GENETICS 08.16.2015

Peering into My Own Genome

— Genomic sequencing held a surprise or two, but no scares

by Michael Smith
North American Correspondent, MedPage Today

I have looked into my genome and seen ... nothing much to worry about.

I carry 15 rare genetic variants that have been associated with some form of disease, but I am heterozygous for all of them -- I have one intact copy of the same gene. So I carry, for instance, variants linked to early onset Parkinson's disease, Blau syndrome, ghosal hematodiaphyseal syndrome, and type 1A achondrogenesis.

For the record, if I am likely to be bedridden after surgery, I will make sure my doctors know about the polymorphism on (one copy of) chromosome 1 that might predispose me to blood clots. And I will be taking baby aspirin for a few days before my next long flight.

But that's about it.

All this comes from my participation in the Personal Genome Project Canada (PGPC), which is part of an international endeavor -- started at Harvard University several years back -- to obtain the complete genetic sequences of hundreds of thousands of people.

That data would be available to researchers and paired with information about the person's physical attributes, family history, and medical records with the goal of being able to understand how genes interact with environment to produce ... well, you or me.

The hope is that tying genotypes and phenotypes together will finally usher in real precision medicine, with each treatment tailored to an individual.

So far, the various PGPs -- there is also one in Austria and one in the United Kingdom -- have only a few hundred sequences done. But whole genome sequencing, as it's known, is about to go mainstream.

Within a year or so, there will be about a million whole genome sequences in various databases around the world, according to Steven Scherer, PhD, of the Hospital for Sick Children in Toronto, who heads the Canadian project.

A lot of that will be, as you might expect, aimed at specific health problems. Scherer, for instance, is well known for his work on the

genetics of autism; he and colleagues are planning to do 10,000 complete genomes to further that research, with 6,000 already completed.

That dwarfs the 30 done for the Personal Genome Project Canada and even the 400-odd completed by the Harvard group.

In turn, it is dwarfed by the 100,000 genomes that genetic pioneer Craig Venter, PhD, of the J. Craig Venter Institute in Rockville, Md., is intending to sequence next year, as part of a project to understand the roots of longevity.

But the personal genome projects underlie that work and similar projects, Scherer said, because it's not just about DNA.

There's a complicated consent process, because sequencing the genomes could have consequences to the individuals involved. The sequences themselves have to be annotated, because there are no signs in DNA that say: 'Here's something important.' Results have to be delivered to the participants in ways they can understand. The environmental information -- the medical history and so on -- has to be connected to the DNA in ways that make it useful.

"The sequence itself used to be the biggest part," Scherer said. "Now it's everything around it that's the really hard part."

An important aspect of the projects is that they offer a template for others to work in the field, according to George Church, PhD, of Harvard University, who is the sparkplug for the whole endeavor.

"Part of it is to demonstrate to other groups that it is valuable and relatively easy to do what we are doing," he said. "And the other part of it is to actually do it and generate large data sets" that can be useful for research.

The key word for Church is causality. Previous genetic research has, very largely, seen correlations between genes and illnesses. But work like the PGP has the potential to show when gene variants cause disease and when they are "innocent bystanders," he said.

For instance, he and colleagues reported last year in *Nature Medicine* on a series of experiments that pinned down the genetic cause of a form of mitochondrial myopathy known as Barth syndrome, using cell lines derived from PGP participants as controls.

Indeed, using the participants' genomes as controls in genetic research is a major benefit of the projects, Church said, because the phenotype data is available. "We're the largest control study, if you will," he said.

Another benefit is the ability to use the information to hone technical abilities. Yearly contests -- dubbed Critical Assessment of Genomic Interpretation -- withhold certain phenotypic data from researchers who then are asked to predict that information from what they know of the genes. Then the hidden data is unmasked and the various interpretation methods can be judged.

The availability of that phenotypic data is an opportunity to make sense of the genes, but it is also a complication.

A key issue is how good that information is, noted ethicist Hank Greeley, JD, of Stanford University in Stanford, Calif., who's noted for his examination of issues surrounding genetics.

For instance, family history of disease is part of that. Did your father die of heart disease or cancer? What about your grandparents?

"Most of us don't have a clue about our family history," Greeley said, "and very often that clue is wrong."

Even the personal stuff -- medications, childhood diseases, vaccinations -- is hard to pin down. I was once tested for allergies; there were some, but I'm not sure now what they were and I can't remember the name of the doctor who did the testing. It's probably in my medical records somewhere, but likely to be in some dusty file folder rather than easy to find and use.

"Maybe George (Church) should start a personal phenome project, if that's a word," Greeley joked.

Perhaps more to the point, even when the data is available, it has to be put in some useful form, which takes time and costs money. That's one of the chokepoints for the Canadian project, according to Michael Szego, PhD, of the Centre for Clinical Ethics in Toronto, who is both an investigator and a participant in PGPC.

He and Jill Davies, MSc, of the Medcan Clinic in Toronto (she's also a participant as well as a genetic counselor) spent an hour with me to talk over the first results, part of the genetic counseling that is a mandatory part of the project.

At the moment, Szego said, the investigators are still working on ways to get the personal history in a useful and easily modified electronic form. Right now, what they have has to be input by hand.

In Ontario, there might be a relatively simple way to get some of that data. The largest Canadian province has a single-payer healthcare system, so broad information on diagnoses and prescriptions is available for millions of people and much of that is already in electronic form.

The investigators are exploring the possibility of using that data to flesh out the personal histories of participants, Scherer said.

In the meantime, they're concentrating on getting the actual sequence data reported to participants in advance of writing a paper on the first 30. For reference, I am Canuck-19; Davies is Canuck-1.

The first slice of data includes the rare variants -- there are some 4,000 known -- which are a bit easier to annotate and also are likely to have large effects. Later reports will cover the more common disease-associated variants and those that might be expected to affect how I metabolize drugs.

Davies said that geneticists have long anticipated that everyone will carry at least some rare variants but the estimates have been slowly rising. Based on their current data, the range is from 11 to 29.

Mine include an A for T substitution on chromosome 1 that is associated with thrombophilia owing to activated protein C deficiency and to factor V deficiency. There's also an A for G switch on chromosome 2 associated with primary pulmonary hypertension and pulmonary occlusive disease.

I also carry a couple of variants that can be tested for before conception and that might be relevant to my son should he have children -- one linked to type 1a achondrogenesis and one to type 1c congenital disorder of glycosylation.

Aside from that, nothing to disturb my sleep. And that seems to be the case for most people.

Szego said one of the things that appears to worry participants most -- including Szego himself -- is that the analysis might turn up some sort of "crystal ball" that predicts when or from what they will die. "The take-home message for me is that the information you get back is not all that scary," he said.

There have been some surprises -- the number of rare variants is one, but also the proportion of people carrying a variant that causes cystic fibrosis, which is higher than expected at about 30%, Davies said.

Whether that's chance or meaningful remains an open question, but it illustrates an important principle: you don't know what you've got until you go and look.

Greeley said there's an important distinction between whole genome sequencing and the commercially available services that offer to tell you about your ancestry or (until recently) your health.

The commercial services rely on maps of known single nucleotide polymorphisms -- single base changes in DNA -- that have been associated with risk of disease or other traits. The advantage, Greeley said, is that they are cheap and easy to use; the disadvantage is that they are only rarely useful.

A whole genome is a stark contrast -- about 6.5 billion data points instead of perhaps a million.

"For large chunks of the sequence the science is completely uncertain," he said, "but there are thousands and thousands of areas where the science is really quite good and the power is going to be strong."

Geneticists are fond of saying that we're mostly alike, tossing around figures like 99% of our genes. But there's a lot of unknown territory, Greeley said, and whole genome sequencing will give us "a better sense of what the 7.3 billion of us look like genomically."

"Just how much variability is there? How many weird things does each of us have?"

On the other hand, even if there are a million whole genomes sequenced by next year, that's still a drop in the bucket. And it's not clear, Greeley said, how representative those genomes will be.

The PGP participants are volunteers -- some of whom are paying their own freight. Almost by definition, they will be relatively affluent and educated. Do they differ from the general population in ways that matter genomically?

Of course, without the non-participants it's hard to tell, but Church said a central goal is to understand the role genes play in disease, and most of the important diseases are shared by everybody.

The issue of representative samples is a bit of "red herring," he said, and in any case can't easily be addressed.

Another issue is the risks associated with what Scherer called the "full monty" -- putting all your information out there.

The risks, while not huge, range from discrimination on the basis of genetics to simple loss of privacy, Greeley said.

While my name will not be attached to my data, I'm pretty much making the link right now for anyone who cares. And it will almost always be possible for people to work back from the raw data to get to the person, Church said.

"Wikileaks got hold of data that had a whole lot more money protecting it than we have," he said.

Is it fair, then, to ask members of minority groups -- many already mistrustful of medical research -- to put themselves on the front lines?

Better, Church says, to wait until he and colleagues can show the work can be done safely.

"It's all about not having too many unintended negative consequences for the cohort," he said, "and so far I think that the record is perfect."

In the long run, "we want 7 billion people", he said, "but in order to get that, the first billion will have to have had a pleasant journey and to get to that the first thousand will have to have had a good journey."

What about the short run? Most people, Church said, have the view that gene-based medicine is "10 years off and always will be" but that's not true.

Already, pre-conception testing is available for gene variants that -- combined -- pose risks to the health of a fetus or child. Prenatal testing can identify several dozen genetic conditions. Post-natal gene testing is widely used to pick up such treatable illnesses as phenylketonuria. And, of course, adults are tested if a physician suspects a genetic condition.

"This is not in the future," he said. "This is today."