Biographical Ancestry Prediction Based on DNA (SNPs) for Investigative Leads

Frederick Bieber: What Jack and Vince have been talking about is very cutting-edge and new, and it's hard to predict when you will see some of these laboratory possibilities become realities in cases that you're all working on, either in your research labs or in your own courtrooms.

And what I'm going to talk about today is sort of a follow up on that to tell you some of the work that we have done on the concept of extending the reach of DNA technology beyond perfect matches between crime scene evidence and individuals who may be known, either known suspects in the community or those who are known because their profiles are stored in one of several databases. Like offender databases, for example.

Before I do that, I want to follow up on something Vince talked about in great detail, and that is biogeographic ancestry prediction or racial or ethnic identification using either STRs or SNPs, the single nucleotide polymorphisms or SNP based testing.

There's actually a number of companies that are offering this kind of SNP assay for biogeographic ancestry prediction, including one in Florida. And I'm not make an advert for *DNA print genomics* in Sarasota, but this triangle plot is an example of the output from that laboratory If you were to send your own cheek swab or that from a DNA sample from crime scene evidence for example to have analyzed.

And in these so-called triangle plots which represent the output of maximum likelihood function analysis one could put one of several ethnic or racial groups that had been previously defined at the points of these triangles. So you could look at three different populations at once. And you could then assign the DNA sample that was obtained somewhere in that triangle plot.

And I'll just give you one example and then give you two examples of how I've used it myself for different reasons.

This red dot here would indicate that that individual DNA sample was principally from European origin, that there was some Native American admixture and that there was some African admixture.

So the position of the red dot in the so-called triangle plots gives you a somewhat of a precise estimate of the amount of mixing or genetic admixture that probably exists in most, many human populations today.

And as Vince said, the so-called African-American population in the United States varies between 10 and 30 or 40 percent admixture between European Caucasians and indigenous people from Africa.

And Jack Ballantyne a couple of years ago at a meeting told us that he had sent his own DNA to one of these companies and that he was relieved to… that he really was a Scotsman. I don't think you need a SNP analysis to tell, from his voice print would tell you that.

But I got interested in that aspect of genealogy. And by the way, the principal use to date of these kinds of ancestry predictions is just that people interested in their own genealogy arrived from the same McKenzie clan in Inverness that my neighbor is with a Scottish name, for example.

And my wife Jane has family lore that her great-grandmother used to store meat under the mattress in New Hampshire.

And I said, "That's a native Indian practice in the Massachusetts' and the Pemigewassett Indians. And I wonder if your father's part Native American."

And she said, "Well yes, there is a story, old family lore that his great-grandmother was a so-called squaw or Indian."

And I said, "Well do you think he'd be interested, why don't you hire a genealogist and find her roots?"

So she and her sister hired a genealogist. And it was a dead-end because the records from the church and burial plots and so on had been burned years before.

I said, "Okay, next time you're down in DC, swab his cheek."

And so she did and sent it down to Sarasota. And it came back and his red dot on this triangle plot was right about here. Closer to the European than the Native American, but predicted in their maximum likelihood analysis that he was about 17% Native American Indian.

And she was very excited as was her dad to realize this family lore was actually true. I was very disappointed because if it were just a few more points, we could've opened a casino. And I could have a second Porsche 911 instead of only one.

So I got interested in this and I thought "you know, we did Jane's family, I really should do mine."

So I sent my own swab down. And they were right on target. It came back 100% Libertarian.

And so I'm a great believer in this kind of profiling that can predict the outcome of elections, I guess by how many of us decide to vote.

But I want to transition my formal comments now to talk about the ways in which geneticists and forensic scientists in crime labs and in research labs use indirect DNA testing.

Most of the time, you're familiar with taking crime scene evidence and comparing it to known individuals. So there's a direct comparison of evidence to a known individual. However, in contrast to those direct analyses many times we need to make what we call indirect genetic comparisons. We don't necessarily have a known source. We're wondering, is this man, the named man, John Paul, the real father of this baby? And so we would get Robin's DNA and the child's DNA and then test John. And sure enough, he's the father. Proud father.

And as you know, these paternity trios are tested regularly in civil cases for child custody and child support issues.

Also in cases of statutory rape, criminal incest, immigration disputes and so on. Paternity testing is this form of indirect testing. You don't really know if all members are really the true source of the biological evidence in that baby, for example.

In a larger category and one that's dear to some of our hearts in this room, including Jack and I who have worked very hard with the New York City medical examiner's office and the Katrina team in Louisiana, when there are mass disasters like in the aftermath of 9/11 or hurricanes Katrina and Rita, the victim identification and the human remains that are to be returned to the families often require indirect testing. We don't have a toothbrush or a known biological sample, a biopsy from a hospital for example. And we need family members to volunteer and come forward. We put the pedigrees together using kinship algorithms that are well known to mathematicians and population geneticists.

And so these indirect methods are exceedingly important in humanitarian human remains identification.

And what I'm going to do is show you how these same technologies that are widely used in those areas could be used in perhaps a controversial way in forensic investigations. So keep in mind, after these mass disasters if we have a toothbrush or a known biopsy we could make a direct comparison to that bone or piece of tissue you found at 9/11 or at the Pentagon or in Shanksville, Pennsylvania or in Louisiana and compare it to the toothbrush of John Smith and find a match.

If John Smith's toothbrush isn't available, maybe his mother or children are available or his siblings, and they'll come forward and we can do the indirect methods.

And we're simply making likelihood ratios of the probability that that child belongs to that couple, the woman and the named man, compared in the denominator of these calculations to the alternate possibility that it belongs to a randomly selected person.

And the formulas get much more complicated in bigger families, but we're basically making indices of paternity index in the simple case of a trio or a kinship index when it's more complicated.

The higher that ratio, the greater the numerator in relation to the denominator, the bigger the weight of the evidence is that favors that stated genetic relationship.

Let me give you an example of how paternity testing could work in this way to solve a criminal case.

You know about the terrible Wichita, Kansas series of homicides and rapes. The so called BTK; bind, torture and kill; serial investigation that was going on for many years. And this person, the perpetrator taunted the police with e-mails and secret letters and they answered him in coded messages in newspapers.

And they eventually suspected this individual, Dennis Rader. And they didn't want to tip him off, so they got his daughter's biopsy without her knowledge or consent through a

court order in the night. She had lived in another state by that time and they simply asked, could the owner of this medical biopsy; most women over 18 are offered pap smears every year and those slides could be obtained from hospitals under court order: could the owner of that Pap smear DNA belong to the biological child of the semen found at these crimes scenes? Simply doing a paternity test without the mother. So this would be called a motherless case in paternity testing lingo.

And the answer to that question in this case was, yes. And based on that and other evidence that the Kansas Bureau of investigation and the FBI developed he was considered a very likely suspect.

As you know, he was approached getting out of his car at his home and later confessed. So even the simple paternity testing, this indirect testing, could be used very effectively in certain targeted situations.

Just to not leave the 9/11 and Katrina issue hanging, I just thought you should see the data that Jack and I and others published in ’05. Of the victims identifications as of 9/11 of ‘05 there were about just shy of 1600 out of 2750 reported missing.

And 25% of all those identifications, of the DNA based identifications, involved indirect testing. I thought that would be interesting to you. That they weren't all direct testing. Fully a quarter of them.

With regard to Katrina for example; and each of these disasters is slightly different and you can read about this in an article coming out fairly soon; almost all of the 146 cases that required DNA for identification were indirect in Katrina. Because as you certainly know, most of the belongings were swept away, including hospital records from those unfortunate victims of those two hurricanes a year ago.

So in some instances most of the work that we do collectively as DNA scientists has to be done using family members who voluntarily come forward and provide a sample to assist in these humanitarian investigations.

I'd like to now shift for the remainder of my comments to how these indirect methods could be used in sort of traditional forensic investigations where you do not have a suspect and you may then go search against one of several databases that may be available to you.

And I'd like you to think with me for a moment on this slide. Just dwell on it until it's intuitively obvious that in mass disasters as we just talked about, on the left, we try to identify the human remains indirectly by the volunteer relatives that come forward to provide a sample.

In the same way, the same analogous way, we could potentially identify key suspects in crime investigations by getting their potential relatives in CODIS databases.

And the major difference in this thought experiment is that these relatives, these potential relatives who are in CODIS didn't come forward voluntarily. They provided a sample under the statutes that exist in all of our 50 states and the federal statute that covers the federal territories and the military.

And in thinking about this I realized, gee we had some data from a few years ago that we published or presented at a American Academy meeting where we looked; “we” meaning Carl Ladd and Michael Burke from the Connecticut State police crime lab and I; looked at a group of sibs, full sibs with the same two parents, about a hundred of those sib pairs and compared the STR profiles using the thirteen CODIS core loci test, cofiler and profiler and unrelated controls. About a hundred pairs of each.

And you can see the overlapping distributions. And you would expect that those who have common ancestors, namely siblings, would share more of their profile in common than would unrelated people.

And this is a summary slide from that poster presentation of six years ago, where we found as many as nine … some sibs had a nine locis match. But on the average they were… shared identity at four loci, compared to only one in unrelated individuals.

And the aleal sharing was of interest to us. That looking at the whole thirteen locis profile; and you could have twenty-six aleals at thirteen loci; on average sibs shared 16 aleals in common across the whole profile.

Whereas unrelated individuals on average shared eight. And there was some overlap. So I filed that away for future reference and began thinking, tried to think a little outside the box about the family issue. Christine Tomsey from the Connecticut State police called me one day shortly after that poster and said, "We beat you, Fred. We have a ten locis match in people in our Pennsylvania database."

And I said, "I'm sure those parents are related."

She said, "Well why would you say that?"

And I said, "Because I'm a geneticist, and if…. it's certainly possible that two unrelated individuals could be identical at ten or eleven or twelve loci, but it would be exceedingly rare. But it wouldn't be so rare if they were consanguineous or related by some other way."