No. 12-207

In The
Supreme Court of the United States

MARYLAND,

Petitioner,

v.

ALONZO JAY KING, JR.,

Respondent.

On Writ Of Certiorari To The Court Of Appeals Of Maryland

BRIEF OF GENETICS, GENOMICS AND FORENSIC SCIENCE RESEARCHERS AS *AMICI CURIAE* IN SUPPORT OF NEITHER PARTY

MICHAEL L. FOREMAN
Counsel of Record
The Pennsylvania State University
School of Law
329 Innovation Boulevard, Suite 118
State College, PA 16803
(814) 865-3832
mlf25@psu.edu

December 28, 2012
Counsel for Amici Curiae

COCKLE LAW BRIEF PRINTING CO. (800) 225-6964
OR CALL COLLECT (402) 342-2831

For questions or comments about this brief, please contact David Kaye, Pennsylvania State University School of Law, University Park PA 16802, 814 865-8974, kaye@psu.edu

TABLE OF CONTENTS

TABLE OF CONTENTS – Continued

TABLE OF CONTENTS – Continued

TABLE OF AUTHORITIES

TABLE OF AUTHORITIES – Continued

TABLE OF AUTHORITIES – Continued

Page

TABLE OF AUTHORITIES – Continued

Page

TABLE OF AUTHORITIES – Continued

Page

INTEREST OF *AMICI CURIAE*

Amici curiae are ten physicians, human geneticists, statistical geneticists, molecular biologists, or other researchers in human genetics or forensic science.[1] They have conducted research at some of the world's leading medical schools, universities, and government laboratories. They include members of the Institute of Medicine, of committees of the National Academy of Sciences and the National Institutes of Health, and the editorial boards of numerous genetics journals. *See* Appendix A.

Amici study the nature and transmission of genetic information. The issue before this Court involves a special kind of genetic information that has proven to be of great value in criminal justice. *Amici* submit this brief, on behalf of neither party, to inform the Court of the possible medical and social significance of the DNA data stored in law enforcement databases.

[1] Letters consenting to the filing of this brief have been filed with the Clerk of the Court. No counsel for a party authored this brief in whole or in part, and no person, other than *amici* or their counsel, made any monetary contribution to its preparation or submission. No *amicus* is employed by any law enforcement agency or any company that has any commercial interest in law enforcement databases. One *amicus* served as a senior scientist for the FBI and was instrumental in the creation of the CODIS database system. Every *amicus* participated in preparing this brief.

SUMMARY OF ARGUMENT

This case concerns "genetic information" – information passed from parent to child according to the laws of genetics. But not all genetic information is equally important – biologically, medically, or socially. Unlike medical genetic tests, law enforcement identification profiles have no known value for medical diagnosis or prediction of future health.

The genetic information stored in U.S. databases is derived from what are called the "CODIS loci." Despite rapidly growing knowledge of the genome, no CODIS locus has been found to cause any genetic disease or trait. Furthermore, the scientific literature does not show that the loci are even moderately correlated with disease status, physical traits, or behavioral predispositions. Although no one can say with certainty what the future will bring and it is possible that specific loci will be found to affect the operation of certain genes or to display correlations to disease states, it is unlikely that the identification profiles will turn into powerful medical diagnostic or predictive tools that can be used to infer disease states or predispositions by examining forensic database records.

That said, the records have other actual or possible uses beyond individual identification. They can be used to establish genetic family relationships (parent-child, sibling, etc.). In addition, they could be used to make rough inferences about geographic or ethnic

origins, although there are far more accurate genetic indicators of this kind of ancestry.

Recent research, notably the ENCODE project, has given rise to speculation in the press that all DNA sequences affect development and health. This is not true. Although the research lowers the percentage of the genome that might colloquially be called "junk," it does not support any use of the CODIS markers to predict genetic diseases or traits. Estimates of the proportion of the genome that is "junk" are only vaguely related to the medical and social relevance of the CODIS markers. Published studies do not establish that the database records are likely to become valuable for medical diagnosis or prediction. The ENCODE data should enhance knowledge of molecular mechanisms and promote more refined definitions of genetic diseases, but this knowledge will not transform those forensic loci that continue to have weak, inconsistent, or undetected disease associations into ones with substantial predictive power.

ARGUMENT

All humans inherit their DNA from their parents. This DNA contains information that can lead to significant predictions about a person's current or future diseases and various traits. Governments have enacted protections against genetic testing because of concerns about discriminatory uses of this kind of

information. *E.g.*, Genetic Information Nondiscrimination Act of 2008, 42 U.S.C. § 2000ff. But not all DNA carries such information.[2] The DNA variations used for criminal identification represent common, normal human variation and have, at most, only weak associations with any diseases or physical or behavioral traits. However, they may provide powerful information about the presence (or absence) of genetic relationships of family members, as well as very weak information about geographic or ethnic categories.

These conclusions come from an analysis of the structure and sequence of DNA in the genome, the mechanisms by which cells use genetic information, and the statistical properties of genetic tests and inheritance.

[2] Many things can be considered as providing "genetic information," such as family medical history or even someone's sex. Indeed, parts of fingerprints are inherited. *E.g.*, Terry Reed, et al., *High Heritability of Fingertip Arch Patterns in Twin-pairs*, 140A Am. J. Med. Genetics 263 (2006). Certain fingerprint features have even been associated with diseases. *See* Appendix B.

I. DNA consists of coding and noncoding sequences.

A. The human genome consists of over three billion base pairs.

Science is unraveling the molecular mechanisms of inheritance.[3] Central to the process are interactions among deoxyribonucleic acid (DNA), ribonucleic acid (RNA), and proteins. These molecules are polymers – chains of smaller units linked to one another. The building blocks of DNA and RNA are "nucleotide bases." DNA contains four distinct nucleotides (abbreviated A, T, C, and G) arrayed inside the famous double helix. An A on one strand binds to a T on the other, and a C pairs with a G. The genetic "sequence" is the order of these nucleotides. Because of the pairing rule, a single letter can denote a pair of bases – an A can be written for A:T, for example.

The vast bulk of the DNA in a human cell is crammed into the cell nucleus, coiled around proteins in structures called chromosomes. Humans normally have 46 chromosomes. Forty-four come in pairs, with one member inherited from the father and one from the mother. These 44 paired chromosomes are called "autosomes," and the pairs are numbered 1 through

---

[3] The exposition here presents material found in many textbooks. For nontechnical presentations tailored to forensic genetics, see David H. Kaye, *The Double Helix and the Law of Evidence* (2010); David H. Kaye & George Sensabaugh, *Reference Guide on DNA Evidence*, *in* Reference Manual on Scientific Evidence 129 (Federal Judicial Center, 3d ed. 2011).

22. The other two chromosomes are the sex chromosomes. Women have two copies of the X chromosome; men have one X (inherited from their mothers) and one Y (inherited from their fathers).[4]

The complete list of all of the bases on all 46 chromosomes is that person's entire genome sequence. In a sense, each person has two genomes, one from the father and one from the mother. Each "haploid" genome (the autosomes plus the sex chromosome from one parent) is about 3.2 billion base pairs long. If all these A's, C's, G's, and T's were written in book form, they would be as long as several copies of the U.S. Reports – not one volume, but the entire 552 volumes.

The two haploid genome sequences a person inherits are very similar to each other and, indeed, to those of every other human. On average, any two genomes differ in only about one base pair in one thousand. The variations are responsible for the genetic differences between people – different blood types, skin colors, disease susceptibilities, and all other genetic traits.

---

[4] The inherited chromosomes are usually not identical to those of each parent. In the process of making sperm and eggs, the parent's two copies of every autosomal chromosome swap some DNA so that each chromosome that gets passed on is a combination of the two copies of the chromosome the parent inherited.

In the early 2000s, the multinational Human Genome Project sequenced most of the 3.2 billion base pairs in the haploid genome from several anonymous individuals. Only a small percentage of those base pairs are involved in what has been understood as the primary function of DNA – "coding" the chemical structure of the proteins and RNAs in the human body. A larger, but still unknown percentage regulates the production of protein and RNA from the coding DNA. Most of the sequence, however, seems to provide no meaningful information. Part B describes the process of "gene expression" by which the body uses coding DNA sequences to produce proteins and RNAs.

B. The exons of genes code for proteins and RNAs.

A gene is not an arbitrary sequence of base pairs. Parts of the sequence specify which chemical units fit together to form a particular protein (or RNA, if that is the product encoded in the gene) and the order in which these units will be arranged.

1. Active protein-coding genes are transcribed into precursor messenger RNA (pre-mRNA).

Protein-coding genes express their sequence-specific information through three major steps: (1) transcription; (2) post-transcriptional modification and transportation; and (3) translation and post-translational processing. In the first step, the base

pairs of the gene are transcribed jot-for-jot into an RNA molecule. This transcript is known as precursor messenger RNA (pre-mRNA). The DNA in the chromosomes must open up for transcription to occur. Transcription factors (specialized proteins) bind to the chromosomal DNA. The bound transcription factors then recruit an enzyme (RNA polymerase) that produces the RNA polymer by "reading" and copying the DNA sequence into this new chemical form. Unlike duplex DNA, RNA is a single helical strand, and a different nucleotide (abbreviated U) is used instead of T.

The small region of the gene to which the RNA polymerase binds is called a promoter. The level of transcription is influenced by activator or repressor proteins that bind to still other small regulatory regions (enhancers and silencers) that also lie outside the part of the gene that specifies how to assemble the protein's building blocks. After emerging from the bound RNA polymerase, the pre-mRNA transcripts must be chemically modified for protein synthesis to progress.

2. Pre-mRNA is processed into messenger RNA.

In the second step, the pre-mRNA transcripts are modified at their ends to protect them from being degraded, and special RNA-protein complexes known as spliceosomes cut away the noncoding parts, generating mature mRNA. The discarded parts correspond

to long stretches of DNA, known as introns. Introns interrupt the much smaller coding parts – the exons – as shown *infra* Figure 1 (not to scale).[5] The RNA transcribed from the exons becomes the mRNA that will be translated into proteins.[6]

3. Messenger RNA is translated into proteins.

In the third phase, the mature mRNA makes its way to ribosomes, microscopic workbenches made of proteins and another RNA (ribosomal RNA). Yet another RNA (transfer RNA) brings the building blocks of proteins (amino acids) to the workbenches. There, they are joined in the order dictated by the mRNA transcript. By combining the amino acids in different orders, a virtually unlimited variety of proteins can be constructed.

The translation of mRNA into amino-acid sequences follows a strict genetic code. Sequences of three mRNA nucleotides code for the 20 amino acids. For example, the mRNA sequence GAA translates into a glutamic acid unit in a protein. Thus, the mRNA is "translated" into proteins according to

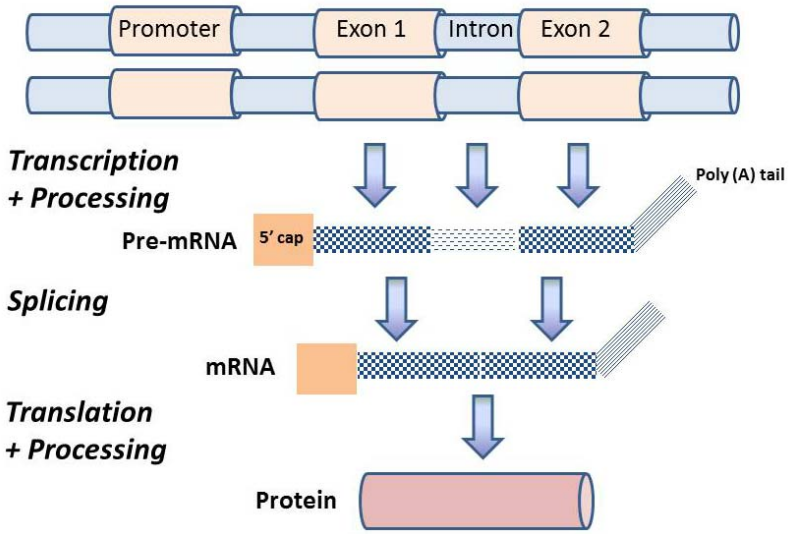
---

[5] The average number of introns in human genes is eight. The size of introns varies much more than that of exons, suggesting that introns are under less pressure from natural selection because they are less important in gene expression.

[6] In some instances, intronic sequences remain in mature mRNA.

"words" such as GAA. Given that the mRNA is a faithful transcription of the sequence of base pairs in the exons, the amino-acid sequence reflects the original sequence of A's, T's, C's, and G's in the exons. Via these three-letter "words," the gene is "expressed" as a protein. Figure 1 provides a schematic representation.

Figure 1. From Gene to Protein.
Transcription, post-transcriptional processing, translation, and post-translational processing.



4. Some genes are only transcribed into RNAs.

Many DNA sequences are not transcribed into messenger RNA. Consequently, they are not expressed as proteins. Some of these sequences are important, however, because they are transcribed into functional RNAs. These DNA sequences are still

considered genes, but their final products are non-protein-coding RNAs (ncRNAs). These ncRNAs include the ribosomal and transfer RNAs that do the work of translation as well as other ncRNAs involved in regulating expression of dozens and even hundreds of protein-coding genes. Furthermore, much shorter RNAs regulate transcription and translation. Thus, it is widely recognized that the genome is abuzz with transcription-to-RNA activity and other events that interact in the expression of the protein-coding DNA.

The discoveries of these processes have generated speculation in the press and some judicial opinions that *all* DNA sequence variations significantly affect development and health. *See infra* Part V. This is not true – even for sequences that are transcribed. Not every biochemical event along the DNA has an impact on health. First, some short ncRNA transcripts are just "noise." They are degraded quickly. Second, intronic parts of the pre-mRNA transcripts normally are removed. Third, even if one transcript could regulate expression, other transcripts may do the same job. Fourth, even if the stable transcript does affect some trait, the effect may be so minor in the context of other genetic and environmental influences as to be of no meaningful predictive or diagnostic value. Finally, even if a stable transcript has a dramatic effect on a trait, that trait may be unrelated to disease status. Consequently, regarding every bit of the genome as if it were a medical record is unwarranted.

C. There are many types of noncoding DNA.

Functional genes have a promoter sequence, exons and introns separated by splicing junctions, and untranslated regions (UTRs) at the start of the first exon and the end of the last one. The protein-coding parts of such genes (the exons) represent only a few percent of the genome. The introns boost the size of human genes by a factor of nearly 20. But genes and gene-related sequences add up to less than 40% of the genome.

The remaining 60+% lies between genes. About 25% of intergenic DNA consists of sequences that occur only once. Some of this 25% consists of apparently nonfunctional relics of evolution such as no-longer-working mutated genes, gene fragments, and other evolutionary relics known as pseudogenes.[7] Other intergenic parts are enhancers and silencers as well as sequences that code for so-called micro-RNAs that alter the stability of the mRNA or its ability to be translated.[8]

The remaining 75% of intergenic DNA – almost half the total human genome – is composed of DNA

---

[7] For further explanation, see James Watson, et al., *Molecular Biology of the Gene* 142 (6th ed. 2008).

[8] The definition of a gene is subject to vigorous debate, and the percentages given here are not exact. Many authorities classify enhancers and silencers as part of the gene-related DNA mentioned in the previous paragraph.

sequences repeated many times in the genome. This DNA mostly falls into two classes. First, microsatellite DNA – also called simple sequence repeats (SSRs) or short tandem repeats (STRs) – represents nearly 3% of the genome. It contains tandemly repeated short sequences. The most common STRs are made of only two nucleotides (e.g., ACACACACACACACAC). For many STR loci, different individuals have a different number of repeats. These length variations result from difficulties that sometimes arise when all the DNA is replicated prior to cell division: imperfectly copied versions of STRs arise with, say, 7 or 9 copies of AC rather than 8.

Second, "genome-wide repeats" have repeat units 10 or 100 times as large as those in STRs. There are numerous classes of genome-wide repeats, but all are forms of transposable elements – sequences that can move from one place in the genome to another, usually leaving the original copy behind. The transposons have been so successful at duplicating themselves over evolutionary time that they constitute about 45% of the human genome.[9]

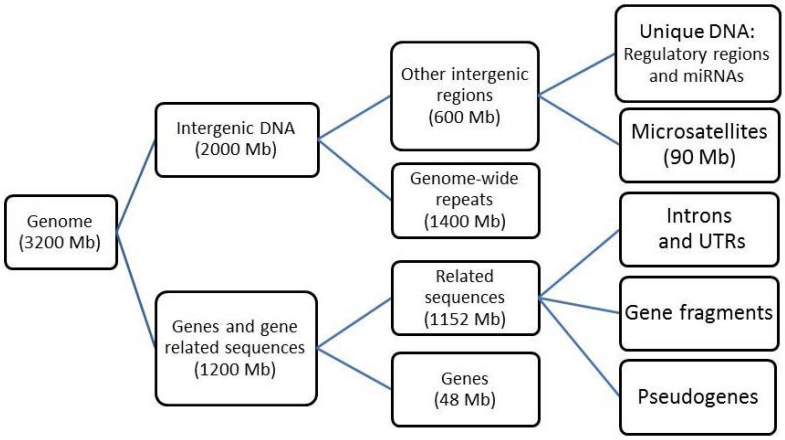
Figure 2 summarizes the organization and content of the human genome. (Early DNA databases employed the type of genome-wide repeats known as minisatellite DNA, or VNTRs, but these were

[9] It omits about 3-4% of the genome associated with structural elements of chromosomes – centromeres and telomeres.

supplanted in the mid-1990s by the STRs described in the next section.)

Figure 2. Organization and Content of the Human Genome

Adapted from James D. Watson et al., *Molecular Biology of the Gene* 143 (6th ed. 2008). "Mb" stands for a million base pairs. CODIS profiles come from selected microsatellites that compose ~0.004 Mb of intergenic and intronic DNA.



II. CODIS loci are one type of repetitive DNA useful for human identification.

In 1990, the FBI introduced software, known as CODIS (Combined DNA Index System), for sharing VNTR and other information among state and local databases. In 1997, the FBI chose 13 STRs for the CODIS database system. There are thousands of STRs, but the 13 were among the most well-defined at the time, and they had no disease associations with predictive power. Currently, the FBI is considering adding more intergenic STRs. Douglas R. Hares, *Addendum to Expanding the CODIS Core Loci in the United States*, 6 Forensic Sci. Int'l: Genetics e135 (2012).

In each of the 13 "CODIS core loci," a short DNA sequence (a "tetranucleotide" of four base pairs) is repeated a variable number of times. All these loci are on autosomes. Because everyone normally has two copies of each autosome,[10] a person should have two copies of each CODIS STR, one from each parent. Five of the loci are in the introns of genes; the other eight are found outside of genes.[11] The 13 STRs combined make up about 4,000 base pairs in the average person, only a little more than one-millionth of the 3.2 billion base pairs in a haploid genome.

Spread over 12 different chromosomes, the CODIS loci have names such as TPOX[12] and D5S818.[13] The CODIS STRs were chosen in part because they are commonly found in a variety of lengths in every major U.S. population group. Each different number of repeats – reflecting a variation in the sequence of bases in that location – is called an

---

[10] A few people are born with extra chromosomes. The limited privacy implications of CODIS profiles for revealing chromosomal abnormalities are noted in David H. Kaye, *Please, Let's Bury the Junk: The CODIS Loci and the Revelation of Private Information*, 102 Nw. U. L. Rev. Colloquy 70 (2007).

[11] *But see* Michael Klintschar, et al., *Physical Location and Linked Genes of Common Forensic STR Markers*, 128 Int'l Congress Series 801 (2006) (suggesting that the D3S1358, D7S820 and D18S51 STRs should be considered intronic).

[12] This STR is found in an intron of the thyroid peroxidase gene.

[13] Names of the eight markers not found in genes are in the form D#S###. The first number indicates the chromosome; the second, a location on the chromosome.

"allele." For each locus, at least six alleles are fairly common in humans. Thus, D5S818 is a place on chromosome 5 where people have a series of repeated AGAT bases. The most common numbers of AGAT repeats are 11, 12, and 13, which make up over 85% of the alleles found in samples of Americans.[14] If a person has a rarer, 7-repeat allele, the sequence will read (AGAT)(AGAT)(AGAT)(AGAT)(AGAT)(AGAT)(AGAT).[15] Each of us will normally have two copies of the sequence. For example, if a boy inherited a 7-repeat STR on his father's chromosome 5 and a 12-repeat one on his mother's, his type for that locus would be 7, 12.

An individual's overall CODIS profile is 13 such pairs of numbers. To produce the profile, the laboratory measures the lengths of the 13 pairs of alleles. It does not examine the other 99.9999% of the genome.

---

[14] Nine- and 10- repeat alleles make up another 10%, with 7-, 8-, 14-, and 15-repeat alleles accounting for another few percent. John M. Butler, et al., *Allele Frequencies for 15 Autosomal STR Loci on U.S. Caucasian, African American, and Hispanic Populations*, 48 J. Forensic Sci. 908 (2003).

[15] Some CODIS loci have compound repeat units (such as TCTA/TCTG in the vWA intron), and some loci have occasional repeat units that are less than the full four base pairs long. These complications do not detract from the forensic value of the loci.

III. Variations in CODIS loci are not known to be causally related to physical or behavioral traits.

Variations in the length of the forensically relevant STRs do not appear to cause any differences in any genetic traits. Indeed, their value in human identification relates to this feature. Regions of DNA with particular functions – whether they encode proteins or RNAs, or are regulatory – usually do not vary much among people (although variations that have minor phenotypical effects, such as hair color, are present within and across populations). Most new variations in functional DNA tend to be harmful because they would result in the production of a functionally different protein or ncRNA or would change the quantity of those products. Individuals with deleterious DNA sequences will usually not have as many children as individuals with functionally normal sequences. Natural selection thus tends to conserve functional sequences from one generation to the next. Conversely, nonfunctional regions usually will be free of this selective pressure. Over time, mutations there can spread more readily through the population.

Most STRs are not under detectable selective pressure. Therefore they can easily vary widely among individuals with little consequence to any individual. The large number of variants of the forensic STRs found in all major population groups makes them particularly valuable for identity testing. At the same time, the lack of evidence of constraints

on these STRs means that most are unlikely to have any function related to genetic traits such as disease susceptibility.

A. CODIS loci are not known to be translated into proteins.

There is no evidence that CODIS loci are translated into amino acids of a protein. All proteins are based on the translation of mRNA, and no one has ever pointed to any mRNA that includes any of the CODIS alleles. The absence of any direct evidence is not conclusive, but it is substantial evidence. After all, these loci have been used as signposts in disease-association studies for some 20 years with no published reports establishing that they have predictive value in any major population group.

The addition or subtraction of a single tetranucleotide repeat unit in an exon could have a dramatic, adverse effect on the protein product.[16] Thus, STR variations in coding DNA are unusual, but they do occur, and a number are known to cause neuromuscular diseases. However, these are almost always triplet repeats – a type not used for forensic loci – with many more expansions than the

[16] The DNA "words" that spell out an amino acid have three base pairs each. Adding or subtracting one set of four letters changes the way all the downstream DNA is read. That some repeat numbers have no adverse effects while others do cause disease by changing protein structure therefore seems unlikely.

tetranucleotide CODIS STRs exhibit. The CODIS loci were chosen because, unlike the disease-producing triplet repeats, they are *not* part of exons.

B. CODIS loci are not known to be transcribed into regulatory ncRNA, but one intronic CODIS locus has been shown to have some regulatory effect.

Within the past decade, much has been learned about the mechanisms by which some fraction of the ncRNA plays a role in regulation – turning genes on and off and modulating the quantity of proteins produced. Scientists know much less about these ncRNAs than they do about mRNA. A CODIS locus might encode some as yet unrecognized ncRNA, but, again, DNA that performs important functions, such as encoding functional ncRNA, usually would be highly conserved, as natural selection tends to weed out damaging variations.

A number of STRs within introns have been shown to regulate splicing and hence to alter mRNA transcripts. In one case, the STR is a tetranucleotide repeat (but in a larger size range and different part of the genome than any of the CODIS tetranuceotide repeats). Laura P.W. Ranum & Thomas A. Cooper, *RNA-Mediated Neuromuscular Disorders*, 29 Ann. Rev. Neuroscience 259 (2006).

One intronic CODIS locus (TH01) has been reported to alter levels of expression of the gene in

which it is located. Véronique Albanèse, et al., *Quantitative Effects on Gene Silencing by Allelic Variation at a Tetranucleotide Microsatellite*, 10 Human Molecular Genetics 1785 (2001). Thus, it is possible that at least one current CODIS locus participates in a regulatory system that at some point might be shown to affect individual health. However, elucidating a mechanism by which a CODIS STR could have a disease-related effect would not automatically make the locus any more predictive than it is now, and the TH01 locus has not been shown to be useful for diagnosis or prediction. *E.g.*, Eric G. Jonnsson, et al., *Failure to Replicate an Association Between a Rare Allele of a Tyrosine Hydroxylase Gene Microsatellite and Schizophrenia*, 248 Eur. Arch. Psychiatry & Clinical Neuroscience 61 (1998).

IV. Variations in CODIS loci could be correlated with physical or behavioral traits, but no substantial correlations have been observed in forensically relevant populations, and database administrators, employers, or health insurers are not likely to find correlations useful for disease diagnosis or prediction.

Since the earliest years of forensic DNA testing, it was understood that noncoding sequences could display "some association" with disease-causing mutations. U.S. Congress Office of Technology Assessment, *Genetic Witness: Forensic Uses of DNA Tests* 132 (1990). Most research trying to find such

mutations relied on these associations, and much of it still does. In fact, the CODIS STRs are a subset of the many STRs used to locate and then develop tests for causal mutations. To see how this gene hunting works, suppose that a man is born with a disease-causing variant in one copy of a gene and that this variant is very close to an STR locus. Assume the disease-causing variant is on a chromosome that the man inherited from his father and that the STR on that chromosome has 10 repeats. The man's maternal chromosome does not have the disease-causing variation, and its nearby STR has, say, 12 repeats. Assume the man has two children and that one inherited from him a copy of the chromosome with the disease-causing variation and the other did not. The child with the disease-causing variation also inherited the 10-repeat allele of the nearby STR; the other child inherited the 12-repeat copy. Observing such co-inheritance of a marker and a genetic disease tells researchers that a gene mutation causing the disease lies near the STR marker. But the STR marker in no way "causes" the disease; it is simply a tool for locating the causal mutation.[17] A genetic test for that

[17] The man's children will not always inherit the disease-causing variation along with the same allele of the nearby STR. As the man's body makes sperm, his two copies of chromosome 4 will exchange some genetic material. *See supra* note 4. The closer the disease-causing variation and the STR are to each other, the more likely they are to be inherited together; the farther apart, the less likely. This frequency of co-inheritance was used to create the first genetic maps nearly a century ago, long before DNA was known to be the carrier of inheritance.

mutation then can be developed. This genetic test may be generalizable to other families even though the STR marker is not.

This association between the 10-repeat STR allele and the disease-causing variation can be used for risk prediction within a family, but as the population grows larger and more diverse, as in the United States, the marker loses its value. Most people with the disease-causing variation will not have a 10-repeat allele in the nearby STR. That was the repeat length in this particular family, but affected individuals in other families, with other genetic histories, are likely to have repeat numbers that span the entire range of alleles. Similarly, most people with the 10-repeat allele will not carry the disease-causing variation. Crude calculations suggest that a positive result on the 10-repeat "test" for the disease-causing mutation in a national database might be correct in less than one case in a thousand.[18]

This problem of low predictive value is particularly acute with a rare disease or trait. Any imperfect marker of a trait becomes less useful when the trait is rare. Take a test for a disease that has a very low "false positive" rate – for example, out of 100 tests, only one is a false positive. Nevertheless, if the condition is only found in one person in a thousand, applied to 1000 people, the test will show about 10 false positives for the one true positive. The positive

[18] *See* Kaye, *Bury the Junk*, at n.26.

predictive value – the percentage of people with a positive test who actually have the condition – will be quite low. As most genetic diseases or genetic variations strongly linked to diseases are uncommon, typically occurring at well under a one percent rate, even an unrealistically strong association between an STR repeat length and a disease typically would not be a good disease predictor.

One might think that there could be greater predictive value if an STR locus is physically linked to a gene involved in a common disease such as diabetes. But common diseases and health problems are affected by multiple genes and strongly influenced by environmental factors. The same is true of behavioral traits that, some have speculated, might become the target of a screening test on identification profiles in a database. David H. Kaye, *Behavioral Genetics Research and Criminal DNA Databanks*, 69 Law & Contemp. Probs. 259 (2006). Even if a particular CODIS STR coincidentally turned out to be strongly associated with one gene in such a system, the correlation with the disease will be greatly diluted.

These are good reasons to believe that any associations between CODIS alleles and genetic traits are likely too small to be useful. In fact, no useful associations have been proven.[19] Some writers claim that

[19] A Vermont trial court found "genetic information [in five CODIS loci] "associated with at [least] eight different medical conditions [and] genetic functions or physical traits associated

(Continued on following page)

studies have identified repeat numbers for the CODIS loci that are powerfully predictive or diagnostic of important diseases in the U.S. population.[20] These claims almost never come from scientists and are not substantiated by scientific studies. *See* John M. Butler, *Advanced Topics in Forensic DNA Typing: Methodology* 228 (2012) (noting the misinterpretation of the literature in a law review exchange said by the Maryland Court of Appeals in this case to demonstrate "considerable current debate," *King v. State*, 42 A.3d 549, 560 n.17 (Md. 2012), *cert. granted sub nom. Maryland v. King*, No. 12-207 (U.S. Nov. 9, 2012)).

At present, there are no published, replicated studies that permit the database records of variations in CODIS STR lengths to be the basis for valid

---

with one or more of the CODIS loci." *State v. Abernathy*, No. 3599-9-11, at 11 (Super. Ct. June 1, 2012), *app. pending*, Vt. No. 2012-102. Neither the opinion nor the testimony presented to the court adequately addressed the predictive value of the putative associations. In addition, the studies listed in the opinion have not been replicated and shown to apply to the U.S. population. Some are merely family studies, which are not expected to provide useful predictive disease markers outside of the families studied. Finally, it is instructive to compare the small number of reports of disease associations to CODIS STRs with the much larger number for fingerprint features. *See* Appendix B.

[20] *E.g.*, Electronic Frontier Foundation, New Research on "Junk" DNA Raises Questions on Eve of Crucial Court Hearing, Sept. 11, 2012, https://www.eff.org/deeplinks/2012/09/new-research-on-junk-dna-raises-questions ("DNA . . . in the form of . . . an extracted profile . . . can reveal . . . medical history, predisposition for disease, and possibly even behavioral tendencies and sexual orientation.").

predictions or diagnoses. To the contrary, the most recent published review concludes that

> The . . . standard and recommended CODIS panels of STR loci . . . continue to be of limited significance for assessing phenotypes. . . . Several . . . overlay predicted sites for genomic regulation, but there is no evidence that any particular repeat [is] indicative of phenotype. The utility of the CODIS profile itself, even in light of the significance of various epigenetic effects and roles of noncoding RNAs, is limited to identification purposes at this time.

Sara H. Katsanis & Jennifer K. Wagner, *Characterization of the Standard and Recommended CODIS Markers*, 41 J. Forensic Sci. (forthcoming 2013).

V. Estimates of the proportion of the genome that is "junk" are largely irrelevant to the privacy interests implicated by CODIS loci.

The term "junk DNA" has both a colloquial and a scientific meaning. Eliding one with the other is dangerous.[21] Only a subset of noncoding DNA is "junk" as the term is used (often with misgivings) in genetics. For the term to apply, the DNA not only must lie outside of genes, but variations in its

[21] *See* John Timmer, *Most of What You Read Was Wrong: How Press Releases Rewrote Scientific History*, Ars Technica, Sept. 10, 2012, http://arstechnica.com/staff/2012/09/most-of-what-you-read-was-wrong-how-press-releases-rewrote-scientific-history/.

base-pairs must have no detectable effect on the fitness of the organism.[22] If substituting a randomly generated sequence would not noticeably influence reproductive success, then the original sequence is, in a sense, "junk." But geneticists always have maintained that such DNA is not necessarily useless or inactive. Sometimes we can salvage parts of our junk for new uses. Thus, even when "junk" sequences of base-pairs do not affect how an individual functions, they can be a kind of evolutionary "treasure." *See* Haig H. Kazazian, *Mobile DNA: Finding Treasure in Junk* (2011).

In this Part, *amici* do not seek to settle the debate over the fraction of the genome that is, in an evolutionary sense, "junk." Rather, we explain why the question is orthogonal to the matter before the Court. The salient issue in this case is whether length variations in the STRs have a direct impact on, or a strong enough correlation with, significant non-observable traits to constitute private information. *Amici*'s position is that differing estimates of the percentage of the genome that is "junk" have no real bearing on this privacy issue. On the one hand, being "junk" is not the same as being uninformative. Therefore, even if the CODIS STRs should be called "junk,"

---

[22] "Fitness" is the average contribution to the gene pool of the next generation that is made by an average individual of the specified genotype. If differences between alleles affect fitness, then alleles with higher fitness will tend to become more common.

they could (in theory) still pose a privacy problem. On the other hand, even if only a tiny fraction of the genome sequence were uninformative, if the CODIS loci reside in that fraction, they would not pose a privacy problem.

A. In genetics, "junk DNA" denotes sequences that lie outside of genes and that are not under detectable selective pressure: that such DNA exists is not in doubt.

In 1972, biologists introduced the term "junk" to refer to "untranscribable and/or untranslatable DNA," Susumu Ohno, *So Much "Junk" DNA in Our Genome*, 23 Brookhaven Symposia in Biology 366, 367 (1972), that does not function as genes and whose specific sequence is "relatively useless." David E. Comings, *The Structure and Function of Chromatin*, *in* 3 Advances in Human Genetics 237, 313 (H. Harris & K. Hirschhorn eds. 1972). Ample data had established that many organisms have "excess" DNA – in copious amounts.[23]

---

[23] For example, it is striking that total genome size does not always increase with the complexity of organisms: "the lowly liverwort has 18 times as much DNA as we, and the slimy, dull salamander known as *Amphiuma* has 26 times our complement of DNA." Comings, *supra*, at 313. Even very similar species have wildly disparate genome sizes. *E.g.*, Sean R. Eddy, *The C-value Paradox, Junk DNA and ENCODE*, 22 Current Biology R898 (2012).

To explain this phenomenon, several researchers proposed that a large fraction of DNA simply has little or no adaptive advantage for the organism. W.F. Doolittle & C. Sapienza, *Selfish Genes, the Phenotype Paradigm and Genome Evolution*, 284 Nature 601 (1980); Leslie E. Orgel & Francis H.C. Crick, *Selfish DNA: The Ultimate Parasite*, 284 Nature 604 (1980). They suggested that this fraction comes from "selfish DNA" that makes copies of itself at the (usually slight) expense of a host genome. *See supra* Part I.C (discussing transposons). Not being subject to natural selection, as untold generations go by, the copies can decay, through many small mutations, into "junk."

That not *all* the DNA sequences of complex organisms are involved in RNA and protein synthesis is indisputable. Conversely, since the early days of molecular biology, that *some* noncoding DNA regulates the operation of genes has been clear. The question that invites disagreement is "how much junk, how much func?"[24]

B. "Junk" DNA sequences could be biologically useful or interesting yet not be useful for disease diagnosis or prediction.

The originators of the "junk DNA" label recognized that some DNA elements containing arbitrary

[24] Early estimates of "mutational load" suggested "that maybe only 1-20% of the human genome could be genic, with the rest evolving neutrally or nearly so." Comings, *supra*, at 313.

DNA sequences could be biologically useful, if only as spacing material. *See* Comings, *supra*, at 316; Ohno, *supra*, at 367–68. However interesting or accurate these theories may be, they have no necessary connection to whether specific sequences are informative of individual traits.

On the one hand, we have seen that correlations could arise from the proximity of STRs to important genes. Therefore, it would be wrong to argue that just because "the genetic markers used for forensic DNA testing . . . show only the configuration of DNA at selected 'junk sites' which do not control or influence the expression of any trait," it clearly follows that they "do not reveal information relating to any medical condition or other trait." House Committee on the Judiciary, *Report on the DNA Analysis Backlog Elimination Act of 2000*, 106th Cong., 2d Sess., H.R. Rep. No. 106-900(1), at 27 (letter from Dep't of Justice to Chairman, House Comm. on the Judiciary, July 21, 2000).

But by the same token, it would be wrong to intimate that just because some noncoding DNA is useful as a "spacer" or actually is biologically active, the CODIS loci probably carry highly personal information. Moreover, even if some length differences of some of these STRs prevented or increased transcription of regulatory RNAs, it would not necessarily follow that they are "essential biological instructions for growth and survival of an organism," Stephen B. Mercer & William G. McLain, *Maryland's DNA Databank*, Md. Bar J., Nov.-Dec. 2004, at 17, 22, and

therefore "sensitive and private." Electronic Frontier Foundation, Letter, *Haskell v. Harris*, No. 10-15152 (9th Cir. Sept. 11, 2012) (en banc). They might have no effect at all because there are alternative paths to the same outcomes, or the affected traits might be no more significant than, say, the thickness of the eyebrows or the width of the nose. One must ask whether the length variations of the particular STRs actually convey meaningful information, and they seem to contain less trait-related information than a photograph of an arrestee.

C. ENCODE data do not reveal that anywhere near 80% of the genome contains medically relevant information.

Just as something can be inactive and functional, so too can something be active and nonfunctional. This fact prompts a few words about recent publicity surrounding the publication of data from the Encyclopedia of DNA Elements (ENCODE) Project. ENCODE is a federally funded public research consortium launched in September 2003, with the goal of identifying "all functional elements in the human genome sequence." National Human Genome Research Institute, The ENCODE Project: ENCyclopedia Of DNA Elements, Sept. 27, 2012, http://www.genome.gov/10005107. In September 2012, the project's leaders announced the release of many data sets on biochemical activity along the genomes of different types of human tissues. They stated that these data "enabled us to assign biochemical functions for 80% of

the genome." Ian Dunham, et al., *An Integrated Encyclopedia of DNA Elements in the Human Genome*, 489 Nature 57 (2012) (abstract).

The value of the data to future research cannot be gainsaid, but interpretations such as the following are unsupportable:

> This research has determined that more than 80% of DNA once thought to be no more than "junk" [controls] how our cells, tissue and organs behave. . . . Based on the research, it is highly likely the genetic markers contained in each Appellant's DNA profile reveal much more information than just his or her identity – information that is sensitive and private. . . .

Electronic Frontier Foundation, *supra* note 20.

The misperception that all noncoding DNA is involved in disease processes or bodily functioning is understandable in light of hyperbolic headlines such as David Brown & Hristio Boytchev, *"Junk DNA" Concept Debunked by New Analysis of Human Genome*, Wash. Post, Sept. 5, 2012. Nonetheless, the ENCODE publications do not state that 100% or even 80% of the genome makes organs function, stimulates tissue growth, turns normal cells into cancerous ones, makes us tall or short, fat or skinny, gay or straight. The main paper defines "a functional element" to include "a discrete genome segment that . . . displays a reproducible biochemical signature," and "functional" merely designates "something that changes a

biochemical property of the cell." Ian Dunham, et al., *supra*.[25]

This definition contrasts sharply with the notion of functional as affecting a nontrivial trait. The ENCODE papers show that 80% of the genome displays signs of certain types of biochemical activity – even though the activity may be insignificant, pointless, or unnecessary. This 80% includes all of the introns, for they are active in the production of pre-mRNA transcripts. But this activity hardly means that all introns are doing useful work as regulators or anything else.

In sum, a "functional element" in ENCODE has no clear or direct implications on how informative the forensic loci are. Even after this phase of ENCODE, there is still a huge number of base pairs that might or might not be regulatory or influence regulation and, hence, gene expression. The ENCODE papers appear to place the percentage of the genome that regulates the 2% or so that encodes proteins at less

---

[25] That the ENCODE researchers are using "functional" in a very technical sense is clear from their own writings. The project's leader explained that this definition of "functional" does not mean "something that changes a phenotypically observable trait as that affects the whole organism." Ewan Birney, ENCODE: My Own Thoughts, Ewan's Blog; Bioinformatician at Large, Sept. 5, 2012, http://genomeinformatician.blogspot.com/2012/09/encode-my-own-thoughts.html. Dr. Birney's personal estimate of the percentage of the genome that affects the organism beyond an inconsequential chemical reaction in a cell is only 20%. *Id*.

than 20%. The CODIS STRs might be (but, more likely, might not be) among this 20%. The ENCODE findings indicate that the system that regulates gene expression is exquisitely complex, but they do little to change the status of "junk DNA" in general, or – to focus on the only relevant point in this case – the CODIS loci in particular.

VI. The CODIS loci can provide information about specific family relationships as well as weak information about racial or ethnic background.

Because children inherit all their DNA from their biological parents, the CODIS loci can be powerful tools for determining whether two people could be genetically related as parent and child. Because a child receives one allele at every autosomal locus from the mother and one from the father, it is often simple to show that a pair of unrelated individuals cannot be parent and child. For instance, if one individual has a 7-repeat and an 11-repeat allele for D5S818, but another individual has a 13-repeat and a 15-repeat allele, then barring mutation, the two cannot be parent and child. This is the most powerful genetic information other than identity that the CODIS profiles contain – that two people are not parent and child.[26]

[26] The CODIS profiles cannot conclusively show whether two people are genetic siblings. It is possible, though highly

(Continued on following page)

Detecting rather than excluding parent-child relationships from profiles in a CODIS database is far more difficult. Some portion of people in the world will share at least one allele at each CODIS locus with any given person. Every one of these many individuals would be on a list of those people who could be a genetic parent of that child. For an "average" CODIS profile (one that contains the two alleles of median incidence at every locus), the probability of a random person sharing one allele at each locus is somewhat above 1/1000. Henry T. Greely, et al., *Family Ties: The Use of DNA Offender Databases to Catch Offenders' Kin*, 34 J.L. Med. & Ethics 248, 252 (2006). Thus, in a database of 10 million people, a person with an "average" profile would have about 10,000 candidates for being his parents or children. Someone with access to the CODIS profiles of two people could tell if they might be among the many thousands of possible parent-child pairs.

Beyond that, someone could use the number of alleles shared and the population frequencies of these alleles to estimate a "kinship index" for any possible genetic relationship (parent-child, siblings, uncle-aunt

---

unlikely, that two full siblings could have, by chance, inherited different alleles at each locus from each of their parents and thus not share *any* CODIS alleles. However, in some instances, the pattern of shared and unshared alleles can establish a distinct probability of relatedness. And, of course, a comparison of two profiles could establish whether two people were identical twins – these twins would share both alleles at each locus of their CODIS profiles.

and nephew-niece, cousins, grandparent-grandchild, and so on) based on any pair of CODIS profiles. The index would almost never be definitive for picking out relatives within a database. Most pairs with a large index value would not actually be relatives, and many unrelated pairs would have larger index values than would true relatives.

One could make much more accurate assessments of family relationships by using more STRs and other markers, but there are too few CODIS markers, with too few alleles, to do this well. *See, e.g.,* David H. Kaye, *The Genealogy Detectives: A Constitutional Analysis of "Familial Searching"*, 51 Am. Crim. L. Rev. No. 1 (forthcoming 2013), available at http://ssrn.com/abstract=2043091.

The frequencies of CODIS alleles often show similar patterns in populations across the globe – the same alleles tend to be common (or rare) in most populations – but there are variations in the frequencies of particular alleles in different populations. At D5S818, for example, the 11-repeat allele is about 50% more common in Caucasians than African-Americans. The 8-repeat allele is rare in both groups, but it is about 16 times more common in African-Americans. Butler, et al., *supra* note 14. These variations have nothing to do with the genetic characteristics that are associated with these groups, such as skin color. They are just the result of migrations and chance genetic history. *See, e.g.,* Richard S. Cooper, et al., *Race and Genomics*, 348 New Engl. J. Med. 1166, 1167 (2003); Kirk E. Lohmueller, *Graydon, et al.*

Provide No New Evidence that Forensic STR Loci Are Functional, 4 Forensic Sci. Int'l: Genetics 273 (2010).

A CODIS profile could be used to calculate probabilities that someone would be described as Caucasian, African-American, or Hispanic, but categorical inferences would not be very accurate,[27] and attempts to predict the census-type race of a person from a CODIS profile would seem pointless considering that apparent race already would be known.

Finally, we should note that both family connections and racial categories can provide minimal information about at least some disease probabilities. Caucasians are more likely than African-Americans to be diagnosed with cystic fibrosis; African-Americans are more likely to be diagnosed with sickle cell disease. But the combination of an inaccurate racial identification, shared genetic backgrounds ("admixture"), and a different incidence of a low-probability disease will not yield valuable or troubling information.[28]

[27] Some of the allele frequency differences used to make inferences will be the result of sampling variance, and census-type categories of "race" do not mirror bio-geographic ancestry. For example, some Hispanic-Americans are of entirely European ancestry.

[28] To take an extreme example, men are much more likely than women to be diagnosed with prostate cancer; women are much more likely to be diagnosed with cervical cancer. Being able to make a guess about whether someone is male or female will not provide much information about whether that person will be diagnosed with one of those diseases.

CONCLUSION

The Court should recognize that CODIS profiling is not the type of genetic testing that supplies significant information on disease risk or other physical or behavioral genetic traits.

Respectfully submitted,

MICHAEL L. FOREMAN
Counsel of Record

APPENDIX A

IDENTIFICATION OF *AMICI*

FREDERICK R. BIEBER has worked as a medical geneticist for 32 years and is Associate Professor of Pathology at Harvard Medical School. He is certified by the American Board of Medical Genetics and is a laboratory director of the Center for Advanced Molecular Diagnostics at Brigham and Women's Hospital. His writing includes papers on human genetics, pathology, forensic medicine, forensic DNA science, and familial searching in DNA databases. He was a member of the World Trade Center Kinship and Data Analysis Panel and the Hurricane Victim DNA Identification Expert Group following Hurricanes Katrina and Rita. He has served on advisory boards or as an advisor for the National DNA Databank of Canada, the National Institutes of Health, and the states of Connecticut, Massachusetts, and Virginia, and was a member of the Congressionally mandated DNA Advisory Board.

BRUCE BUDOWLE is Professor, Department of Forensic and Investigative Genetics, and Executive Director, Institute of Applied Genetics, at the University of North Texas Health Science Center. As a senior scientist and researcher for the FBI for 26 years, he was an architect of the CODIS system, a chair and member of the Scientific Working Group on DNA Methods, Chair of the Scientific Working Group on Microbial Genetics and Forensics, and Chair of the DNA Commission of the International Society of

Forensic Genetics. He was a principal advisor in efforts to identify victims from the World Trade Center attack and helped establish a mitochondrial DNA sequencing program to enable high throughput sequencing of human remains. He has published hundreds of articles and authored or coauthored books on molecular biology techniques, electrophoresis, protein detection, and microbial forensics. He also has carried out research on genetic risk factors for diseases such as insulin-dependent diabetes mellitus, melanoma, and acute lymphocytic leukemia. At UNTHSC his research focuses on human forensic identification, microbial forensics, and emerging infectious disease.

ARAVINDA CHAKRAVARTI is Professor, Departments of Medicine, Pediatrics, and Molecular Biology and Genetics, Johns Hopkins University Bloomberg School of Medicine and Department of Biostatistics. His laboratory, which is part of the Institute for Genetic Medicine, focuses on the development and applications of molecular genetic, genomic, and computational methods for understanding the features of complex disease gene architecture in birth defects, cardiovascular disorders, and mental illness. A member of the National Academies' Institute of Medicine, he has served on advisory committees for the National Institutes of Health and is an editorial board member of four genetics journals.

HENRY T. GREELY is the Deane F. and Kate Edelman Johnson Professor of Law, Stanford Law School. Director, Center for Law and the Biosciences; Professor (by courtesy) of Genetics, Stanford School of

Medicine; Chair, Steering Committee of the Center for Biomedical Ethics; and the founder and Director of the Stanford Interdisciplinary Group on Neuroscience and Society. A Fellow of the American Association for the Advancement of Science (AAAS), he is the coauthor of a chapter of the FJC *Reference Manual on Scientific Evidence* and has written extensively on ethical and legal issues raised by genetics, including both genetic genealogy and family forensic DNA searching. He is an additional author of this brief.

MITCHELL M. HOLLAND is Associate Professor, Biochemistry and Molecular Biology Department, and Director, Forensic Science Program, The Pennsylvania State University. A Fellow of the American Academy of Forensic Sciences, Dr. Holland was Senior Vice President for Operations and Laboratory Director, The Bode Technology Group; Scientific Laboratory Director and Head of Research, Armed Forces DNA Identification Laboratory, Armed Forces Institute of Pathology. His publications include papers on mitochondrial DNA and STR loci as used in human identification. He participated in many human remains cases involving military personnel as well as victims from the World Trade Center disasters and commercial airline accidents.

DAVID H. KAYE is a member of the graduate faculty of The Pennsylvania State University Forensic Science Program and Distinguished Professor and Weiss Family Scholar, School of Law. He also is Regents' Professor of Law Emeritus at Arizona State University, where he was the founding director of the Center

for the Law, Science, and Innovation and a Professor (by courtesy) in the School of Life Sciences. He served on the National Academy of Sciences Committee on DNA Forensic Science: An Update, was the legal reporter for National Commission on the Future of DNA Evidence, and is a coauthor of chapters on both statistics and DNA evidence in the Federal Judicial Center's (FJC's) *Reference Manual on Scientific Evidence*. With support from the Human Genome Project, he has written on the significance of law enforcement DNA databases for behavioral genetics research and on CODIS profiles as medical information. His research or commentary has appeared in journals of medicine, genetics, forensic science, and statistics, and psychology as well as law. He is the principal author of this brief.

HAIG H. KAZAZIAN, JR., is Professor of Genetics, Institute of Genetic Medicine, Johns Hopkins University. Formerly, he was Seymour Gray Professor of Genetics, Perelman School of Medicine, University of Pennsylvania and Chair of the Department of Genetics. Before going to Penn in 1994, he was Director of the Johns Hopkins Center for Medical Genetics. Dr. Kazazian discovered that transposable elements are active in human beings and cause disease through insertional mutagenesis. His laboratory studies the frequency of new insertions in the population, in cancers, and in other diseases, and is investigating the molecular diagnosis and treatment of hemophilia A. He is a member of the American Academy of Arts and Sciences, the Association of American Physicians,

the National Academies' Institute of Medicine, and a past president of the American Board of Medical Genetics. He served on both National Academy of Sciences Committees on DNA Technology in Forensic Science.

KENNETH K. KIDD is Professor of Genetics, Psychiatry, and Ecology and Evolutionary Biology, Yale University School of Medicine. He has pursued research in many areas of human genetics, including medical genetics (studies of neuropsychiatric disorders and simple Mendelian disorders), gene mapping (both physical and genetic), database design for modern genetic data, and a variety of molecular methodologies. His laboratory examines human genome diversity at the DNA level. With support from the National Science Foundation, he established ALFRED, the ALlele FREquency Database as a resource for the international scientific community. Dr. Kidd is a certified Medical Geneticist of the American Board of Medical Genetics and a Fellow of the American Association for the Advancement of Science. He has served on multiple editorial boards and federal advisory committees, including advisory panels for DNA identification of victims of the World Trade Center attack and Hurricane Katrina.

TIM D. SPECTOR is Professor of Genetic Epidemiology, Kings College London and director of the TwinsUK twin register of 12,000 adult twins with genetic information. He trained as a physician and in Epidemiology and Genetics and runs the register, which is the most deeply genotyped and phenotyped

in the world. Through Genome Wide Association Studies his group has found over 300 novel gene loci in over 30 disease areas including osteoporosis, osteoarthritis, and melanoma. He has published over 600 research articles on common diseases and traits. A Senior Investigator for the National Institute for Health Research of the United Kingdom's National Health Service, he also led the Wellcome Trust MuTHER study, which discovered how a "master regulator" gene can cause a cascade of metabolic effects in other genes that could contribute to diabetes and other conditions. Dr. Spector also is joint chair of the VISIGEN visible traits global genetics consortium and has published articles on ancestry, hair and eye color.

HUNTINGTON F. WILLARD is Director, Institute for Genome Sciences & Policy, and Nanaline H. Duke Professor of Genome Sciences at Duke University. He previously held faculty positions at the University of Toronto, Stanford University, and Case Western University, where he was Chairman of the Department of Genetics. He is the author or coauthor of over 300 scientific publications in human genetics and genomics, including the widely used textbook Genetics in Medicine. He has served on the editorial boards of numerous scientific journals and was co-founder of Human Molecular Genetics. Dr. Willard is a previous President of the American Society of Human Genetics and is a member of the American Academy of Arts & Sciences. He served on the Secretary's Advisory Committee for Genetics, Health & Society for the

Secretary of the Department of Health and Human Services.

APPENDIX B

SOME STUDIES REPORTING DISEASES ASSOCIATED WITH FINGERPRINT FEATURES

Alter, Milton & Robert Schulenberg, *Dermatoglyphics in Congenital Heart Disease*, 41 Circulation 49 (1970)

Avila, M.T., et al., *Neurodevelopmental Interactions Conferring Risk for Schizophrenia: A Study of Dermatoglyphic Markers in Patients and Relatives*, 29 Schizophrenia Bulletin 595 (2003)

Bramona, E., *Dermatoglyphics and Schizophrenia: A Meta-analysis and Investigation of the Impact of Obstetric Complications upon a-b Ridge Count*, 75 Schizophrenia Research 399 (2005)

Bukelo, M.J., *Palmar Dermatoglyphics in Children with Acute Lymphoblastic Leukemia – A Preliminary Investigation*, 18 J. Forensic Leg. Med. 115 (2011)

Chakraborty, D., et al., *Dermatoglyphic Analysis in Malay Subjects with Bipolar Mood Disorder*, 56 Med. J. Malaysia 223 (2001)

Chintamani, et al., *Qualitative and Quantitative Dermatoglyphic Traits in Patients with Breast Cancer: a Prospective Clinical Study*, 13 BMC Cancer 7 (2007)

Cvjeticanin, M. & A. Polovina, *Quantitative Analysis of Digitopalmar Dermatoglyphics in Male Children*

with Central Nervous System Lesion by Quantification of Clinical Parameters of Locomotor Disorder, 53 Acta Medica Croatica 5 (1999)

Dogramaci, A.C., et al., *Dermatoglyphs in Patients with Beta-thalassemia Major and Their Thalassemia Carrier Parents*, 33 Collegium Antropologicum 607 (2009)

Hiroshi Schiono, *Dermatoglyphics in Medicine*, 7 Am. J. Forensic Med. & Pathology 120 (1986)

Kahn, H.S., et al., *A Fingerprint Marker from Early Gestation Associated with Diabetes in Middle Age*, 38 Int'l J. Epidemiology 101 (2009)

Kelly, B.D., et al., *Neurological Soft Signs and Dermatoglyphic Anomalies in Twins with Schizophrenia*, 19 European Psych. 159 (2004)

Markow, Therese Ann & Kevin Wandler, *Fluctuating Dermatoglyphic Asymmetry and the Genetics of Liability to Schizophrenia*, 19 Psychiatric Research 323 (1986)

Nezhad, Hossein Rezaei & Nasser Mahdavi Shah. *Application of Dermatoglyphic Traits for Diagnosis of Diabetic Type 1 Patients*, 1 Int'l J. Environmental Sci. & Development 36 (2010)

Polovina, S., et al., *Qualitative Dermatoglyphic Traits in Brachial Plexus Palsy*, 31 Collegium Antropologicum 1077 (2007)

Pulliam, T.J. & M.M. Schuster, *Congenital Markers for Chronic Intestinal Pseudo-obstruction*, 90 Am. J. Gastroenterology 922 (1995)

Reed, Terry & John M. Opitz, *Dermatoglyphics in Medicine – Problems and Use in Suspected Chromosome Abnormalities*, 8 Am. J. Med. Genetics 411 (1981)

Rosa, A., et al., *Further Evidence that Congenital Dermatoglyphic Abnormalities Are Associated with Psychosis: A Twin Study*, 28 Schizophrenia Bull. 697 (2002)

Schaumann, B.A. & J.M. Opitz, *Clinical Aspects of Dermatoglyphics*, 27 Birth Defects Original Articles Series 193 (1991)

Schiono, Hiroshi, *Dermatoglyphics in Medicine*, 7 Am. J. Forensic Med. & Pathology 120 (1986)

Shemetova, M.V., et al., *Specific Dermatoglyphic Traits of Patients with Occupational Allergic Bronchitis*, 36 Genetika 100 (2000)

Solhi, H., et al., *Diagnostic Value of Fingerprint Patterns: An Explorative Study on Betathalassemia Diagnosis*, 36 Bangladesh Med. Res Council Bull. 27 (2010)

Tay, J.S., et al., *Dermatoglyphs in Children with Mitral Valve Prolapse*, 22 J. Med. Genetics 135 (1985)

Todd, E.S, *Characterization of Dermatoglyphics in PHOX2B-confirmed Congenital Central Hypoventilation Syndrome*, 118 Pediatrics 408 (2006)

van Oel, C.J., et al., *Differentiating Between Low and High Susceptibility to Schizophrenia in Twins: the Significance of Dermatoglyphic Indices in Relation to Other Determinants of Brain Development*, 52 Schizophrenia Research 181 (2001)

Verbov, Julian, *Clinical Significance and Genetics of Epidermal Ridges – A Review of Dermatoglyphics*, 54 J. Investigative Dermatology 261 (1970)

Weinreb, Herman J., *Fingerprint Patterns in Alzheimer's Disease*, 42 Archives Neurology 50 (1985)

Weizman, Z., et al., *Dermatoglyphic (Fingerprint) Patterns in Coeliac Disease*, 10 J. Pediatric Gastroenterology & Nutrition 451 (1990)