Accountability in Research, Vol. 3, pp. 47–54
Reprints available directly from the publisher
Photocopying permitted by license only

©1993 Gordon and Breach Science Publishers S.A.
Printed in the United States of America

Unresolved Issues in the Forensic Use of DNA Profiling

Ronald Ostrowski and Dan E. Krane†*

**The University of North Carolina at Charlotte, Department of Biology, Charlotte, NC 28223*

†Department of Biological Sciences, Wright State University, Dayton, OH 45435

The enormous amount of genetic diversity in humans allows for a powerful form of individual identification. This "DNA profiling" is based on the fact that sites within the human genome have variable numbers of tandem repeats (VNTRs) and has been hailed in forensic sciences as the greatest discovery since fingerprinting. The techniques involved are virtually the same as those used in all molecular biology laboratories. A major difference however is that in forensic science DNA samples can be less than ideal in both quality and quantity. Furthermore, in basic molecular biology the origin of the sample is known while in forensic testing it is not. Thus, the challenge is to reconcile a "match" between a crime scene DNA sample and one from a suspect(s). Presently, a debate exists regarding the use of the unmodified product rule versus a more conservative ceiling principle approach to calculate the probability of a coincidentally matching DNA profiles. The latter was endorsed in a recently published report by the prestigious National Research Council but has not received widespread support from testing laboratories. Further exacerbating the debate over how much weight should be attached to DNA profile evidence is a lack of widely accepted standards for forensic laboratories especially in the areas of proficiency testing, publication of error rates and laboratory personnel certification.

Keywords: VNTRs, DNA profiling, product rule, ceiling principle, laboratory standards

Each human genome, the sum total of each of our sets of genetic information, has been estimated to differ from every other human genome at three million nucleotide sites [Hofker, et al., 1986]. The enormous amount of genetic diversity that those differences represent means that is should be possible to uniquely identify every individual (barring only monozygotic twins) based only on the sequence of their DNA [Botstein, et al., 1980]. Obtaining the sequence of a human genome, however, is a daunting task—current estimates are that it will take many hundreds of millions of dollars and the dedicated effort of hundreds of scientists to complete the sequencing of a composite human genome over the course of roughly fifteen years. Obviously, an analysis of an individual's complete genome is not a practical means of identification at the present time. However, research has shown that a sizable portion of the three million differences between individuals at the level of their DNA is concentrated in regions containing *v*ariable *n*umbers of *t*andem *r*epeats (VNTRs) [Jeffreys, et al., 1985a, 1985b; Gill, et al., 1985]. By looking specifically at these easily examined loci, scientists have found a powerful new means of human identification

and their discovery has been widely hailed as one of the greatest advances in forensic science since the discovery of fingerprints almost one hundred years ago.

VNTRs are highly polymorphic regions of the genome in which small sequences of DNA, often between 15 and 30 nucleotides long, are tandemly repeated a variable number of times. For instance, one allele may have 31 copies of the repeated sequence while another may have 43. The utility of these loci in studies such as paternity testing, forensic analyses and anthropological research in which the ability to distinguish between individuals or groups of individuals is important was quickly appreciated. The use of VNTRs in forensic analyses is typically referred to as "DNA profiling" and it is in that arena that their use has received the greatest attention, both in the scientific community as well as in the popular press.

In general, the techniques involved in DNA profiling are the same as those used in virtually any laboratory that works in the field of molecular biology. Testing labs do occasionally employ procedures that defense attorneys have argued make some of their conclusions suspect but continual refinements in the protocols used are making such challenges increasingly rare and the reliability of the methods is not often an issue in courts any longer. Unlike the experiments that scientists typically perform however, the quality of material in forensic cases is often poor and the quantities of that material can prohibit the resolution of ambiguous results by simply repeating the experiment. This inability to satisfy one of the basic tenants of science poses one of the greatest challenges forensic scientists must face and requires that they are scrupulously careful in their initial, and usually only, analysis of material. The results obtained when that challenge is met however can be most impressive and can usually be easily distinguished from cases in which samples were too small or degraded to allow an unambiguous inclusion or exclusion of an individual. Additionally, the significance of the exclusion of an individual as a suspect that could have contributed a biological sample to a crime scene has been quite clear from the start—in all normal circumstances (as in those in which laboratory error and degradation are not a concern), if a suspect's VNTR banding pattern does not match the banding pattern seen in the evidence, the suspect simply could not have contributed it.

What significance should be attached to the finding of a matching DNA profile, however, has been at the heart of a heated and lengthy debate between scientists and attorneys alike. Few would argue that finding two unrelated individuals who coincidentally possess matching DNA profiles at several of these highly polymorphic loci is an extremely rare event. Initially, it was easy to convince a jury that a match in DNA profiles between a suspect and an evidentiary sample was a damning piece of evidence and both prosecutors and defense attorneys left the issue at that. But DNA obtained from a sample and a suspect is never determined to be unequivocally identical and increasing emphasis has been placed on stating precisely how damning such a match is. It is unlikely that the true probability of a coincidental match will ever be determined with absolute certainty and that emphasis has led to arguments about not only the manner in which the likelihood of random individuals having the same DNA profile are calculated, but also the frequency at which testing laboratories make serious errors in their analyses.

The manner in which forensic labs have traditionally estimated the probability of a VNTR match between two randomly chosen individuals has been based on the

multiplication of the frequencies of the relevant VNTR alleles occurring in a database of a particular racial group [Budowle, et al., 1991]. If the frequency in a general population of finding a certain VNTR allele is 0.2, and that of another allele at the same locus is 0.1, the frequency of finding a single individual with both those bands is said to be 0.2×0.1 or 0.02. If an individual with those two bands was also tested at an additional locus and the bands found there were present in the general population with a frequency of 0.2 and 0.2, the combined genotype frequency would be calculated as $0.2 \times 0.1 \times 0.2 \times 0.2$ or 8×10^{-4} which corresponds to odds of one in 1,250. Multiplication of frequencies in this way is often referred to as invoking the "product rule"—a well accepted statistical principle in and of itself. The more loci probed and the less common the alleles found, the more dramatic the numbers in the end. Odds as impressive as one in ten billion are commonly introduced in court after having been calculated in this way. Such odds of coincidental matches, even those as low as one in 1,250, would seem sufficient to compel a conviction.

There has been extensive argument within the scientific community however regarding whether or not such calculations are correct [Lander, 1989; Lewontin and Hartl, 1991; Chakraborty and Kidd, 1991]. One of the fundamental assumptions of the product rule is that each of the events observed (i.e. detecting a particular band at a given locus) occurs independently. If having one band at a VNTR locus makes an individual more likely to have a particular band at that or another locus then the product rule, in its simplest form, cannot properly be used. In reality, scientists have known for a very long time that certain genetically inherited traits associate non-randomly [Lewontin, 1972; Mourant, 1976]. Many human traits are correlated (and are often associated with an individual's ethnic background) and can be used as evidence of human population substructuring. Blonde hair and blue eyes, for instance, are each found in about one in five Caucasian individuals (a frequency of 0.2) but the probability of finding a Caucasian with both blonde hair and blue eyes is not 0.2×0.2 (0.04, or one in 25) but, rather something more like 0.2 since those traits are so often linked. In addition to other physical traits such as complexion and hair type, other loci (such as blood group and enzyme coding genes) which would appear to be under less direct natural selection have also been found to have gene frequencies that are consistent with significant substructuring within the human races [Lewontin and Hartl, 1991]. At the same time, other loci have been found to be indicative of little or no human substructuring [Smouse et al., 1982]. These studies leave the possibility of evidence for population substructuring at the VNTR loci used in forensic analyses an open issue.

Several alternatives to the product rule have been put forward with the concerns associated with population substructuring at these loci in mind. One alternative has been referred to as a "1/N approach" which entails estimating the frequency of occurrence of a given DNA profile as the reciprocal of the number of times that the particular profile actually is seen in a reference database [NRC report, 1992]. Another approach avoids arguments over specific frequencies simply by referring to matches with frequencies of occurrence less likely than one in 1,000 as "very strong" evidence [Evett and Werrett, 1990]. Still another alternative, referred to as "the ceiling principle," suggests determining VNTR allele frequencies for a large number of ethnic groups and using only the highest frequencies seen in any group

as the numbers to be multiplied, thereby arriving at a result that is almost certain to be conservative regardless of an individual's ethnicity [Lander, 1991; NRC report, 1992].

The National Research Council (NRC), a branch of the prestigious National Academy of Science, recently issued a report [NRC report, 1992] designed to address many of the issues related to the use of DNA technology in forensic setting. Its conclusion in regard to the appropriate way to calculate combined genotype frequencies is clear: since it is possible that VNTR allele frequencies can differ from one human subpopulation to another, it recommends that the ceiling principle be used in all genotype frequency determinations. Of all the methods considered, only the ceiling principle means of calculating frequencies simultaneously accounts for whatever effects human population substructuring might have as well as yields final probabilities which are still reflective of the considerable power of DNA profiling in general. Realizing that some time would be required to ascertain the frequency of VNTR alleles in a number of ethnic groups, the NRC also proposed an interim means of calculating combined genotype frequencies that involves reporting both a conservatively modified ceiling principle number and the 1/N number.

Unfortunately, the NRC's report has not succeeded in its attempt to resolve the controversy surrounding the manner in which combined genotype frequencies are determined [Ayala, 1992; Devlin et al., 1993]. Supporters of the ceiling principle, essentially those who have opposed the use of the unmodified product rule previously, cite it as an acceptable compromise that results in a consistently conservative estimate of the true probability of a coincidental match that is still powerful enough to allow convictions [Aldhous, 1993]. However, labs involved in forensic DNA profiling have maintained that other parts of their analyses are sufficiently conservative that small amounts of population substructuring need not be taken specifically into account by the ceiling principle. They also point out that the ceiling principle makes relatively arbitrary assumptions regarding the lowest possible frequency that can be used in determining a combined genotype frequency [Devlin et al., 1993]. It is true that data needed to assess the true extent of ethnic variation at VNTR loci is only just beginning to be reported in the scientific literature [Balzas, et al., 1992; Krane, et al., 1992; Sajantila, et al., 1992]. Those initial studies however indicate that the level of substructuring seen is small but not insignificant and that the admittedly arbitrarily chosen limits on allele rarities of the ceiling principle do in fact result in a method that is not unfairly biased against defendants and yield probabilities that are actually only slightly less damning than those generated by the product rule [Krane, et al., 1992]. The traditional product rule however is to some degree unfairly prejudicial against defendants in at least half of the cases it is used according to that same study [Krane, et al., 1992].

It is not surprising that the initial studies of the interim ceiling principle indicates that it is consistently more conservative (and, as a result, yields less impressive probabilities of a coincidental match) than the unmodified product rule. The initial indications are that slight differences in allele frequencies between ethnic groups can be found and, in a strict sense, constitutes a violation of a fundamental assumption of the product rule. The differences however do not seem to be nearly as pronounced as they are for the earlier example of blonde hair and blue eyes. In some respects the argument has now become "What does it matter if the odds are

one in 100,000 or one in 1,000,000?" Prosecutors say, "Very little," while defense attorneys would maintain, "In a legal system that demands we err on the side of the accused, very much—especially in cases where the numbers are one in ten vs. one in 100."

Consequently, the debate over the proper way to determine the probability of a coincidentally matching DNA profile in a forensic case has raged on despite the release of the NRC report. The net result in court rooms since the release of the report has been the admittance of ceiling principle and 1/N generated numbers as evidence in criminal cases since experts on both sides of the argument generally agree that those calculations generate conservative assessments of the true probability of a coincidental match. Most forensic labs however have not abandoned their effort to introduce numbers generated by the unmodified product rule and are consistently finding themselves challenged in that as defense attorneys becoming increasingly familiar with the methodology involved. Unfortunately, the forensic laboratories' position that human populations are not sufficiently substructured to merit the conservative approach of the ceiling principle has little experimental support along the lines called for by the NRC report to date [Lewontin and Hartl, 1991]. The issue appears to have little prospect of resolution until a number of additional studies such as the initial test of the ceiling principle [Krane, et al., 1992] are completed. Many such studies are currently underway [Krane et al., in preparation; Evett et al., personal communication; Budowle et al., personal communication] and may contribute to that resolution. More likely though, the ultimate answer to the question phrased in very uncompromising terms by scientists on both sides of the issue will not be arrived at until after the issue becomes moot due to the introduction of promising new technologies in the next three to five years [Jeffreys et al., 1991].

An additional concern which may make the debate over the proper method of calculating VNTR profile frequencies seem like splitting fine hairs is also beginning to receive increased attention in court rooms due to the recommendations of the NRC report. Relatively few opportunities exist for forensic laboratories to demonstrate their proficiency in a clearly unbiased way and, occasionally, errors in forensic DNA reports have been found. If a laboratory misinterpretation of a VNTR profile results in a false inclusion of a suspect with a frequency of one case out of 10,000, then it can be argued that it is that number which should be reported to a jury rather than any more damning VNTR frequency number which may have been generated.

It is not unusual for innovative techniques to be pressed into service before all experts have agreed to what constitutes proper regulation and standards for the newly developed method. The forensic use of DNA profiling has not proven to be an exception. Typical standardization problems caused by enthusiasm to employ a new methodology have been greatly compounded by the very nature of forensic sample collection in that little is known about the condition of the material to be tested from one case to the next.

In addressing this issue, Roberts [1991] has stated that forensic data must be reliable and not necessarily perfect. This statement is disconcerting. The extensive proficiency testing and quality control standards applied to most analyses performed in clinical laboratories appear to be built on the assumption that perfection,

or at least a close approximation, is in fact an attainable goal. And, while the misdiagnosis of a genetic disease can have serious implications, it is also true that the results of a mistaken DNA profile analysis can be equally as grievous from the perspective of both the victim and the accused. It certainly seems reasonable to apply at least the same standards to forensic DNA profile analyses as those which are applied to much more common clinical tests such as those for genic and chromosomal abnormalities.

Admittedly, perfection is virtually unattainable in an area such as the forensic use of DNA typing where the quality and quantity of the crime scene sample can be less than desirable. Furthermore, no formal standardization of laboratory procedures exists at this time. These short comings are further exacerbated by the lack of agreement by forensic laboratories to submit to standard types of proficiency testing and personnel certification. This form of quality control and assurance is a traditional cornerstone of scientific and medical laboratories. Unaffiliated, unbiased agencies such as the College of American Pathologists, the Centers for Disease Control and the American Society of Histocompatibility, to name but a few, conduct oversight programs that insure the quality of laboratory results and thereby confidence in data for tests ranging from genotype analyses for genetic diseases to karyotyping.

The oversight of these agencies is tailored to a laboratory's particular specialty(s), however, some general formats can be seen. One method is to provide unknown control samples for comparative analysis by several labs. The general success and error rates of each lab using such samples are subsequently published. A second method, and often one that is complementary to the former is on site visits by inspectors. The accuracy and timeliness of protocols and manuals, safety procedures in use, personnel training and rectification of previous deficiencies, etc. are considered. Forensic laboratories have maintained that such visits may compromise the identity of their employees in certain cases. They have also argued that exposing evidence from other cases may render it inadmissible in a court of law [Ostrowski and Bender, 1993]. Furthermore, private laboratories are naturally concerned about giving away trade secrets, be they new methods of analysis or tediously constructed data bases, in a competitive market.

Most forensic laboratories have begun to take steps in the right direction though. Since 1985, a voluntary laboratory-accreditation program has been conducted through the auspices of the American Society of Crime Laboratory Directors-Laboratory Accreditation Board. Unfortunately, this program is not flawless in that it does not list laboratories that have failed examinations or had deficiencies [NRC report, 1992].

A practitioners group called the Technical Working Group on DNA Analysis Methods (TWGDAM) have published guidelines for proficiency testing [TWGDAM, 1989, 1990]. The NRC report [1992] however, takes some issue in that at the present time these are nothing more than recommendations. To be most effective these recommendations would have to be implemented in a formal, detailed program in which all DNA forensic laboratories would be expected to participate in order to be credible.

It is clear that the forensic use of DNA profiling is still in a transitional period for both laboratory procedures and statistical analysis of the data. Until laboratories are

expected to participate in properly administered unbiased and external accreditation and proficiency testing programs, their credibility and reliability will remain an open issue—much to the detriment of many laboratories that already adhere to high standards. Additionally, studies such as those suggested by the NRC into the extent of human population substructuring apparent at VNTR loci remain incomplete. Until those studies are performed, forensic laboratories will not be able to convincingly argue that their methods are already conservative enough to make using the ceiling principle unnecessary. These issues should be resolvable before VNTR analyses are superceded by the introduction of new methodologies as additional work is done in this area and if the parties involved can set aside their bitterly contentious stances. In the mean time, caution should be used to insure that any calculations favor defendants and that the regulatory agencies recommended by the NRC are quickly put into place.

LITERATURE CITED

Aldhous, P. (1993) Geneticists attack NRC report as scientifically flawed. *Science* 259:755–756.

Ayala, F.J. (1992) DNA Law. *J. Mol. Evol.* 35:273–276.

Balzas, I., Neuweiler, J., Gunn, P., Kidd, J., Kidd, K.K., Kuhl, J. & Mingjun, L. (1992) Human population genetic studies using hypervariable loci l. Analysis of Assamese, Australian, Cambodian, Caucasian, Chinese and Melanesian populations. *Genetics* 131:191–198.

Botstein, D., White, R.L., Skolnick, M. & Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32:314–331.

Budowle, B., Giusti, A. M. Waye, J.S. Baechtel, F.S., Fourney, R.M., Adams, D.E. Presley, L.A., Deadman, H.A. & Monson, K.L. (1991) Fixed-bin analysis for statistical evaluation of continuous distributions of allelic data from VNTR loci, for use in forensic comparisons. *Am. J. Hum. Genet.* 48:819–823.

Chakraborty, R. & Kidd, K.K. (1991) The utility of DNA typing in forensic work. *Science* 254:1735–1739.

Devlin, B., Risch, N. & Roeder, K. (1993) Statistical evaluation of DNA fingerprinting: A critique of the NRC's report. *Science* 259:748–749.

Evett, I.W. & Werret, D.J. (1990) in *Proceedings of the International Symposium on Human Identification* (Promega, WI). pp. 77–91.

Gill, P., Jeffreys, A.J. & Werret, D.J. (1985) Application of DNA "fingerprints." *Nature* 318:577–579.

Hofker, M.H., Skraastad, M.I., Bergen, A.A.B., Wapenaar, M.C., Backker, E., Millington-Ward, A., van Omenn, J.B. & Pearson, P.L. (1986) The X chromosome shows less genetic variation at restriction sites than the autosomes. *Am. J. Hum. Gen.* 39:438–449j.

Jeffreys, A.J., Wilson, V. & Thein, S.L. (1985a) Hypervariable minisatellite regions of human DNA. *Nature* 314:67–73.

Jeffreys, A.J., Wilson, V. & Thein, S.L. (1985b) Individual-specific "fingerprints" of human DNA. *Nature* 316:75–79.

Jeffreys, A.J., MacLeod, A., Tamaki, K., Neil, D.L. & Monckton, A.J. (191) Minisatellite repeat coding as a digital approach to DNA typing. *Nature* 354:204–209.

Krane, D.E., Allen, R.W., Sawyer, S.A., Petrov, D.A. & Hartl, D.L. (1992) Genetic differences at four DNA typing loci in Finnish, Italian, and mixed Caucasian populations. *Proc. Natl. Acad. Sci. USA* 89:10583–10587.

Lander, E.S. (1989) DNA fingerprinting on trial. *Nature* 399:501–505.

Lander, E.S. (1991) Research on DNA typing catching up with courtroom application. *Am. J. Hum. Gen.* 48:819–823.

Lewontin, R.C. (1972) The appointment of human diversity. *Evol. Biol.* 6:381–398.

Lewontin, R.C. & Hartl, D.L. (1991) Population genetics in forensic DNA typing. *Science* 254:1745–1750.

Mourant, A.E., Kopec, A.C. & Domaniewska-Sobczak, K. (1976) *The distribution of human blood groups and other polymorphisms*. New York, Oxford University Press.

National Research Council (1992) DNA technology in forensic science. Washington, D.C. National Academy of Science Press.

Ostrowski, R.S. & Bender, H. (1993) DNA profiling in forensic sciences. *Trialbriefs*, in press.

Roberts, R.J. (1992) The societal impact of DNA fingerprint data. *Acc. Res.* 2:87–92.

Sajantila, A., Budowle, B., Strom, M., Johnsson, V., Lukka, M., Peltonen, L. & Ehnholm, C. (1992) PCR amplification of alleles at the D1S80 locus: comparison of a Finnish and a North American Caucasian population sample, and forensic casework evaluation. *Am. J. Hum. Gener.* 50:816–825.

Smouse, P.E., Spielman, R.S. & Park, M.H. Multiple-locus allocation of individuals to groups as a function of the genetic variation within and differences among human populations. (1982) *Am. Nat.* 119:445–463.

Technical Working Group on DNA Analysis Methods (TWGDAM). (1989) Guidelines for a quality assurance program for DNA restriction fragment length polymorphism analysis. *Crime Lab Dig.* 16: 40–59.

Technical Working Group on DNA Analysis Methods (TWGDAM). (1990) Guidelines for a proficiency testing program for DNA restriction fragment length polymorphism analysis. *Crime Lab Dig.* 17: 50–60.