Personal Genome Service Providers
George Church, Ph.D.

MS. AU: Our next speaker is George Church. Dr. Church is founder of Knome and the director of the Personal Genome Project. Dr. Church is a professor of genetics and the director of the Center for Computational Genetics at Harvard Medical School. He has been at the forefront of DNA sequencing technologies, including the development of the first genomic sequencing method, in collaboration with his dissertation advisor Walter Gilbert in 1984. His research continues to foster high throughput technologies for molecular biology.

Welcome, Dr. Church.

DR. CHURCH: Thank you. I want to thank all of these government agencies, as well as companies that we work with very closely, and also disclose possible influences that we have had.

From the discussion so far, I wanted to say that when we say "personal empowerment requires prior validation," which was a conversation that came up earlier, one of my take-homes here is that a lot of what we are doing in the Personal Genome Project and at Knome and to some extent in my advisory role at 23andMe is research. It is empowering people to do research rather than empowering them to influence their medicine right at the moment.

I think that is incredibly important in the sense that there is a very strong attitude among many people, certainly not everybody, where we want to learn about the world at some risk to ourselves. We will explore the planet and risky areas of the planet individually. We will look at investments. We will look at the Internet. These are all risky environments in different ways that affect your quality of life, and they are probabilistic decisions that aren't necessarily any more complicated or less complicated than genetics, and they are moving targets.

This is the other end of the spectrum, I think, from the big four or five we are talking about here today, where we have various ways that people are doing their own genetics. Many people know about the genographic project, which is mainly ancestry. But in addition, Hugh Rienhoff was on the cover of Nature for trying to understand his daughter's illness, and he is doing this basically in his home.

But what is happening is that there are people that, rather than hiding from their personal genomics, for which there is no cure, they are embracing it, they are becoming activists, and they are saying we can do something for our family by doing research on our ourselves and people like this, ranging from my colleague Doug Melton, whose family has diabetes, to Hollywood blockbusters about lipid biochemistry that Nick Nolte became, representing Augusto Adone and his son. Nancy Wesler. We have already heard about Michael Fox and so forth.

Next slide. So, in this context of course, all of those and some of the people we have heard about today are saying privacy is not their top concern. But even when it is, there are many ways that privacy is compromised when you put things anywhere other than in a vault. You can have a laptop theft where 26 million veterans' data got out. You can get a case where a 15-year-old person wanted to know his anonymous sperm donor father and took a cheek swab and did a genealogy which narrowed it down to an individual that he found and confirmed was his father.

There are many, many ways that data get out, and it is unrealistic to overpromise. We certainly want to try to make it as private as possible.

When we talk about the research landscape here, we have standard research on the far right here, where we have open access as long as there is no trait data, such as the 1,000 Genome Project HapMap, and we have various types of approaches that are increasingly returning data back. We have already heard about the Reveal study with my close colleague and so forth.

But typically, the data are kept de-identified or safe from the individuals that donated it.

At the other end, you have it only available to the individuals with marginal ways of getting it into the public domain. Then we are exploring ways in the middle here where we can make it both publicly available, connecting DNA and traits, and yet not overpromise on privacy, particularly recruiting people who have passed an exam with 100 percent on the questions. That is Item No. 4 here.

One of the goals of this project, which has had IRB approval since 2005, is to really try to get ahead of the curve. Of course, the curve has well caught up with us at this point. But the idea is to bring technology to bring down the cost of not just the coding sequence but the regulatory data -- by "regulatory," now we are talking about RNA regulation, not the kind of regulations we are talking about here -- that Teri mentioned. Maybe 6 percent is coding in GWAS studies, but closer to 90 percent in the rarer diseases that populate online Mendelian inheritance.

We want full subject participation, which is not unusual in this context. We have multiple samples to make sure we have the identity. We have open access. We have a trait questionnaire. We have stem cell RNA I will mention in a moment, and we have now IRB approval to scale up to 100,000 individuals.

We are focusing on sequencing, and I think Ryan Phelan will talk in a little while. This is one of many tests that constitute the best of the genetic diagnostics, but this illustrates that, in contrast to the big three that are producing chip-based analyses, when you really want to go into detail on a test like BRCA1 and BRCA2, you are typically talking about sequencing, not chips. That is not necessary forever, but that is typically the practice now.

DNA Direct has something where you actually will see causative alleles which change the reading frame for tumor-suppressor like BRCA1. This has to be very carefully interpreted at the DNA sequence level because the consequences are very serious even the preventative sense, where people will do a bilateral mastectomy if they trust the interpretation of the data.

Now, we have alluded to but haven't really touched on yet this next generation sequencing, which has changed our perspectives of what is possible tremendously, possibly by a factor of 1,000 drop in price. You will see this in a later slide.

There are at least two classes of chemistry. We are trying to produce a platform that will support multiple versions of each of these classes of chemistry based on DNA polymerase or ligase. Rob Mitra helped worked on the polymerase version in 1999, and a couple of companies, Illumina and Intelligent Biosystems, use this.

The same thing can be done for ligase. We are using fluorescently colored monomers or multimers which can be discriminated by these enzymes. This is something that Jay Shendure and Greg Porreca developed.

Now, in addition to those commercial instruments, we have an instrument which is kind of at the fringe in between academic and commercial which we call a polonator which is intended to be an

unusual model and which is completely open-source hardware, software, wetware. We are just opening it up to the community so they feel empowered to change any part, and it is intended to be easily modular.

Now, maybe only 5 percent of the research community will want to change it, but that 5 percent will greatly aid the other 95 percent.

So this is $155,000, which is about four times less than our previous contribution to the applied biosystems solid device, which is $600,000, and similar to even lowering in prices of the reagents and reagent use.

Next slide. What does that kind of technology result in. It results in plummeting costs which are faster than the already very rapid Moore's Law for Computing. Moore's Law for Computing is about a 24-month improvement in service for a given price point for computers, and this is more like a six- to 12-month doubling time, going from a fairly low estimate cost of $100,000 per million base pairs towards the end of the Human Genome Project, plummeting -- this is a logarithmic plot, as you can see -- down so that we are getting close to $1,000 a genome very, very soon. Multiple technologies are going along this pathway at slightly different points.

We can see how this plays out in the consumer market here in the genographic project, which is arguably the most popular out. Two hundred thousand people have done it. It has a very high price tag per base pair or per bit of information, but still people are very curious about their ancestry and they are willing to pay a lot, $99 for 12 bits of information.

DNA Direct has very high quality and medically actionable information, mostly done with DNA sequencing technology which historically has been expensive but has been plummeting, according to this plot here.

We are already familiar with these. Then, the Personal Genome Project has a cluster of four points here because we are not just doing genomics, we are doing coding regions, regulatory, microbiomics, and so forth. But they all have roughly similar price per mega base pair.

Then Knome is the only company that really offers full genome sequencing. It is currently $350,000 and likely to go down on that same curve very soon.

So it is not just genomes, as David mentioned in his talk earlier. There are environmental components which are very important. When you say "personal genomics," you should be thinking about the regulatory elements which might be less expensive and more interpretable if analyzed at the RNA level.

Some of the environmental components can be measured either by measuring the microbiological components, allergens, microbes, viruses, or their impact on the immune system, which, rather than being a spike of microorganisms, it might be clear from the system will be a longer term persistence leading to traits. So we don't just go from genome to traits. We go through this regulatory and environmental filter.

Next slide. In order to get at some of these RNA regulatory interactions with the environment, in the Personal Genome Project we have included multiple cell types from adults whether they are healthy or diseased, and we don't do it by assaying all of the different tissues from the PGP volunteers. Even though they are really gung ho, they really draw the line at a thousand biopsies.

[Laughter.]

DR. CHURCH: Instead we take one biopsy from the skin from which we have established stem cell lines, and we are making these available to the community, from which you can reprogram to almost any tissue you want. This is of course a very fast-moving target as well.

We want to be able to do biology on these individuals as well as inherited germ line genomics. At the extreme of that is looking at the microbiological components in general, viruses and bacteria, and not necessarily the whole genome but selected parts. Just like we might want to do different assays for the inherited genome that go beyond SNPs, we might want to go beyond SNPs for microorganisms.

Here we have studied the resistance settlement to 18 different major classes of antibiotics over 140-some days in some of the Personal Genome Project volunteers. A big solid blue means that each of these isolates along the X-axis is resistant to multiple antibiotics along the Y-axis. This was a surprising result and was actually an outlier both for this individual and for other individuals done on the same day. But this is the kind of background information that you could do by highly targeted analysis of microbiomics.

MS. AU: Dr. Church, could you just wrap it up in about 15 seconds?

DR. CHURCH: The next slide is the last slide.

The questions I wanted to add to the questions that were given are, how do we fund these association studies in education. Is there a role for direct-to-consumer companies. How do we celebrate and incentivize the best protocols, not just scare the worst and reinforce the oldest.

What about do-it-yourself genetics; is that going to be completely outside the direct-to-consumer we have been talking about. There is this risk of gene information. There are many other things that people do that are probabilistic that I mentioned. There is the risk of not educating. I don't think anybody is seriously considering that. What kind of model do we have. Is insurance an interesting model, where healthy people like David still have a finite risk. Thank you.

MS. AU: Thank you.