Transcript of Dr. Mark Perlin's talk on "Scientific combination of DNA evidence: A handgun mixture in eight parts" delivered on 8 September 2010 in Sydney, Australia at the Twentieth International Symposium on the Forensic Sciences of the Australian and New Zealand Forensic Science Society.

Dr. Perlin: There will be no equations of any kind. Everything is done with pictures and hand waving. Peter Gill has brought up the concept that maybe attention could be paid to the interpretation of challenging DNA evidence. John Buckleton has suggested that continuous DNA, as opposed to thresholds, would be interesting. So, this is an exercise in seeing how that works.

(Next Slide)

The case that was provided to Cellmark was a handgun. On that handgun, swabbings were taken from four different locations

(Next Slide)

They were taken from the base of the gun, from the back, from the trigger, and from the top of the gun.

(Next Slide)

When they got the swabs to the lab, amplifications of each swabbing were done in duplicate.

(Next Slide)

I am going to walk through this diagram and use these pictures as labels. The reason that they are called 3, 4, 5, 6 is because that is the last number of the case item ID throughout. So, for consistency, those are the numbers, and the two colors, green and blue, just show the two amplifications of each swab. There were eight amplifications done all together. This is what the data looks like. All of the pictures that we are seeing are taken from the TrueAllele VUIer interface. Here, we have two amplifications for the base of the gun. Here is the SGMplus profile that was done. Similarly, for swabbing 4, we have two amplifications. All together there are eight amplifications.

(Next Slide)

It is good to get a close-up. So, we are going to zoom in on locus D18, and throughout the talk, we are going to focus on this because it is nice clean mixture data. What do we see in the data? For the base of the gun, we see two EPGs that look roughly similar to on another, but there is some difference in the peak variation. When we move to a different template from a different swabbing, we again see two amplifications that are sort of similar to on another, but they are

somewhat different from the base of the gun (template 3). As we continue down, we are seeing that the duplicate variations, like for swabbing 5, are capturing some of the peak variation, but each template pair is different from every other pair because they are four different templates. This will become important in the interpretation as we go.

(Next Slide)

Human review was done. The three solid squares show when a full allele pair was produced. For the two loci where there are dotted lines, one allele and a wildcard “F”, as they do in England, was produced. A statistic was computed using inclusion, which is a threshold-based method. The CPI, which as we know is a likelihood ratio just with a particularly weak likelihood function, was 17,000. The likelihood ratio is a great way of expressing ratios of probabilities. It is also interesting that the logarithm of a likelihood ratio is a standard measure of information in much of science, statistics, and computer science. We can think of the logarithm as the order of magnitude or of the number of zeros after the first digit. We are going to use that measure of information. Whenever we see log(LR), that is the information content. Since this is about 10,000, it has 4 zeros, and the log is a 4 with something after it. That is the order of magnitude. About 4 information units were extracted from human review. That was good enough for government work. As the case went on, the individual was convicted and so on. They were able to get a lot from this major contributor, such as a

database hit and so on.

(Next Slide)

Let us take a look at the data in a different way. I would like to introduce, for those who have not seen it before, the notion of a quantitative likelihood function where we review the data continuously instead of having thresholds. This is taken from the Explain interface of the TrueAllele system. Any good system should visually tell us what it does. In the two orange bars, we are looking at the alleles from the major contributor, [15,19], and that is there at an 85% amount. In the blue bar at allele 13, suppose that this is a homozygote for a 15% minor contributor. All I have done is overlain the amount of DNA that might be at those alleles from the different contributors (orange from the major contributor and blue from the minor contributor) and there is a pattern. This is all taken from one amplification of the base of the gun (number three).

(Next Slide)

Now, when the computer tries that possibility out, it will expand that and create a pattern. What we are seeing (in green) is the original EPG data. It is continuous. There are no thresholds. It is whatever it is, whether it is 100 or 1000 or 10 rfu. That is the data. In most statistical reasoning, we do not touch the data. We do not do anything to it. The data is what we are conditioning on because we are

trying to determine the genetic identity conditioned on the data. Once we change the data, we are determining something, but it may not quite be the genetic identity. Now, the computer is showing those as triangles in the Explain interface (in gray). That is a model that can get created where we say, “Suppose that we have this cartoon with what the genotypes would be. How much for this two-person mixture (major and minor)? How much DNA would be at each allelic location? How much stutter might be there? How much relative amplification? How much decay?” (This is not modeled here because it was not needed in this case.) The question we are asking of a likelihood function is how well does our model, pattern, prediction explain the observed data.

On this case, we can visually see and the computer can mathematically see that there is very little deviation between the peak heights (shown in the data in green) and the proposed pattern (shown in gray). It is a very good fit. If we had chosen different alleles with very different values, then we might get no match at all. What we are seeing visually is very much what the computer is seeing numerically. In this case, the data would have a high likelihood. Valid statistical inference requires us or a computer to consider every possibility and combination, particularly for computers, of what the allele pairs would be for each of the two contributors, the different amounts of the DNA, the mixing proportions at each experiment, the stutter, the relative amplification, and so on. The computer does that over and over. It just tries everything out, and when it is done, what is most likely when combined with priors ends up being most

probable.

(Next Slide)

What we get is a probability distribution, which at every different locus in this case in SMGplus has 100% probability. From the computer's perspective, that was a very easy problem looking at one EPG, or one amplification, from the base of the handgun. It had a definite genotype at every locus, and as a result, it pulled out the full random match probability with the likelihood ratio having information or its logarithm of 16. We see 1 trillion-fold increase over CPI of 10^4 to 10^{16}.

(Next Slide)

This is nice, but what also happened in this case is that the computer inferred a minor genotype. What we are seeing at each locus here is a genotype probability distribution. When there is uncertainty, all mixture interpretation methods produce a probability distribution. The more the probability bar is to the right, the more definite it might be about some allele pair possibilities over others. We do not get that with inclusion or RMNE to the same extent as a method. We have genotype uncertainty, and when we compare it against the suspect profile, we get a log(LR) information of 5, which is a 100,000-fold increase. This would be good enough to stop given that the person was convicted with just 17,000, and they

never found the minor person. However, we are going to continue looking at this minor contributor. I should stress that we do not need to continue with this for a criminal case. It is just that we can, and the question scientifically is, “How much information can we get out of the data?”

(Next Slide)

Let us take a look. Clearly, if we have one swab, then we can look at both amplifications. People regularly look at two amplifications from the same item together. The computer does that as well. It assumes these particular allele pair values, tries them out, and generates all possible patterns for different variables. It tries out every possible allele pair, and it keeps asking, “How well does the proposed hypothesis explain the observed data?” Because the results are conditionally independent on the assumptions, we are allowed to multiply the two likelihood numbers together. We see the likelihood as a pattern. The computer sees the likelihood as a number. It multiplies those two numbers together and tries it out for all different genotype possibilities.

(Next Slide)

The computer ends up with a little bit more information. Looking at two amplifications from the one PCR template, we have a little more certainty, and the information content went up to a little over five. There was a gain from the

human result of nothing to under five to over five.

(Next Slide)

Well, we have four templates. This is the mixture weight interface from TrueAllele. It is showing the mixture weight probability distribution of the template. There are separate mixture proportions at each locus, but the template itself also has a mixture variable. We see that they are around 15%. What we are looking at is 0% to 100% mixture weight, on the x-axis, and then the y-axis is scaled to show a histogram. The probability distribution is around 15% for the minor and 85% for the major. It is different for each template. So, what we are going to do now is combine in the interpretation all the data from items 3 and 4.

(Next Slide)

Now, we have a joint likelihood function where we see the first item and its two amplifications at the locus D18. We can look at the data. Here, for the back of the gun, are the two amplifications. Notice that these two patterns are similar with some peak variation. Those two patterns are similar amongst themselves with peak variation, but the patterns between the two templates are different, which we would expect. There is a different proportion of how much is in each of the templates of the two contributors. The computer tries out all parameter possibilities and sees which are more probable.

(Next Slide)

Now, when it infers the minor genotype, we are seeing that there is an increase in probability towards one particular allele pair at every locus. We are using two amplifications of two PCR templates, and we end up with a match strength of 10^9, which is a billion to one. This is a larger increase, and as we keep moving along, we will do it again.

(Next Slide)

We next look at all four items and how they get interpreted jointly with duplicate amplifications. The computer assesses how well the proposed genotype and the pattern that it produces fit the data. Those genotypes and the patterns they produce have to satisfy the constraints of the data from all eight experiments. From a probabilistic perspective, all we are trying to do is get what the genotype is using probability. When there is uncertainty, there is probability conditioned on the data. The more data that we have, the more informative our answer might be because of the way the modeling methods work. The data are conditionally independent. Each fitting or comparison of a proposed pattern to the observed data gives a number when it is based on a genotype and other variables. That is the likelihood, which is the probability of observing this particular data given the genotype and dozens of other parameters. Those produce a number at each

data pattern, and those eight numbers get multiplied together. The data here are so constraining that there is really only one answer that could fit all of the data constraints.

(Next Slide)

What we end up with is a unique profile. That is the genotype. There is only one allele pair possibility now listed at each of the SGMplus loci, and it has 100% probability. The result is that we achieve the full match strength of 10^{12}, or a trillion to 1. We see in the succession of LR increase that we are basically done. What is the expression? "Don't try this at home." Obviously, a detailed mathematical theory and some good calculators are needed to do this. These are four PCR templates with two amplifications of each.

(Next Slide)

What I would like to show now is the overall trend in information as we introduce more and more data. This is taken from the Report interface of TrueAllele. The stuff on the left I just added so it would be easier to see. Here are the first eight items. They are grouped. These two were from swabs 3, 4, 5, and 6. They are single amplifications. How much information do we get? We are looking at the top now. We see already that we are getting a minor contributor giving a match strength of 10^{8}, which is 100 million, from one of those swabs from TrueAllele's

perspective. Again, in the real world we would have stopped unless we wanted to know the profile exactly.

(Next Slide)

We move on and ask jointly, “What happens when we look at both amplifications for each template with a joint analysis with two amplifications (3, 4, 5 and 6) from the handgun?” We see that overall there is an increase in information. That group of four is moving over to the right a little bit. We are interested the most informative one that is going up to 10^9.

(Next Slide)

We move down one more row grouping and ask now, “What if we did all of the different pair-wise combinations of the two templates?” Again, the orange four-wise combinations are shifting over to the right, and we are going up to LRs of 10^{10} and 10^{11} until we finally max out with all eight up to 10^{12}.

(Next Slide)

These are the observations from the study. Then, I will give some conclusions. We looked at co-ancestry, which we can do in the Report interface, and that cost about half of a log unit. It is still around a trillion to one. With three unknown

contributors, which we ran through the system, we essentially got the same results. We see that multiple amplifications can be good if our goal is to get as much information as we can, and multiple templates are even better. Why? It is because the data is what constrains the genotypes. The inferred genotypes are conditioned on the data. The more data that we have and the more conditionally independent amplifications, then the more constraints we have. So, if we are wondering when doing pair-wise comparisons (which we are doing studies on), which are good amplifications to combine, it is often ones that are dissimilar because the data constraints are different. Data that is unlike other data when we put it together imposes different constraints and restricts the possibilities of what the inferred genotypes can be.

(Next Slide)

In conclusion, we looked at two mixture genotype examples. We looked at the major contributor and the minor contributor. This is what we found. With the major contributor, we saw that there is a quantitative likelihood function, and that model uses all of the data. It does not use what some threshold says, such as maybe we should go over 40 rfu or maybe over 50 or some other lab would say 150. I will not go into the philosophical issue of what is an allele. Let us not go there. Ask later.

Quantitative likelihood modeling compares predicted patterns with peak height

against the full continuous data that are present. It is more informative than qualitative threshold-based methods. How much more so? Well, we saw that we went from the reported 10^4 to 10^{16}. So, the major contributor is 1 trillion times more informative.

For the minor contributor, we saw that starting with that same one amplification of the item of the base of the gun (number 3) that we had a log(LR) of 0 because there was no information available from the human review at all. Using the joint likelihood function, which took products of the quantitative likelihoods over more and more of the data, we eventually found that was more informative than taking any one particular item in isolation. How much more informative? Well, we went from 10^0 with human review to 10^{12} with a unique profile with a joint analysis. Again, it is about trillion times more informative than human review.

In conclusion, it is good to use all of the data, and we get more information when we do. Thank you.