quite dramatically, so that we can now do these studies on a much larger scale. And that's why many of us are very optimistic that the time for human genomics in drug discovery is now.

• What lessons have you learned since launching the Regeneron Genetics Center? The GWAS era provided us with a large menu of common variants that have relatively small effect size. And these have been good to help inform biology. But it's really the rare alleles that have a large effect size. And one of the things we are learning is that these loss-of-function alleles that have the greatest opportunity to inform biology are extremely rare. This is one of the reasons why we have upped our goals in terms of the number of subjects that we ultimately plan to sequence.

A second lesson that we've learned is that it is possible to extract useful phenotype data from real-world electronic health records. The devil is in the detail, obviously, but there has been some doubt about whether electronic records can be used for research and I think we are demonstrating that they can. This is a very important finding, given what will be emerging through the Precision Medicine Initiative in the years to come.

A third lesson is that it is possible to leverage the allelic architecture of disease - from Mendelian disorders and families to larger patient populations - to identify and validate genetic findings. We now have a number of examples in which we take a genetic finding in an individual or a family with a rare Mendelian genetic disorder, find a gene that we believe is involved and causative of that rare disorder, and then look at variation in that gene in a broader patient population to find related variants that are associated with moderate forms of the same disease. We've used this approach to discover dozens of novel genes that are associated with various phenotypes, including early-onset pulmonary arterial hypertension and early-onset inflammatory bowel disease.

Some companies sequence the whole genome in their drug discovery efforts. Why are you keeping your focus on only the exome?

I have no doubt that at some point we will be able to interpret the whole genome. But right now what people do when they sequence the whole genome is then create a synthetic exome that they study. They essentially put the remaining 98% of the sequence on a hard drive or in the cloud. Maybe a few years from now they will bring it back out. But by then it will probably be cheaper to just resequence the whole genome. I just don't think we are in a position to be able to interpret regions outside the exome yet.



Are you at the point yet where you can identify clusters of variants with small effect sizes that cumulatively point to broader pathways that can be targeted?

I think this is a very interesting approach, and one that we plan on exploring. But we have not really put much time into it yet. For one thing, the statistical approaches that are needed to demonstrate that there are more disease-associated variants in one particular pathway than in another are complicated. And, secondly, I don't think we understand all the pathways well enough to really put much confidence in these findings yet.

• How do you see the Precision Medicine Initiative affecting pharmaceutical genomics projects?

We don't know exactly what the Precision Medicine Initiative is going to look like yet. But let's assume it is going to be a million-plus patient cohort in which phenotypic data are extracted from electronic health records. I expect that those public data sets as well as private data sets that various pharmaceutical and other industry partners have harnessed — are going to help us to identify many new drug targets. And because different pharmaceutical companies have different research focuses and core competencies in terms of what targets they can go after, I think there will be plenty of room for multiple pharmaceutical companies to differentially exploit these genomic discoveries.

We are also all recognizing that genomic sequencing, at the scale that we are doing it, is huge. It's a team sport, and one in which industry partners may be able to work together in a pre-competitive way.