

The molecular reinscription of race: unanticipated issues in biotechnology and forensic science

TROY DUSTER

ABSTRACT In the last five years, there has been a considerable increase in the number of published articles documenting the health disparities between the majority white population and various other racially designated groups in the United States. This was a direct result of the congressional mandate given to the National Institutes of Health in 2000 to conduct research and report findings on the topic. It was inevitable that reports of patterned disparities between racially designated groups would resuscitate an old debate about whether there were important biological differences between such groups. There is now a new role in this debate for biotechnology firms committed to finding markets for pharmaceutical products that have failed to get past clinical trials aimed at the general population. By refocusing marketing strategy on racially designated populations, the industry has gathered support for medicines that were previously 'racially neutral'. And this strategy has in turn inspired 'strange bedfellow' support from clinicians who claim selective advantage for their patients from such 'racialized medicine'. There has been a parallel development in forensic science. There are new claims that DNA analysis of crime scene data will assist criminal investigations by narrowing the search for suspects along racial lines. Because these tests are privately conducted, it is not possible for critics to assess the sampling procedure or the overall methodology that is used in categorizing human populations by race. Nonetheless, these claims dovetail with those in pharmacogenomics that assert the importance of patterns in the DNA for 'predicting' ethnic and racial membership. In sum, these developments are ushering in an era of the molecular reinscription of race in the biological sciences.

KEYWORDS biobanks, DNA databanks, forensic science, human genome, health disparities, race, racial medicine

No one could have predicted that the Defense Department's commissioning of a secure high-speed communication network for military use would 'spin off' into the World Wide Web of the contemporary Internet. While the Defense Department achieved its goal, that achievement has been dwarfed by the scale of the social, economic and political consequences of the way the Internet has developed. From daily commercial transactions to downloads of streaming videos, from distance learning to search engines that can mine the Library of Congress, from e-mail to news media clips to blogs, the list could go on for several pages. The Internet is such a dominant

feature of the lives of so many that, on a worldwide scale, if banks or hotels or restaurants do not have their own websites, they are consigned to outlier status in the backwaters of an integrated global network.

No one could have predicted the speed and penetration of these developments. But, if we had stopped and thought about it, we might have expected dramatic changes were the known synergistic ingredients to the infrastructure to be placed on to the conveyor belt of modern life. In a parallel fashion, one of the deeply consequential spin-offs of the mapping and sequencing of the entire human genome has been the use of these technologies to identify individuals at the level of the microchip. This capacity for the identification of millions, even billions, across the globe is going to produce spin-offs that will ultimately dwarf the original intentions of the early advocates of a genome map. The first of those is just now emerging: the subtle, sometimes inadvertent, reinscription of race at the molecular level. A second development—the use of markers for individual identification and for claims to ‘authenticity’ of group membership—has already penetrated deep into the criminal justice system by way of forensic applications. This includes the use of DNA in post-conviction cases to determine whether there has been a wrongful conviction, the kind of situation that might result in freeing the innocent. Another use involves the collection of DNA from suspects or arrestees in pretrial circumstances to increase the DNA database, which in turn is designed to help law enforcement officers find matches between the DNA samples of those suspects or arrestees and tissue samples left at some unsolved crime scene: a net to catch the guilty. This is a long way from the original reason for mapping the genome.

The rationale behind the Human Genome Project, and for the Haplotype Map Project that followed,¹ has always been the search for ways to improve our health.² There have already been some health benefits, and there will certainly be more. Nonetheless, in this paper I will point to mounting evidence that the inadvertent and unintended spin-offs (into domains far removed from health concerns and clinical medical applications) of the revolution in human molecular biology will dwarf the medical achievements.

In the last five years, there has been a peculiar and fateful irony in the convergence of the desire (and pressure) to use genetics to improve our health, and the decision by the US Congress to require that the National Institutes of Health record data and engage in research to lessen the health disparities between racial and ethnic groups. In 2000 Congress passed the

1 International HapMap Consortium, ‘The International HapMap Project’, *Nature*, vol. 426, 18 December 2003, 789–96.

2 The rationale for the Icelandic data base, for the Howard University database and for the Estonian database is also in each case explicitly about strategies for improving health.

Minority Health and Health Disparities Research and Education Act of 2000 (Public Law 106-525), which mandated the National Institutes of Health to support research on health disparities between groups categorized by race and ethnicity.

80 As a direct consequence, the years since have seen a sharp increase in the number of articles that report health disparities between members of the majority white population and the various groups racially and ethnically designated. That was to be expected. Moreover, since the National Human Genome Research Institute is a branch of the National Institutes of Health, it would follow that research on human genetics would enter the fray, with
85 scientists poised and ready to assert the unique contribution of molecular genetic differences to an explanation of these health disparities. For example, because the rate of prostate cancer in African Americans is more than double that in white Americans, it was inevitable that some would attempt to explain this through the lens of genetics. This in turn would lead the way to
90 rescuing old racial taxonomies and their relationship to genetic profiles and genetic conditions. It was not expected that this strategy would thereby inadvertently resuscitate the idea that genetic differences between those we place in racial categories might well explain different health outcomes. This is territory full of minefields for obvious reasons that date back to the eugenics movement in the United States and its promulgation and extension into Nazi Germany.³

To navigate around such problems, research scientists have developed two strategies. The first is to use only an *ex post facto* deployment of race and ethnicity, after data have been collected in large clinical trials in which race
100 had not been a defining category of selection. Two recent examples of this are the marketing of the racialized drugs BiDil (for relief of hypertension and congestive heart failure amelioration in African Americans) and Iressa (a late-stage drug for treatment of lung cancer in Asians). The second strategy for avoiding the explosive combination of race-genetics-disease is to propose
105 a large-scale research project aimed at determining the genetic *v.* environmental causes of health problems. Of course this proposal presumes that one can successfully disentangle the genetic from the environmental. It is one of the most fundamental axioms of contemporary life sciences that there is an
110 interactional element at the core of the relationship between genes and environment. To assert that some human health condition, trait or behaviour is, say, 60 per cent genetic and 40 per cent environmental is to express a static

3 Philip R. Reilly, *The Surgical Solution: A History of Involuntary Sterilization in the United States* (Baltimore: Johns Hopkins University Press 1991); Stefan Kühl, *The Nazi Connection: Eugenics, American Racism, and German National Socialism* (Oxford and New York: Oxford University Press 1994); Robert N. Proctor, *The Nazi War on Cancer* (Princeton, NJ: Princeton University Press 1999).

and archaic version: a statistical artefact of the static assumption.⁴ Such a statistical 'parcelling out' denies or obscures the profoundly interactional relationship between race, genes and disease, now generating one of the most vexing and visceral debates in contemporary science. In order to understand why, we must first try to understand what it means to 'isolate' race as a variable (or constant) in a research protocol. The first part of this paper addresses the subtle reinscription into the biological sciences and clinical medicine of an old notion, pre-dating molecular biology, of human taxonomies of race. The second part follows on from a discussion of the population-specific DNA databases to address their forensic use as markers for identification.

Navigating with (and around) race

Biotechnology firms have found an unusual and effective way around the problem of confronting 'race' as a 'biological category'. Their strategy is *not* to deal with race by means of a full-scale, case-control research design, but to 'back into' a clinical study that was never designed to test whether race played any role only to discover *ex post facto* that the race of the clinical population, however defined, *did* play a role in drug efficacy. The reinterpretation by racial categories of already collected data sets conveniently circumnavigates the problem of having to define terms. After all, to do a case-control study would require the researcher to define terms and to specify the boundaries of the relevant populations. For 'race', this would be a knotty problem these days when the primary criterion for racial classification is self-reporting.

BiDil and the medicalization of the sources of hypertension

BiDil is a combination of isosorbide dinitrate and hydralazine, designed to restore low or depleted nitric oxide levels in the blood to treat or prevent cases of congestive heart failure. It was originally designed for a wide population base; race was irrelevant. But the early clinical studies revealed no compelling results, and a Food and Drug Administration (FDA) advisory panel voted 9 to 3 against approval.

In a remarkable turn of fate, however, BiDil was suddenly born again as a racialized intervention. One of the investigators went back into the data, and found that African Americans in the original clinical trial seemed to fare better than Whites. Because the study was not designed to test that hypothesis, a new clinical trial would have to be approved. However, rather than setting up a study designed to see whether BiDil worked better in one

4 Richard C. Lewontin, 'Analysis of variance and analysis of causes', *American Journal of Human Genetics*, vol. 26, no. 3, 1974, 400–11.

150 group than in another, in March 2001 the FDA approved a full-scale clinical trial, the first to be conducted exclusively in black men and women suffering from heart failure.

155 The BiDil Blacks-only trial reflects two problematic assumptions about race and medicine. The first is that African Americans' risk of developing and dying from heart failure is twice that of Whites. This claim has been floating around in the scientific literature for more than a decade, mainly uncontested. Jonathan Kahn recently completed some remarkable scientific sleuthing on the topic, however, and showed that the claim by NitroMed, the company that developed BiDil, about the scale of black and white differences is simply untrue. Kahn traced the citation sources back nearly 160 two decades, and demonstrated conclusively that the difference between Blacks and Whites is actually closer to 1.2:1.⁵ There is a difference, but it is nowhere near the 2:1 ratio that would warrant special trials for one population group. Thus, a substantial section of the scaffolding of the BiDil clinical trial is built on incorrect statistical data about racial disparities. The 165 second claim is that BiDil has a special effect that is greater in African Americans than in Whites.⁶ The clinical trials were not designed to test that hypothesis. Rather, by concentrating only on Blacks, the study can have little or nothing compelling to say about comparative results, by race.

170 In the early spring of 2005, anticipating the FDA decision to approve the drug in late spring, NitroMed released a statement that was an attempt to provide a justification for the approval of the drug as race-specific: 'The African American community is affected at a greater frequency by heart failure than the corresponding Caucasian population. African Americans 175 between the ages of 45 and 64 are 2.5 times more likely to die from heart failure than Caucasians in the same age range.'⁷ The figures are technically correct, but the age-group 45–64 only accounts for about 6 per cent of heart failure mortality, while those over 65 constitute 93.7 per cent of the mortality. Moreover, over 65, the statistical differences between 'African Americans' and 'Caucasians' almost completely disappear. Why the focus on race? What 180 is at stake and why is NitroMed so invested in selectively portraying these sharp racial differences? Part of the answer lies in the role of prospective markets for biotech products. While the new mantra of biotechnology is the claim that pharmaceuticals will someday soon be marketed to individuals

5 Jonathan Kahn, 'How a drug becomes "ethnic": law, commerce, and the production of racial categories in medicine', *Yale Journal of Health Policy, Law, and Ethics*, vol. 4, no. 1, Winter 2004, 1–46.

6 Anne L. Taylor, Susan Ziesche, Clyde Yancy, Peter Carson *et al.*, 'Combination of isosorbide dinitrate and hydralazine in Blacks with heart failure', *New England Journal of Medicine*, vol. 351, no. 20, 11 November 2004, 2049–57.

7 NitroMed press release, 'FDA accepts NitroMed's new drug application resubmission for BiDil', 2 February 2005, available at <http://investors.nitromed.com/phoenix.zhtml?c=130535&p=irol-newsArticle&ID=670434&highlight=> (viewed 8 August 2006).

185 based on their DNA, the fundamental truth is that selling drugs is about
markets. These markets are not about individual designer drugs, but about
groups and population aggregates that become target markets.

190 The original BiDil patent was not race-specific, but it expires in 2007. By
making the drug race-specific, the patent will be extended another thirteen
years to 2020. This gives the company NitroMed the exclusive right to
market the drug as a therapy for heart conditions. This whole process has
succeeded in biologizing and medicalizing 'race' in ways that could have
195 been easily avoided . . . if market forces were not dominating the decision-
making. By simply testing generic drug combinations of hydralazine and
isosorbide dinitrate in comparison populations, the question of drug efficacy
would have been answered. But such a strategy would have undermined the
marketing monopoly and would have, instead, brought relatively cheap
200 drugs to the market. This strategy would have 'saved lives' in a more
comprehensive and race-neutral way, in sharp contrast to the rhetoric being
deployed by those marketing BiDil, who say that, regardless of the socio-
political aspects of marketing a drug that is race-specific, the drug 'saves
lives'.

205 In a classic piece of epidemiological research, Michael Klag and his
associates showed a decade ago that, in general, the darker the skin colour,
the higher the rate of hypertension for American Blacks, even within the
African American community.⁸ Klag *et al.* indicated that this was not
biological or genetic in origin, but biological in effect, due to stress-related
outcomes of reduced access to valued social goods such as employment,
210 promotion, housing stock and so on. The effect was biological, not the
cause.

Iressa, lung cancer and Asians

215 The British/Swedish firm AstraZeneca conducted a large clinical trial across
the full population of patients suffering from advanced stages of lung cancer
to test the efficacy of Iressa, one of its most promising drugs. The drug was
first introduced in 2002, and over 45,000 patients worldwide have taken it. It
is designed to treat the most common form of lung cancer by blocking an
enzyme that otherwise causes cancer cells to proliferate. In the full study, the
difference between those on placebo and those taking Iressa was statistically
insignificant. When the results were announced, the FDA began reviewing
220 the situation to determine whether the drug should be pulled from the
market. However, in viewing the data by race and ethnicity, it was
determined that, while the average number of months that life was
prolonged was 5.5 for the general population, Asians had a 9.5-month

8 Michael Klag, P. K. Whelton, J. Coresh, C. E. Grim and L. H. Kuller, 'The association of skin color with blood pressure in US Blacks with low socioeconomic status', *Journal of the American Medical Association*, vol. 265, no. 5, February 1991, 599–602.

225 prolongation, nearly twice as long. Immediately, AstraZeneca touted the findings as significant, and began preparations for shifting their marketing strategies and sales to Asian countries.⁹

230 Two elements have now combined in a synergistic way to set the stage for the 'hurricane just offshore' that is waiting for the winds to change. The first concerns the new diagnostic technologies of bio-informatics, computer-assisted analyses of patterns of DNA. The second concerns new research on specific populations and allele-frequency markers for those populations that just happen occasionally to coincide with the socially defined groupings we used to call 'ethnic' or 'racial', but are now increasingly referred to as groups aggregated by 'population-specific allelic frequency patterns'. With the latest computers, we can now put the DNA of several clusters of people on computer chips, and see what patterns in their DNA might emerge.

240 We can perform hundreds, even thousands, of such experiments in a few hours. Sometimes this proves to be a useful technology in the hunt for particular regions in order to help explain some illnesses. For example, if we get a few hundred patients, all with prostate cancer, then we can look for patterns in their SNP profiles using the chip technology.¹⁰ With the new SNP profiles, it is possible for a researcher—by simply sorting those affected, and not driven by any theoretical question—to come up with new taxonomies of people who share certain patterns in their DNA. That is the way much of clinical genetics tries to find 'candidate genes' for a particular condition, say, prostate cancer, among those who suffer from it. But what happens when rates of prostate cancer differ significantly between groups that have been racially designated? This way of framing the problem leaves one vulnerable to making a profoundly subtle interpretive error, namely, that, by holding other factors constant statistically (such as income, education, class etc.) and finding persistent racial differences, one has substantial grounds for concluding that race is more biological than social and political. The following example provides an illustration of how this error is made, both in the natural sciences and the social sciences.

The interpretive error: controlling for other factors and the biology of race

In 1986 the Drug Enforcement Administration initiated Operation Pipeline, a programme designed in Washington, D.C. that ultimately trained 27,000

9 Nicholas Zamiska and Jeanne Whalen, 'Cancer drug, deemed failure, helps Asians', *Wall Street Journal*, 5 May 2005, B1.

10 Single nucleotide polymorphisms (SNPs) are DNA sequence variations that occur when a single nucleotide (A, T, C or G) in the genome sequence differs between members of a species (or between paired chromosomes in an individual). For example a SNP might change the DNA sequence AAGGCTAA to ATGGCTAA. SNPs make up about 90 per cent of all human genetic variation.

434 *Patterns of Prejudice*

260 law enforcement officers in forty-eight participating states over the ensu-
ing decade. The project was designed to alert police and other law
enforcement officials as to the 'likely profiles' of those who should be
stopped and searched for possible drug violations. High on the list were
265 young, male African Americans and Latinos driving in cars that signal-
led that something might be amiss. For example, a nineteen-year-old
African American driving a new Lexus would sound an 'obvious' alarm
because the assumption is that the family could not afford such a car, and
that the driver must therefore be 'into drugs'. In a study of the I-95
270 highway corridor just outside of Baltimore, police records indicated that
Latinos and Blacks were eight times more likely to be stopped by the
police than Whites.¹¹ A young white male in his early twenties driving
a Lexus would not be considered 'suspicious', since the police are more
likely to assume that he is driving the family car. So, if class is statisti-
cally held constant, and a racial difference is still found, it would be a
massive interpretive error to conclude that, since race is still a factor, it
275 must be operating as a biological variable. The same would be true for the
study of a medical condition, say, hypertension. If the researcher has
statistically controlled for social class and race still shows up as a risk
factor, it does not mean that the social meaning of race is not present.
Middle-class African Americans are subject to more stressors than middle-
280 class Whites, and that is a social reality that may have biological outcomes,
not origins.¹²

The segue to forensics, criminal justice and 'molecular race'

285 It is possible to make arbitrary groupings of populations—defined by
geography, language, self-identified faiths, other-identified physiognomy
and so on—and still find statistically significant allelic variations between
those groupings. For example, we could simply pick all of the people in
Chicago, and all of the people in Los Angeles, and find statistically
significant differences in allele frequency at *some* loci. Of course, at many
loci, even most loci, we would not find statistically significant differences.
290 When researchers claim to be able to assign people to groups based on allele
frequency at a certain number of loci, they have chosen loci that show
differences between the groups they are trying to distinguish.

11 Troy Duster, 'Selective arrests, an ever-expanding DNA forensic database, and the specter of an early twenty-first century equivalent of phrenology', in David Lazer (ed.), *DNA and the Criminal Justice System: The Technology of Justice* (Cambridge, MA: MIT Press 2004), 315–34.

12 Raymond Franklin, *Shadows of Race and Class* (Minneapolis: University of Minnesota Press 1991).

295 The work of Ian Evett *et al.*, Alex Lowe *et al.* and others suggests that only
about 10 per cent of DNA sites are 'useful' for making distinctions.¹³ This
means that, at the other 90 per cent of the sites, the allele frequencies do not
vary between groups such as 'Afro-Caribbean people in England' and
300 'Scottish people in England'. But it does not follow that, because we cannot
find a single site at which allele frequency matches some phenotype—we are
trying to make identifications for forensic purposes, we should be
reminded—there are not several (four, six, seven) sites that could be effective
in aiding the FBI, Scotland Yard or the criminal justice systems around
the globe in formulating highly probable profiles of suspects, and the
likely ethnic, racial or cultural populations in which they can be found ...
statistically.

305 So, when molecular biologists assert that 'race has no validity as a
scientific concept', this is apparently in contradistinction to the practical
applicability of research on allele frequencies in specific populations. It is
possible to sort out and make sense of this—and even to explain and resolve
the apparent contradiction—but only if we keep in mind the difference
310 between using a taxonomic system with sharp, discrete, definitively
bounded categories, and one that shows only patterns (with some overlap)
but that may prove to be empirically or practically useful.

When representative spokespersons from the biological sciences say that
315 'there is no such thing as race', they mean, correctly, that there are no discrete
racial categories that come to a discrete beginning and end, that there is
nothing mutually exclusive about our current (or past) categories of 'race',
and that there is more genetic variation within categories of 'race' than
between them. All this is true. However, when Scotland Yard or the
Birmingham police force or the New York Police Department wants to
320 narrow the list of suspects in a crime, they are not primarily concerned with
tight taxonomic systems of classification with no overlapping categories.
That is the stuff of theoretical physics and philosophical logic, not the
practical stuff of crime-solving or the practical application of molecular
genetics for health delivery via genetic screening, and all the messy
325 overlapping categories that will inevitably be involved with such enter-
prises. That is, some African Americans have cystic fibrosis even though the
likelihood of that is far greater among Americans of North European descent
and, in a parallel if not symmetrical way, some American Whites have sickle
cell anaemia even though the likelihood of that is far greater among

13 I. W. Evett, J. S. Buckleton, A. Raymond and H. Roberts, 'The evidential value of DNA profiles', *Journal of the Forensic Science Society*, vol. 33, no. 4, 1993, 243–4; I. W. Evett, P. D. Gill, J. K. Scranage and B. S. Weir, 'Establishing the robustness of Short-Tandem-Repeat statistics for forensic application', *American Journal of Human Genetics*, vol. 58, 1996, 398–407; Alex L. Lowe, Andrew Urquhart, Lindsey A. Foreman and Ian W. Evett, 'Inferring ethnic origin by means of an STR profile', *Forensic Science International*, vol. 119, no. 1, 2001, 17–22.

330 Americans of West African descent. But in the world of cost-effective
decision-making, genetic screening for these disorders is routinely based on
335 commonsense versions of the phenotype. The same is true with regard to the
quite practical matter of naming suspects.

In an article in the 8 July 1995 issue of *New Scientist* entitled 'Genes in Black
and White', some extraordinary claims are made about what it is possible to
335 learn about socially defined categories of 'race' from reviewing information
gathered using new molecular genetic technology.¹⁴ In 1993 a British forensic
scientist published what is perhaps the first DNA test explicitly acknowledged
to provide 'intelligence information' along 'ethnic' lines for 'investigators of
340 unsolved crimes'.¹⁵ Ian Evett, of the Home Office's forensic science laboratory
in Birmingham, and his colleagues in the Metropolitan Police claimed that
their DNA test could distinguish between 'Caucasians' and 'Afro-Caribbeans'
in nearly 85 per cent of cases. Evett's work, published in the *Journal of the
Forensic Science Society*, draws on apparent genetic differences in three sections
345 of human DNA. Like most stretches of human DNA used for forensic typing,
each of these three regions differs widely from person to person, irrespective of
race. But, by looking at all three, the researchers claimed that under select
circumstances it was possible to estimate the probability that someone
belonged to a particular racial group. The implications of this for determining,
350 for practical purposes, who is and who is not 'officially' a member of some
racial or ethnic category are profound.

The legal and social applications of these technologies are already in
considerable use by the *cognoscenti*, and they are poised to 'take off' even
355 further. For example, more than a decade ago, several states began keeping
DNA database files on sexual offenders. Three factors converged to make
this a popular decision by criminal justice officials and one that would be
backed by politicians and the public: 1) sex offenders are those most likely to
leave bodily tissue and fluids at the crime scene; 2) they rank among the
360 most likely repeat offenders; and 3) their crimes are often particularly
reprehensible in that they violate persons, from rape to molestation and
abuse of the young and most vulnerable. Today, all fifty states store DNA
samples of sex offenders, and most states do the same for convicted
murderers. However, thirty-four states now store DNA samples of all
365 felons.¹⁶ Since 1 September 1999, the state of Louisiana has taken a DNA
sample of all those merely arrested for a felony.¹⁷ In compliance with its 1998

14 Gail Vines, 'Genes in black and white', *New Scientist*, no. 1985, 8 July 1995.

15 Ian W. Evett, 'Criminalistics: the future of expertise', *Journal of the Forensic Science Society*, vol. 33, no. 3, 1993, 173-8.

16 Tania Simoncelli, 'Dangerous excursions: the case against expanding forensic DNA databases to innocent persons', *Journal of Law, Medicine and Ethics*, vol. 34, no. 2, 2006, 390-7.

17 Paul E. Tracy and Vincent Morgan, 'Big Brother and his science kit: DNA databases for 21st century crime control', *Journal of Criminal Law and Criminology*, vol. 90, no. 2, 2000, 635-90 (682).

state law, Virginia now has a website listing its registered sex offenders.¹⁸ This has been followed by websites for Indiana, Florida and Alaska. So, what started off as concern with sexual offences has now widened to include many other felonies.

Function creep: from Social Security to DNA databanks

Social Security in the United States was originally intended only as a federal retirement programme. The entire scheme was vigorously debated in the 1930s, and some members of Congress argued vehemently that the Social Security identification card should not be a national identification card. The vote in the Congress was close. However, the Internal Revenue Service (IRS) soon began using the social security number to track citizens for tax-collection purposes. It was later used by many public services as a required personal identification number and, still later, private institutions began to demand it for identification purposes. This process is called 'function creep': the process by which the original function may remain, but newer uses expand into ever-widening spheres.

Since it is the state that is primarily involved in criminal law enforcement, there have been state-by-state variations in the use of DNA databanks and storage. Fifteen years ago, most states were only collecting DNA samples from sexual offenders. Now, seven states require the DNA databanking of *all* felons, including those involved in white-collar felonies. When Governor Pataki proposed such a scheme for New York state, the state legislature forced him to jettison the idea. Louisiana was the first state to pass a law mandating the collection of DNA samples from anyone arrested for a felony but, in the past five years, it has been joined by Virginia and several other states. That is, just like 'function creep' with regard to Social Security identification numbers, the use by states of DNA databanks is rapidly expanding.

Today, thirty-eight states store samples from varying categories of lawbreakers in a DNA databank. Twenty-nine states now require that tissue samples be retained in their DNA databanks after profiling is complete.¹⁹ Only one state, Wisconsin, requires the destruction of tissue samples once profiling is complete. What started as a tool for dealing with sex offenders has now 'crept' into a way to deal with lawbreakers and those merely arrested.

18 See the 'Sex offenders and crimes against minors registry' on the Virginia State Police website, at <http://sex-offender.vsp.virginia.gov/sor/index.htm> (viewed 16 August 2006).

19 Jonathan Kimmelman, 'Risking ethical insolvency: a survey of trends in criminal DNA databanking', *Journal of Law, Medicine and Ethics*, vol. 28, no. 3, 2000, 209–21 (211).

438 *Patterns of Prejudice*

While thirty-nine states permit the expunging of samples if charges are dropped, almost all of those states place the burden of initiating the destruction of samples on the individual. Thus, civil privacy protection, which in the default position places the burden on the state, is reversed. With the reauthorization of the 2005 Violence against Women Act, the burden for expunging samples shifts to the individual who has been placed in the federal CODIS (Combined DNA Index System) database.

Twenty states authorize the use of databanks for research on forensic techniques. Based on the statutory language in several of those states, this could easily mean assaying genes or loci that contain predictive information, even though current usage is restricted to analysing portions of the DNA that are only useful as identifying markers. Since most states retain the full DNA (and every cell contains all the DNA information), it is a small step to using these DNA banks for other purposes. The original purpose has long been pushed to the background, and the 'creep' extends not only to crimes other than sexual offenses, but to misdemeanours and even arrestees.

On 5 January 2006, the President of the United States signed into law HR 3402, the Department of Justice Reauthorization of the Violence against Women Act of 2005. For the first time this legislation gives state and federal law enforcement officials the right to enter DNA profiles of those merely arrested for federal crimes into the CODIS database. Previously, only convicted felons could be included. Those DNA profiles will remain in the database unless and until those individuals who are exonerated or never charged with the crime request that their DNA be expunged. Thus the default position will be to store these profiles, and expunging requires the proactive agency of those arrested. This announcement was a cause for celebration by one of the leading providers of DNS testing services, Orchid Cellmark, Inc., of Princeton, New Jersey. The president and chief executive officer of Cellmark, Paul J. Kelly, immediately issued a statement applauding this development, stating:

This is landmark legislation that we believe has the potential to greatly expand the utility of DNA testing to help prevent as well as solve crime . . . It has been shown that many perpetrators of minor offenses graduate to more violent crimes, and we believe that this new legislation is a critical step in further harnessing the power of DNA to apprehend criminals much sooner and far more effectively than is possible today.²⁰

20 'Orchid Cellmark hails passage of landmark legislation expanding use of forensic DNA identity testing in the U.S.', press release, 6 January 2006, available at www.orchid.com/news/view_pr.asp?ID=393 (viewed 17 August 2006).

440 Britain has been in the vanguard of these developments, but there is every
indication that this will not be so for long.²¹ In April 2004 a law was passed
in the United Kingdom permitting police to retain samples from anyone
arrested for any reason, including people who are not charged with a crime.
Anyone can have their DNA taken and stored. The database already
445 contains 2.8 million DNA 'fingerprints' taken from identified suspects,
plus another 230,000 from unidentified samples collected from crime
scenes.²² Samples are being added at the rate of between 10,000 and
20,000 per month.²³ The aim is to have on file a quarter of the adult
population's DNA, a figure that exceeds ten million, making it by far the
450 largest DNA database in the world. In keeping with the 'racialization' of
DNA forensic databases, it was recently disclosed that nearly 4 in 10 Blacks
are in the UK forensic database, compared with fewer than 1 in 10 Whites.²⁴
Perhaps the most striking statistic comes from the fact that police record 'the
ethnic appearance' of each person placed in the database: 82 per cent of male
profiles are white, while only 7 per cent are black, according to the Home
Office. Thus, comparing the proportion of each racial grouping in the
455 database against their proportion of the whole population reveals the sharp
disparity. Because many of these individuals are involved only in arrests,
civil libertarians have wondered out loud about the implications for the
presumption of innocence. Dominic Bascombe, a writer for a London-based
newspaper (*The Voice*), expressed his concern that these data reflected the
460 greater likelihood that Blacks face arrest: 'It is simply presuming if you are
black you are going to be guilty—if not now but in the future.' He then noted
that this constituted what he called 'genetic surveillance' of Blacks: 'We
certainly don't think it reflects criminality'.²⁵ 'Anyone in the database—and
family members—can more easily be linked to a crime scene if their DNA is
465 found there. This may be because they are a criminal, or because they visited
the scene prior to the crime.'²⁶

21 The British may be in the lead, but the Portuguese have even bigger plans: in early April 2005, the Portuguese government announced that it intended to collect DNA on all of its residents, all of its inhabitants. Maria João Boavida, 'Portugal plans a forensic genetic database of its entire population', *Newropeans Magazine* (online journal), 8 April 2005, at www.newropeans-magazine.org/index.php?option=com_content&task=view&id=2059&Itemid=121 (viewed 17 August 2006).

22 'Police aim for DNA on 25% of UK population', *British Journal of Healthcare Computing & Information Magazine*, February 2005, abstract at www.bjhc.co.uk/news/1/2005/n502020.htm (accessed 17 August 2006).

23 And this was before the bombings in London on 7 July 2005.

24 James Randerson, 'DNA of 37% of black men held by police', *Guardian*, 5 January 2006, 1.

25 Quoted in Randerson, 'DNA of 37% of black men held by police'.

26 Randerson, 'DNA of 37% of black men held by police'.

440 *Patterns of Prejudice*

470 The New York Police Department, with the urging and support of Mayor
Rudolph Giuliani, was chafing to make use of a portable DNA lab kit.²⁷ At
the time this was being proposed, in 2000, New York Police Chief Howard
Safir said that a DNA sample should be taken from 'anyone who is arrested
for anything'.²⁸ This lab kit, the convergence of molecular genetics and
contemporary concerns, was developed by the Whitehead Institute for
Biomedical Research. By the year 2010 it is likely that we will be using these
technologies far more in relation to forensics and the law than to medicine
475 and therapies. California passed a ballot proposition in 2004 that permits, by
2008, the collection and storage of DNA data not only on all felons, but on
arrestees and those committing some categories of misdemeanours.

480 New York state collects fingerprints from persons receiving public
assistance (to prevent duplicate claims) and has done so since 1994, but
does not generally share those data with law enforcement. In that state, the
record of one's arrest must be sealed, and employers cannot ask if someone
has ever been arrested. However, an employer can ask if someone has been
convicted. When a mere arrest triggers DNA sampling and storage, when
those samples are not expunged, the record becomes analogous to radio-
485 active storage in that it keeps emitting its effect indefinitely.²⁹

Population-wide DNA databases

490 It is now relatively common for scholars to acknowledge the considerable,
and documented, racial and ethnic bias in police procedures, prosecutorial
discretion, jury selection and sentencing practices—of which racial profiling
is but the tip of an iceberg.³⁰ Indeed, racial disparities penetrate the whole
system and are diffused throughout it, all the way up to the racial disparities
involved in seeking the death penalty. If the DNA database primarily
includes those who have been touched by the criminal justice system, and
that system operates practices that routinely select more from one group
495 than another, there will be an obvious skew or bias towards this group. Some
have argued that the way to handle the racial bias in the DNA database is to
include everyone.

500 But this does not address the far more fundamental problem of the bias
that generates the configuration and content of the criminal (or suspect)
database. If the lens of the criminal justice system is focused almost entirely
on one part of the population for a certain kind of activity (drug-related
street crime), and ignores a parallel kind of crime (fraternity cocaine sales a

27 Ayana Mathis, 'Stop, drop, and swab: New York police ponder portable DNA labs',
Village Voice, 6 June 2000, 26.

28 Tracy and Morgan, 'Big Brother and his science kit', 665.

29 Kimmelman, 'Risking ethical insolvency'.

30 Marc Mauer, *Race to Incarcerate* (New York: New Press 1999).

505 few miles away), then, even if the fraternity members' DNA is in the databank, they will not be subject to the same level of matching or of subsequent allele frequency profiling research to 'help explain' their behaviour. *That behaviour will not have been recorded.* That is, if the police are not arresting the fraternity members, it does not matter whether their DNA is in a national database because they are not *criminalized* by the selective targetting of the criminal justice system.

510 Thus, it is imperative that we separate arguments about bias in the criminal justice system at the point of contact with select parts of the population from 'solutions' to bias about 'cold hits'. It is certainly true that, if a member of the fraternity committed a rape, left tissue samples at the scene and—because he was in a national DNA database—was nabbed by the police with a 'cold hit', that would be a source of justifiable affirmation. But, by ignoring powder cocaine and emphasizing street sales of cocaine in the African American community, the mark of criminality is thereby generated, and this is not altered by having a population-wide DNA database.

520 However, the surface fiction of objectivity will lead to a research agenda concerning DNA. There is a serious threat regarding how these new technologies are about to be deployed that is masked by the apparent global objectivity of a population-wide DNA database. I am referring to the prospects for SNP profiling of offenders. As noted, even if everyone were in the national database, this would not deter the impulse to do specific and focused research on the select population that has been convicted.

530 **Troy Duster** is Director of the Institute for the History of the Production of Knowledge and Silver Professor of Sociology at New York University. His most recent books are *Backdoor to Eugenics* (2nd edn, 2003) and (with Michael K. Brown *et al.*) *Whitewashing Race: The Myth of a Colorblind Society* (2003).