Global genomic data-sharing effort kicks off

A coalition of researchers, funders and businesses aims to facilitate the exchange of genomic data. Broad Institute researcher David Altshuler gives *Nature* the run-down.

Ewen Callaway

06 March 2014

The number of sequenced human genomes will soon jump from the thousands to the millions. But a recently established high-profile coalition says that if scientists want to make the most of this deluge, they need to pool their data. Composed of leading researchers, funders, businesses and advocates, the Global Alliance for Genomics and Health plans to establish technical, ethical and legal and clinical guidelines to make it easier to share genomic data.

The Global Alliance was formed last year and held its first official meeting this week in London. The group a boost last week, when Google announced that it had joined the effort and created a programming interface, Google Genomics, to analyse genome data.

After the 4 March meeting, *Nature* caught up with David Altshuler, a genomicist at the Broad Institute of Massachusetts Institute of Technology and Harvard in Cambridge, Massachusetts, who is a member of the Alliance's steering committee.

Why was the Global Alliance established?
We're living at a very important moment in history, where it's becoming possible to collect large amounts of information about genome sequences in individuals, along with clinical information. By understanding the relationship between genotypes and phenotypes we can discover mechanisms of disease, we can develop diagnostics and risk predictions, we can identity new leads for drug discovery. But the other thing that's become clear is that in order to learn from the data, you need very large data sets, larger than any entity can collect on its own.

Broad Institute

David Altshuler is a genomicist at the Broad Institute of Massachusetts Institute of Technology and Harvard.

What kind of progress has this group made thus far?
We've gone from a concept to having 151 partners. If you look at the list you'll see academic medical centres and people involved in health delivery; you'll see some disease advocacy organizations; you'll see for-profit companies in life science and information science. This is a much broader array of participants than have ever been in the room before.

What came out of this week's meeting?
It was a big tent that brought together stakeholders to identify the most pressing problems the alliance would solve. For example, there is a set of file formats currently used that came out of the 1,000 Genomes Project because we needed it. We think the current generation requires not file formats but machine-readable application programming interfaces (APIs), and this group is developing — together with academics and for-profit companies — an open-source public API for genome sequencing reads and for genetic variants. That's a very concrete set of things that we think the field needs.

Is this where Google fits in?
It's important to realize that the Alliance is 151 members, of which Google is one. We think it's important to have all the best experts at the table.

Craig Venter just announced a new company that will sequence and analyse 40,000 human genomes a year, and many other significant data generators are not part of your alliance. Are you worried that genomics has already become too much of a Wild West to reign in and standardize?
It would be surprising in such a dynamic field not to see new activities. I fully expect to see millions of genomes sequenced. We all fully

expect and embrace and think it is a very good thing that there will be many more people generating data — genomic, clinic, environmental and other types. That's the time in which we live.

But there are certain kinds of human activities, such as the World Wide Web, where information being in an inter-operable framework with appropriate security and privacy has really transformed things in a very positive way. We're not in that world yet for genomics.

Why should researchers share the data they've spent a lot of money to generate?
My view is that many researchers want this. Many engaged in the clinical uptake of genomics want better information to help interpret variants for their patients, and most of all — and this is what motivated the people in this room — they want to maximize the public benefit of genomics.

You've come to England in the midst of a controversy over the collection and centralization of patient health data. Are you worried about public reaction to a global effort to share personal data and genomes?
It's really important that we frame this as "there are real benefits", and we need to be concrete about them: a child who might have a rare genetic disease who can't get a diagnosis, but who might if information could be exchanged; a patient with cancer for whom there might be a drug that would be more effective if we had some understanding of their tumour, but we don't have enough data yet to recognize the pattern. There are real benefits. And then there are real potential harms. If one takes the view of only focusing on the harms, then there are approaches to put data behind high walls so no harms can occur and no learning can occur.

What are the next steps?
We seek engagement with a broader range of countries and parts of the world. We want to engage more with disease advocacy groups and with the public. There are these specific to-dos for each of the working groups. We've tried to identify a small number of high-priority goals that we can accomplish in 2014.