Clinical Chemistry

Q&A: Privacy in Direct-to-Consumer Genetic Testing

Article: Jason Y. Park, et al.
Q&A: Privacy in Direct-to-Consumer Genetic Testing
Clin Chem 2019;65:612-17.
http://clinchem.aaccjnls.org/content/65/5/612

Guests: Dr. Jason Park is Associate Professor at the University of Texas Southwestern Medical Center. He is also the Medical Director of the Advanced Diagnostics Laboratory at the Children's Medical Center in Dallas, Texas.

Bob Barrett: This is a podcast from *Clinical Chemistry,* sponsored by the Department of Laboratory Medicine at Boston Children's Hospital. I am Bob Barrett.

For about the past 15 years, the public has become aware that genetic testing can provide medical diagnosis and inform specific medical therapies. In addition to being used for medical purposes, genetics has also become a popular recreational tool for genealogy and general wellness. The number of customers for direct-to-consumer genetic testing has now grown to greater than 12 million and continues to rise.

In the May 2019 issue of *Clinical Chemistry*, a Q&A feature titled "Privacy in Direct-to-Consumer Genetic Testing" asked four experts with different roles in this field to discuss the current state of genomics and privacy. The moderator for this Q&A feature was Dr. Jason Park, Associate Professor at the University of Texas Southwestern Medical Center in the Department of Pathology, and the Eugene McDermott Center for Human Growth & Development. He is also the Medical Director of the Advanced Diagnostics Laboratory at the Children's Medical Center in Dallas, Texas and he is our guest in this podcast. So, Dr. Park, overall, what are the risks to privacy from genomic data?

Dr. Jason Park: Genomic data contains the personal privacy risk of traditional medical information such as diagnosis and test results. In addition, genomic data creates unique vulnerabilities since it contains information which is relevant to biologic family members. For example, if I have a positive cancer genetic test result, then my biologic family members have a substantially increased risk of having the same positive cancer genetic test result. In the genomic era, we are entering a period where families have shared responsibility for confidentiality.

Bob Barrett: Is such genomic data protected by current U.S. law?

Jason Park: A patient's genomic data is protected under the U.S. HIPAA Privacy Rule, but only when the data is considered individually identifiable. Once the genomic data is stripped

of the traditional identifiers such as the patient's name, date of birth and discharge date, the data is then considered de-identified and not subject to the HIPAA Privacy Rule. What is unique about genomic data is that it can be re-identified back to the patient. Although this is difficult, several studies have shown that using a combination of public databases, genomic data which has been de-identified, can be re-identified back to an individual.

Bob Barrett: So, are there legal protections for patients with positive genomic test results?

Jason Park: Yes, and the most important U.S. law is GINA, the Genetic Information Nondiscrimination Act of 2008. This law prohibits discrimination based on genetic information for employment and health insurance. In addition, GINA extends protection to biologically related family members. Importantly, GINA does not provide protection for life insurance, disability insurance, or long-term care insurance. For example, a life insurance company could use an individual's genomic test result to decide insurance premiums or general insurability. Again, the risk applies not only to the individual with the original test result, but all biologic relatives.

Bob Barrett: Are there efforts to create centralized human genomic databases?

Jason Park: I'm not aware of coordinated efforts by insurance companies to use genomic data to decide insurance premiums. However, the U.S. has had a long history of using genetic data for law enforcement. There are over 15 million DNA profiles in government databases in the form of STR profiles. STR profiles are the type of DNA data we hear of most commonly in criminal cases and identification of human remains.

In addition to this traditional type of law enforcement DNA database, there are multiple research efforts to do genomic sequencing of individuals and retain this data in centralized databases. One project generating information is the NIH All of Us Research Program which will characterize health and genomics of up to one million Americans. International efforts for genomic databases include the UK 100,000 Genomes Project which has completed the profiles of 85,000 individuals. There is also a planned Chinese Precision Medicine initiative which seeks to sequence 100 million individuals by 2030.

Bob Barrett: What are some of the uses of genomic databases?

Jason Park: The main historic use of genomic databases, such as STR profiles, has been in law enforcement identification of

individuals. This activity of identification has expanded to international intelligence activities with the gathering of the genomic data of enemy combatants in war theaters. The past 20 years have seen genomic data used to identify new pharmaceutical targets for diseases such as cancer. What is clear is that genomic data in any form is financially useful information. As an example, in 2018, the pharmaceutical company, GlaxoSmithKline invested $300 million with the consumer genetics company, 23andMe, to gain exclusive rights to the genetic data of 23andMe customers.

Bob Barrett: What are some of the new developments in genomic privacy and consumer genetics?

Jason Park: The most remarkable recent development has been the use of consumer genetic databases for criminal law enforcement. Traditionally, only one type of genetic data, the STR profile, was used for criminal law enforcement. However, over the past five years, SNP profiles, which are the commonly used genetic test in consumer tests have been repurposed to solve criminal cases.

Bob Barrett: So, how are SNP data used? Is it similar to Short Tandem Repeat profiles?

Jason Park: SNP data is different and not interchangeable with the STR profiles. However, SNP data can be as powerful as STR profile data in identifying individuals. As a prominent example, last year, a suspect was identified in the 40-year-old cold case of the so-called Golden State killer. Crime scene DNA had previously been run through traditional STR databases and had been negative. The investigators of that crime recently took the novel approach of using crime scene DNA to generate a SNP profile compatible with consumer genetic databases.

They entered this profile into a recreational genealogy database to identify partial matches. They found over a dozen partial matches that are best described as third cousins to the killer. The investigators then used geneologic data such as newspaper clippings, birth announcements, marriage records, to then reconstruct the history of the family tree from these third cousins to identify a common ancestor. Once they had the common ancestor, they then attempted to identify all living descendants. This massive project created a family tree of 1,000 individuals dating back to the early 1800s.

Based on this family tree, they then applied their existing criminal profile of a white male, 60 to 75 years old with a military interest or background. They identified two suspects, one of which they confirmed using traditional STR profiling and subsequently arrested. Since this SNP

consumer genetic search technique debuted in 2018, over 30 suspects have been positively identified from previously so-called cold cases.

Bob Barrett: Well, finally Dr. Park, what do you recommend for maintaining genomic privacy?

Jason Park: A simple recommendation would be to not participate in consumer genomics. However, this becomes unrealistic because it requires a coordinated effort of all biologic relatives not participating. This might be feasible for your core family, but becomes impossible when you think of all the second and third cousins you do and do not know of. In addition, there have been recent population-based family trees that have been created which contain greater than 10 million individuals in a single family tree going back hundreds of years. Right now, 60% of Americans of European ancestry are believed to already be partially identifiable using SNP data in consumer genetic databases. I personally believe that the future of genomic privacy will not be in the protection of the data; rather, genomic data protections will have to be implemented in the control of how the data is used.

Bob Barrett: That was Dr. Jason Park, from the University of Texas Southwestern Medical Center and the Children's Medical Center in Dallas, Texas. He has been our guest in this podcast from *Clinical Chemistry* on privacy in direct-to-consumer genetic testing. That Q&A feature appears in the May 2019 issue of *Clinical Chemistry*. I'm Bob Barrett. Thanks for listening!