Am. J. Hum. Genet. 54:931-933, 1994

INVITED EDITORIAL
DNA Data Banking: A Cautionary Tale

Barry Scheck

Benjamin N. Cardozo School of Law, New York

The forensic DNA data-bank legislation passed so far in 19 states raises disturbing questions about whether the scientific community in general—and the genetics community in particular—can effectively influence the use of DNA technology by law enforcement. The weaknesses in the data-bank statutes reviewed here by McEwen and Reilly (1994) range from very practical and technical matters to questions of ethical and constitutional principle. What stands out, however, is that clear warnings from scientists about the problems involved in setting up DNA data banks have been substantially ignored by legislators and law enforcement.

The scientific community has consistently called for meaningful quality assurance in forensic DNA laboratories, which would involve accreditation of laboratories, certification of personnel, and regular proficiency testing by an independent group of scientists (Ad Hoc Committee on Individual Identification by DNA Analysis, The American Society of Human Genetics 1990; National Academy of Sciences 1992). Such independent quality assurance, whether carried out by a private professional group or by a state-created entity, costs money. Unfortunately, serious quality assurance is a prosaic legislative proposal, unlikely to garner funds from budget-conscious legislators unless there is a major lobbying effort from those who recognize its importance. A DNA data bank, on the other hand, can be promoted by law enforcement, as a magic bullet in the war on crime, an initiative that will lead to the spectacular apprehension of dangerous recidivist criminals.

Yet, as a practical matter, the short-term decision to fund DNA data banks, rather than quality assurance, will create serious long-term problems for forensic DNA laboratories, both in the analysis of casework and in the operation of the data banks. The current RFLP methods being used to build the data banks are comparatively expensive. Typing a blood sample taken from a convicted individual will cost about $100 to perform, and storage of that sample will run about $20. Thus, the cost of just typing and storing an RFLP data bank of more than 10,000 samples would be more than $1.2 million (National Academy of Sciences 1992). There will also be additional expenses involved in gathering the blood samples from thousands of prisoners and conducting the extensive and continuous intra- and interlaboratory reproducibility tests that are necessary for reliable operation of data banks, in light of the measurement-variability problems associated with the RFLP methods used by the forensic laboratories. It is quite doubtful that most state laboratories can now afford to do much more than collect blood samples, especially if the state legislation calls for data banking in broad range of cases (all felonies or all violent felonies, as opposed to just sex offenders).

On the other hand, the development of informative PCR-based forensic tests using discrete allele systems seems imminent, and such testing would almost certainly be a cheaper and more efficient technology to employ in building a data bank. Consequently, the National Research Council (NRC) of the National Academy of Sciences (1992) warned against "premature" development of DNA data banks based on current RFLP technology, because "it runs the risk of perpetuating a 'dinosaur' technology in the face of better techniques" (p. 116).

A further practical complication of investing heavily in data banks rather than in quality-assurance programs is that, increasingly, judges—and juries—will be demanding that forensic DNA laboratories produce reliable estimates of laboratory error rates. The recent landmark Supreme Court case articulating a new standard for the admission of scientific evidence, Daubert

Received March 30, 1994; accepted for publication March 31, 1994.

Address for correspondence and reprints: Mr. Barry Scheck, Benjamin N. Cardozo School of Law, Criminal Law and Appeals Clinics, 55 Fifth Avenue, Room 1701, New York, NY 10003.

0002-9297/94/5406-0001$02.00

v. Merrell Dow Pharmaceutical, Inc. (1993), is bound to focus serious attention on this issue, since the decision explicitly emphasizes error rate as an important factor that courts should consider in assessing the reliability of a scientific technique (Scheck 1994).

The population genetics debate over the methods used by forensic DNA laboratories to calculate the frequency of a matching DNA profile in some specified population has distracted attention from a consensus that has emerged on the error-rate issue (Hagerman 1990; Lempert 1991; Lewontin and Hartl 1991; National Academy of Sciences 1992; Weir 1992; Koehler 1993; Mueller 1993; Thompson 1993). When a "match" between DNA trace evidence found at a crime scene and a suspect is offered into evidence, it is just as important to know the probability that a false-positive "match" might be declared because of laboratory errors—such as sample handling problems—as it is to know the probability that coincidentally matching DNA profiles might occur. Indeed, it is commonly expected that the false-positive error rate of a forensic DNA laboratory is, in most cases, likely to be much higher than the frequency of a coincidental match.

There has been some disagreement about whether the false-positive error rate and the frequency of a coincidental match should be presented separately to juries (National Academy of Sciences 1992) or combined in one likelihood ratio (Hagerman 1990; Weir 1992). There has also been concern expressed about the difficulties involved in measuring laboratory error (Weir 1992). Yet there has been no serious dissent about the necessity of ascertaining laboratory error rate.

A reasonable estimate of laboratory error rate (both false positives and false negatives) requires a much more substantial commitment to quality assurance than the Federal Bureau of Investigation (FBI) or state forensic laboratories have been willing to make so far. The NRC strongly recommended that error rates be derived from having forensic laboratories undergo regular blind proficiency tests designed by external entities, on samples that replicate casework. Despite the force of that recommendation, there has yet to be one blind external proficiency test on any forensic DNA laboratory, much less a regular program of blind external proficiency testing.

A strong quality-assurance program is also indispensable to the reliable construction and operation of the forensic data banks. The principal problem here is not a false-positive data-bank match (McEwen and Reilly [1994] correctly point out that a false positive is likely to be discovered by subsequent testing of a suspect), but a false-negative data-bank match. A false negative could irretrievably steer investigators away from the apprehension of a violent offender whose DNA profile was missed in a data-bank search because of either faulty typing of the data-bank sample or a sample handling error.

The DNA Identification Act now pending in Congress will not necessarily do much to solve the quality-assurance problem (DNA Identification Act of 1993). Although this legislation requires that state laboratories must meet certain minimal quality-assurance standards if they are to be part of the national DNA data bank (the so-called CODIS system), these standards will be set by the Director of the FBI, in consultation with an advisory committee selected by the FBI director. Without an independent group of scientists and forensic scientists setting quality-assurance standards, as recommended by the NRC, it is simply unrealistic to expect that the FBI will impose rigorous quality-assurance standards on itself and state laboratories, especially if implementation of those standards would be costly. Indeed, the NRC found that the quality-assurance guidelines produced so far by TWGDAM (Technical Working Group on DNA Analysis Methods), a forensic laboratory group led by the FBI, did not go far enough, that the Department of Justice (DOJ) lacked expertise in quality assurance in molecular and population genetics, and that public confidence would be undermined by having the FBI regulate itself and all the other laboratories involved in setting up a national DNA data bank.

Even more disturbing, the state DNA-data-bank statutes reflect, as McEwen and Reilly (1994) point out, inadequate consideration of the privacy, ethical, and constitutional concerns that have worried the scientific community. To the extent that the statutes limit either access to or use of DNA data-bank profiles or samples, it is for "law enforcement purposes." This term is disturbingly broad. It might encompass personnel other than state police (e.g., immigration officials, child support bureaus, or welfare agencies), and it might permit DNA data to be used for "law enforcement" purposes beyond identification testimony in criminal cases. Nor is it alarmist to be concerned that, under these data-bank statutes, researchers might be able to use either the DNA-data-bank blood samples that state laboratories are warehousing—or even DNA profile data—to conduct "law enforcement" studies on the genetic makeup of sex offenders, drug addicts, alcoholics, or violent offenders generally.

As McEwen and Reilly (1994) point out, many of the data-bank statutes do not restrict the collection of sam-

ples to sex offenders or persons who have committed crimes of violence, but include nonviolent felons and some misdemeanants. Since there is little evidence to show that nonviolent offenders are likely to commit rapes or violent crimes where DNA typing is useful, the rationale for including them in a DNA data bank is weak. Indeed, there is a serious constitutional question as to whether data banking of samples from nonviolent offenders—or even from non-sex offenders—violates the Fourth Amendment's prohibition against unreasonable searches and seizures. There have been just a few cases challenging the constitutionality of DNA data-bank statutes, and, although the statutes were upheld, some judges expressed the view that the Fourth Amendment would curtail collection from nonviolent inmates unless there was a stronger showing that such a seizure of bodily fluids and private genetic data would be productive in the prospective apprehension of criminals (Jones v. Murray 1992; Ryncarz v. Eikenberry 1993).

Certainly, use of DNA data-bank information or samples, either to develop investigative leads on relative of convicted offenders or for "law enforcement" reasons other than the future identification of convicted offenders in the data bank would be subject to constitutional challenge, notwithstanding the failure of statutes to delineate strict limitations on use or access. But extended constitutional litigation is a very blunt and ineffective instrument when it comes to solving the often subtle ethical and privacy problems that the scientific community has already anticipated will arise from the operation of DNA data banks.

It must be emphasized, of course, that DNA data banks promise to play a very important role in effective law enforcement, particularly in the solution of violent crimes. This potential has never been in doubt; nor has the potential threat to civil liberties that is occasioned by such a massive collection of DNA samples and data. The key issue, however, is one of process and power. Law-enforcement officials must not be the sole custodians, administrators, and policy-makers who control access to and use of DNA data banks. Yet no state, with the possible exception of Michigan, has taken up the suggestion of the NRC that an advisory board of scientists, ethicists, and other professionals exercise meaningful oversight over the operation of DNA data banks. The genetics community will get only so many wake-up calls before it is too late to control the proliferation of genetic data by law enforcement.

References

Ad Hoc Committee on Individual Identification by DNA Analysis, The American Society of Human Genetics (1990) Individual identification by DNA analysis: points to consider. Am J Hum Genet 46:631–634

Daubert v Merrell Dow Pharmaceuticals, Inc, 113 S Ct 2786 (1993)

DNA Identification Act of 1993, HR 829, S 497, 103d Con, 1st Sess (1993)

Hagerman PJ (1990) DNA typing in the forensic arena. Am J Hum Genet 47:876–877

Jones v Murray, 962 F 2d 302 (4th Cir 1992)

Koehler J (1993) Error and exaggeration in the presentation of DNA evidence at trial. Jurimetrics 34:21–26

Lempert R (1991) Some caveats concerning DNA as criminal identification evidence: with thanks to the Reverend Bayes. Cardozo Law Rev 13:303, 325

Lewontin R, Hartl D (1991) Population genetics in forensic DNA typing. Science 254:1735, 1749

McEwen JE, Reilly PR (1994) A review of state legislation on DNA forensic data banking. Am J Hum Genet 54:941–958 (in this issue)

Mueller L (1993) The use of DNA typing in forensic science. Accountability Res 3:1–4

National Academy of Sciences, National Research Council, Commission on Life Sciences, Board of Biology. (1992) DNA technology in forensic science. National Academy of Sciences, Washington DC

Ryncarz v Eikenberry, 824 F Supp 1493 (ED Washington 1993)

Scheck B (1994) DNA and Daubert. Cardozo Law Rev 15:1201–1239

Thompson W (1993) Evaluating the admissibility of new genetic identification tests: lessons from the "DNA war." J Criminal Law Criminol 84:22, 91–92

Weir B (1992) Population genetics in the forensic DNA debate, Proc Natl Acad Sci USA 89:11654, 11658