KAMALA D. HARRIS
Attorney General

State of California
DEPARTMENT OF JUSTICE

455 GOLDEN GATE AVENUE, SUITE 11000
SAN FRANCISCO, CA 94102-7004

Public: (415) 703-5500
Telephone: (415) 703-5830
Facsimile: (415) 703-1234
E-Mail: Daniel.Powell@doj.ca.gov

September 12, 2012

Molly Dwyer
Clerk of the Court
Ninth Circuit Court of Appeals
95 Seventh Street
San Francisco, CA 94103-1526

RE: *Elizabeth Aida Haskell, et al. v. Kamala Harris, et al.*
Ninth Circuit Case No. 10-15152

Dear Ms. Dwyer:

Pursuant to Federal Rule of Appellate Procedure 28(j), appellees submit this response to amicus curiae Electronic Frontier Foundation's (EFF) September 11, 2012 letter and also submit the attached article, Katsanis et al., *Characterization of the Standard and Recommended CODIS Markers*, J. Forensic Sciences (forthcoming) and a Letter to the Editor to be published in that same issue.

In response to the California Court of Appeal's decision in *People v. Buza*, 129 Cal. Rptr. 3d 753 (2011), scientists reviewed the existing literature to determine whether the existing core CODIS loci continued to meet the standard that they have "[n]o known association with medical conditions or defects." Katsanis, *et al*., *supra*, at 1. Significantly, among the data that the scientists examined was data from the ENCORE project that is the subject of EFF's letter. *Id*. at 2.

While it is true that recent studies have shown that certain DNA that was thought to be "junk" DNA has a greater role than previously believed, none of those studies have examined the *specific loci* used by the federal government and California in developing a DNA profile. The Katsanis study, however, did examine those specific loci, concluding that "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." *Id*. at 3.

Rather than rely on "popular science articles", *id*. at 1, this Court should be guided by rigorous science that examines the loci that are actually being used by law enforcement officials. Only the Katsanis study examined those loci using all available scientific information and it

"affirm[ed] that individual genotypes are not at present revealing information beyond identification." *Id*. at 3; *see also* Letter to the Editor at 3.

Sincerely,

s/ Daniel J. Powell

DANIEL J. POWELL
Deputy Attorney General
State Bar No. 230304

For KAMALA D. HARRIS
Attorney General

DJP:sc

SA2010100220
20605894.doc

EXHIBIT 1

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, 2012
doi: 10.1111/j.1556-4029.2012.02253.x
Available online at: onlinelibrary.wiley.com

TECHNICAL NOTE

CRIMINALISTICS; JURISPRUDENCE

Sara H. Katsanis,[1] *M.S. and Jennifer K. Wagner,*[2] *J.D., Ph.D.*

Characterization of the Standard and Recommended CODIS Markers*

ABSTRACT: As U.S. courts grapple with constitutional challenges to DNA identification applications, judges are resting legal decisions on the fingerprint analogy, questioning whether the information from a DNA profile could, in light of scientific advances, reveal biomedically relevant information. While CODIS loci were selected largely because they lack phenotypic associations, how this criterion was assessed is unclear. To clarify their phenotypic relevance, we describe the standard and recommended CODIS markers within the context of what is known currently about the genome. We characterize the genomic regions and phenotypic associations of the 24 standard and suggested CODIS markers. None of the markers are within exons, although 12 are intragenic. No CODIS genotypes are associated with known phenotypes. This study provides clarification of the genomic significance of the key identification markers and supports—independent of the forensic scientific community—that the CODIS profiles provide identification but not sensitive or biomedically relevant information.

KEYWORDS: forensic science, genetic identity, DNA typing, forensic genetics, Combined DNA Index System, short tandem repeats

The culmination of the 1996–1997 STR Project was the selection of the 13 core CODIS markers, all highly polymorphic tetra-nucleotide short tandem repeats (STRs). In 2010, the U.S. Federal Bureau of Investigation revisited the panel composition, creating the CODIS Core Loci Working Group to consider the expansion of the core CODIS marker panel to minimize the likelihood of adventitious matches, improve international compatibility for data sharing, and improve the discriminatory power for missing persons cases and familial searching (1). This culminated in the proposal of an additional 11 STRs to be used alternatively in various identification contexts (1). Several criteria were considered in selecting markers for an expanded panel, the first of which being that they have "[n]o known association with medical conditions or defects" (1, p. e52, 2, p. 1). The primary rationale behind the emphasis on the development of panels that contain no association with biomedically relevant phenotypes is clear, as the statutory authority for CODIS itself (the DNA Identification Act of 1994 [3]; DNA Analysis Backlog Elimination Act of 2000 [4]; Justice for All Act of 2004 [5]; and the DNA Fingerprinting Act of 2005 [6]) is restricted to identification purposes. The Department of Justice has reiterated that CODIS profiles are to be "sanitized 'genetic fingerprints' that can be used to identify an individual uniquely, but do not disclose an individual's traits, disorders, or dispositions" (7). Thus, the rationale behind the criterion requires little explanation by the Working Group. On the other hand, the criterion used by the Working Group for the selection and ranking of the markers is unclear, and the literature offers little information relevant to whether (and the extent to which) any of these markers are causally related to phenotypes (1,2). Moreover, a quick review of the literature of linkage analyses and genome wide association studies (GWAS) may yield deceptive and exaggerated reports of linkage with some of these markers, because the number of reports may be simply a relic of the convenient markers' inclusion in commonly used linkage screening panel sets—such as the Marshfield linkage maps (8,9)—and, thus, the results may not be indicative of any actual causal relationship or biological function (10).

Motivated by recent court opinions (11) in which judges (frequently referencing popular science articles [12]) call the phenotypic irrelevance of the CODIS profile into question, we seek to clarify the role of phenotype in the selection criteria of markers. A myriad of important criticisms of forensic DNA analysis including the legal considerations and implications of molecular photofitting and phenotyping (13), the history and substance of criticisms aimed at forensic identification using blood grouping, HLA testing, and more recent methods, the suitability of the fingerprint analogy (which has considerable legal importance), and the appropriateness of selected CODIS markers (2) are each worthy of discussion. The focus of this technical note, however, is to address only the feasibility of creating a DNA database restricted to identity information. As the forensic community grapples with the technical and statistical benefits of the original 13 CODIS loci and the additional 11 loci for prioritization and selection, here, we examine the role of the markers and their regions as elements of the human genome. The selection of markers for identification is an ongoing process varying by population and regulatory control; as such, some markers (e.g.,

[1]Genome Ethics, Law & Policy, Duke Institute for Genome Sciences & Policy, Duke University, 304 Research Drive, Box 90141, Durham, NC 27708.

[2]Division of Translational Medicine and Human Genetics, Center for the Integration of Genetic Healthcare Technologies, Perelman School of Medicine, University of Pennsylvania, 1112 Penn Tower, 399 S. 34th St., Philadelphia, PA 19104.

*Partially funded by grant number P50HG004487-05 from the National Human Genome Research Institute (NHGRI).

Received 14 Mar. 2012; and in revised form 18 June 2012; accepted 30 June 2012.

D6S1043, D14S1434) relevant in select sub-populations are not reviewed herein.

Methods

We used the UCSC Genome Browser (build GRCh37/hg19) (14) to analyze each STR region. We conducted BLAT searches (15) of primers from each STR to locate the precise region of the repeat. We collected data on (i) phenotype and disease associations (GAD view-pack; DECIPHER-full; Online Mendelian Inheritance in Man [OMIM] AV SNPs-full; OMIM genes-full; OMIM pheno loci-full; GWAS catalog-full; RGD human quantitative trait loci [QTL]-full); (ii) genes and gene prediction tracks (UCSC genes-pack; RefSeq-dense); (iii) mRNA and EST tracks (human mRNAs-pack; spliced ESTs-pack); (iv) variation and repeats (common SNPs(132)-pack; simple repeats-pack; microsatellites-full); and (v) regulation (ENCODE Regulation-show; ENC RNA Binding-show; ORegAnno-full; Vista Enhancers-full). We used Ensembl (16) to determine intronic regions and note any reported phenotypic associations for STR genotypes. Disorders associated or linked within 1 kb of the STR were noted; chromosomal anomalies were not noted. To examine the potential relevance of the markers as noncoding genomic elements, we examined sequences for predicted enhancers and noted RNA-binding protein sites as well as predicted DNase I hypersensitivity and transcription binding sites. SNPs documented in dbSNP (build 132) overlapping and linked within 1 kb of the STR were noted. SNPs from the 1000 Genome Project were searched in SNPedia. If the STR was within a gene locus, we noted the gene name and examined the positioning with regard to the surrounding exons. Extragenic STR regions were examined to document proximity to the nearest transcript. We subsequently searched relevant genes in Database of Genotypes and Phenotypes (dbGaP) (17), OMIM (18), and GeneTests (19) to confirm genetic associations and document the availability of a genetic test for any related gene. We also examined the Marshfield linkage maps to determine which markers are used in the human genetic screening panels (20). All searches were conducted in October 2011.

Results and Discussion

Individual genotypes of the 24 STRs were not found to be associated with any documented phenotypes (note exception: DYS391 is on the Y chromosome, which if present in a DNA profile may indicate male sex). None of the 24 STRs are located within protein-coding exons (see Table 1) (see also Ref. 10). Two of the STRs (VWA and D12S391) are colocated on the same arm of chromosome 12 (12p13) within 6 Mb (21,22). Twelve are located within introns of genes, with six of those being genes with known phenotypic associations (see Table 2). Mutations in the six genes are well documented as causative of the corresponding syndromes, but no mutations have been found to be in linkage disequilibrium with any tetra-nucleotide repeat genotypes. Of the intronic STRs, two (FGA and VWA) were within 400 bp of a splice site. All STR loci were associated (within 1 kb) with at least one phenotype according to published GWAS or quantitative trait loci (QTL) studies. TH01 was associated with the most phenotypes (18 traits) ranging from alcoholism (23) and schizophrenia (24) to autosomal recessive spinocerebellar ataxia (25), while DYS391 is believed to be associated only with hairy ears (26). Such genome wide studies often span large regions of the genome; our findings demonstrate that CODIS STR loci are located within such regions, and hence potentially linked to such traits. However, association with these traits does not imply necessarily that individual CODIS marker genotypes are

TABLE 1—*Genomic characterization of CODIS markers.*

	CODIS Marker	Cytogenetic Location	Intragenic or Distance from Nearest Gene	Included in Marshfield Human Genetic Linkage Maps	Number of (#) SNPs (dbSNP Build 132) Within 1 kb
1	D18S51	18q21.33	Intron 1	Included	13
2	FGA	4q28	Intron 3		4
3	D21S11	21q21.1	>100 kb from nearest gene	Removed	7
4	D8S1179	8q24.13	>50 kb from nearest gene	Included	14
5	VWA*	12p13.31	Intron 40		27
6	D13S317	13q31.1	>100 kb from nearest gene	Included	10
7	D16S539	16q24.1	~10 kb from nearest gene	Removed	29
8	D7S820	7q21.11	Intron 1	Included	7
9	TH01	11p15.5	Intron 1		8
10	D3S1358	3p21.31	Intron 20		11
11	D5S818	5q23.2	>100 kb from nearest gene	Removed	8
12	CSF1PO	5q33.1	Intron 6		15
13	D2S1338	2q35	~20 kb from nearest gene	Included	11
14	D19S433	19q12	Intron 1	Included	10
15	D1S1656	1q42	Intron 6	Removed	22
16	D12S391*	12p13.2	~40 kb from nearest gene	Included	18
17	D2S441	2p14	~30 kb from nearest gene	Removed	13
18	D10S1248	10q26.3	~3 kb from nearest gene	Included	16
19	Penta E	15q26.2	Within uncharacterized EST; ~50 kb from nearest gene		19
20	DYS391	Yq11.21	~5 kb from nearest gene		0
21	TPOX	2p25.3	Intron 10		33
22	D22S1045	22q12.3	Intron 4	Included	19
23	SE33	6q14	psedogene, ~30 kb from nearest gene		8
24	Penta D	21q22.3	Intron 4		6

Markers are shown in their relative rank according to Hares (1).
*VWA and D12S391 are colocated on 12p13 within 6 Mb.

TABLE 2—*Reported phenotypic relevance of genomic regions of CODIS markers.*

	CODIS Marker	Gene Name	Disorder(s) Caused by Gene Mutations	Number of (#) Phenotypes Associated Within 1 kb	Predicted DNA Elements
1	D18S51	*BCL2* (B-cell CLL/lymphoma 2)	Leukemia/lymphoma, B-cell	11	ELAV1 binding site
2	FGA	*FGA* (fibrinogen alpha chain)	Congenital afibrinogenemia; hereditary renal amyloidosis; dysfibrinogenemia (alpha type)	17	PABPC1 binding site
3	D21S11	None		1	None
4	D8S1179	None		17	None
5	VWA*	*VWF* (von Willebrand factor)	Von Willebrand disease	12	ELAV1 binding site
6	D13S317	None		5	None
7	D16S539	None		8	None
8	D7S820	*SEMA3A* (sema domain, immunoglobulin domain, short basic domain, secreted (semaphorin) 3A)		8	CELF1, ELAV1 and PABPC1 binding site
9	TH01	*TH* (tyrosine hydroxylase)	Segawa syndrome, recessive	18	ELAVL1, PABPC1 and SLBP binding site
10	D3S1358	*LARS2* (leucyl-tRNA synthetase 2, mitochondria)		15	None
11	D5S818	None		5	None
12	CSF1PO	*CSF1R* (colony stimulating factor 1 receptor)	Predisposition to myeloid malignancy	15	eGFP-GATA2 transcription factor; PABPC1 binding site
13	D2S1338	None		9	None
14	D19S433	*C19orf2* (uncharacterized gene)		7	DNase I hypersensitivity site; SLBP binding site
15	D1S1656	*CAPN9* (calpain 9)		10	PABPC1 binding site
16	D12S391*	None		6	None
17	D2S441	None		6	None
18	D10S1248	None		6	DNase I hypersensitivity site
19	Penta E	EST: BG210743 (uncharacterized EST)		8	None
20	DYS391	None		1	None
21	TPOX	*TPO* (thyroid peroxidase)	Thyroid dyshormonogenesis 2A	5	PABPC1 and SLBP binding site
22	D22S1045	*IL2RB* (interleukin 2 receptor, beta)		11	None
23	SE33	None		9	None
24	Penta D	*HSF2BP* (heat shock factor 2-binding protein)		6	PABPC1 and SLBP binding site

Markers are shown in their relative rank according to Hares (1).
*VWA and D12S391 are colocated on 12p13 within 6 Mb.

predictive or causative of any particular trait. As expected, all regions were sprinkled with documented SNPs (see Table 1), with the region of TPOX having the most (33 SNPs) and the region of FGA having the fewest (four SNPs). Four SNPs (rs3829986 and rs41338945 near CSF1PO, rs34120165 near VWA, and rs28359647 near D1S1656) were among those commonly queried for the 1000 Genome Project; none of these are annotated in SNPedia. None of the STRs overlapped predicted enhancers. Ten of the STRs (CSF1PO, FGA, TH01, TPOX, VWA, D7S820, D18S51, D19S433, D1S1656, and Penta D) lay within predicted RNA-binding protein sites. Two STRs (D19S433 and D10S1248) lay within DNase I hypersensitivity sites and one (CSF1PO) lay within a transcription factor. The role of tetra-nucleotide repeats in RNA binding and DNase I hypersensitivity is unknown, although expanded tetra-nucleotide repeats may destabilize transcription factor binding sites (27). At this time, no correlation has been made between STR repeat sizes in humans and the impact on transcription factor binding. The Marshfield human genetic linkage maps include 14 of the 24 markers, with nine still in use and five identified as "cryptic duplicate markers" and removed from subsequent panels.

The current understanding of the standard and recommended CODIS panels of STR loci summarized here highlights that these markers continue to be of limited significance for assessing phenotypes. Indeed, we found no documentation of individual genotypes for the 24 STRs to be causative of any documented phenotypes either in the literature or in the interrogated databases. Several of the STRs overlay predicted sites for genomic regulation, but there is no evidence that any particular repeat genotypes are 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. The existence of the predicted DNA elements suggests that some STR loci may be involved in genomic regulation. However, even for CODIS marker genotypes statistically associated with biomedically relevant phenotypes, statistical association is not synonymous with positive or negative predictive value (24). While we cannot say that the standard and recommended CODIS markers are wholly absent and forever immune from any implications for potentially sensitive or medically relevant information, we can affirm that individual genotypes are not at present revealing information beyond identification (1,2,5).

References

1. Hares DR. Expanding the CODIS core loci in the United States. Forensic Sci Int Genet 2011;6(1):e52–4.
2. Ge J, Eisenberg A, Budowle B. Developing criteria and data to determine best options for expanding the core CODIS loci. Investig Genet 2012;3:1.
3. DNA Identification Act of 1994, Pub Law 103–322, 108 Stat. 1796, 2065–71.

4. DNA Analysis Backlog Elimination Act of 2000, Pub Law 106–546, 114 Stat. 2726–37.
5. Justice for All Act of 2004, Pub Law 108–405, 118 Stat. 2260.
6. DNA Fingerprinting Act of 2005, Pub Law 109–162, 119 Stat. 2960, 3085
7. 73 Fed. Reg. at 74937.
8. Weber JL. Informativeness of human (dC-dA)n.(dG-dT)n polymorphisms. Genomics 1990;7(4):524–30.
9. Ghebranious N, Vaske D, Yu A, Zhao C, Marth G, Weber JL. STRP screening sets for the human genome at 5 cm density. BMC Genomics 2003;4(6):1–10.
10. Butler JM. Genetics and genomics of core short tandem repeat loci used in human identity testing. J Forensic Sci 2006;51(2):253–65.
11. *People v. Buza*, San Francisco Co. Super. Ct. SCN 207818 (First App. Dist, Ct. of App. Cal., Aug. 4, 2011) at 5 (quoting Haskell v. Brown, 677 F.Supp.2d 1187, 1190 (N.D. Cal. 2009)).
12. Gibbs WW. The unseen genome: gems among the junk. Sci Am 2003;289(5):46–53, 48.
13. Koops B-J, Schellekens M. Forensic DNA phenotyping: regulatory issues. Columbia Sci Technol Law Rev 2008;9:159–202.
14. Kent WJ. BLAT—the BLAST-like alignment tool. Genome Res 2002;12 (4):656–64, http://genome.ucsc.edu/index.html (accessed May 11, 2012).
15. Butler JM, Reeder DJ. Overview of STR fact sheets. Short Tandem Repeat DNA Internet Database. www.cstl.nist.gov/strbase/str_fact.htm (accessed May 11, 2012).
16. Flicek P, Amode MR, Barrell D, Beal K, Brent S, Chen Y, et al. Ensembl 2011. Nucleic Acids Res 2011;39(Database issue):D800–6.
17. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, et al. The NCBI dbGaP database of genotypes and phenotypes. Nat Genet 2007;39(10):1181–6.
18. Online Mendelian Inheritance in Man (OMIM®). McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD, http://omim.org/ (accessed May 11, 2012).
19. GeneTests: medical genetics information resource (database online). Copyright, University of Washington, Seattle, 1993–2011, http://www.genetests.org (accessed May 11, 2012).
20. Broman KW, Murray JC, Sheffield VC, White RL, Weber JL. Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am J Hum Genet 1998;63:861–9, http://research.marshfieldclinic.org/genetics/GeneticResearch/compMaps.asp (accessed May 11, 2012).
21. O'Connor KL, Hill CR, Vallone PM, Butler JM. Linkage disequilibrium analysis of D12S391 and vWA in U.S. population and paternity samples. Forensic Sci Int Genet 2011;5(5):538–40.
22. Gill P, Phillips C, McGovern C, Bright JA, Buckleton J. An evaluation of potential allelic association between the STRs vWA and D12S391: implications in criminal casework and applications to short pedigrees. Forensic Sci Int Genet 2012;6(4):477–86.
23. Dahmen N, Völp M, Singer P, Hiemke C, Szegedi A. Tyrosine hydroxylase val-81-Met polymorphism associated with early-onset alcoholism. Psychiatr Genet 2005;15(1):13–6.
24. Jacewicz R, Szram S, Galecki P, Berent J. Will genetic polymorphism of tetranucleotide sequences help in the diagnostics of major psychiatric disorders? Forensic Sci Int 2006;3:24–7.
25. Breedveld GJ, van Wetten B, te Raa GD, Brusse E, van Swieten JC, Oostra BA, et al. A new locus for a childhood onset, slowly progressive autosomal recessive spinocerebellar ataxia maps to chromosome 11p15. J Med Genet 2004;41:858–66.
26. Lee AC, Kamalam A, Adams SM, Jobling MA. Molecular evidence for absence of Y-linkage of the hairy ears trait. Eur J Hum Genet 2004;12:1077–9.
27. McIvor EI, Polak U, Napierala M. New insights into repeat instability: role of RNA-DNA hybrids. RNA Biol 2010;7:551–8.

Additional information and reprint requests:
Sara H. Katsanis, M.S.
Genome Ethics, Law & Policy
Duke Institute for Human Genome Sciences & Policy
Duke University
304 Research Drive, Box 90141
Durham, NC 27708
E-mail: sara.katsanis@duke.edu

EXHIBIT 2

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, 2012
doi: 10.1111/j.1556-4029.2012.02252.x
Available online at: onlinelibrary.wiley.com

Letter to the Editor—Out with the "Junk DNA" Phrase

Sir,

What started as a clever talk title by Susumu Ohno (1) to describe non-protein-coding DNA (ncDNA) quickly became a ubiquitous phrase ("junk DNA") causing substantial confusion and distraction from a more sophisticated and accurate appreciation of the majority of the human genome that does not encode for proteins. While much of the scientific community rejected the vernacular—with some prominent scholars calling the characterization "ambiguous and even derogatory" (2)—the term persisted widely, in concept and informally, in academic literature and popular media. The scientific intricacies of the many humanomes (i.e., the genome, exome, transcriptome, and proteome) are relatively poorly understood by those outside the relevant disciplines, though those within the relevant disciplines cannot deny the importance of better understanding the molecular and cellular roles of ncDNA. While popular media are picking up on the general scientific sentiment that ncDNA has some importance (see, e.g., [3]), there is considerable confusion about (i) what the "junk" vernacular originally referred, (ii) what the implications of the subsequent scientific rejections of that vernacular are, and (iii) what the current characterizations of CODIS markers non-protein-coding DNA are. This confusion is apparent in important court opinions (e.g., [4–6]). Misconceptions of "junk DNA" are shaping the judiciary's perception of the loci used to in the standard CODIS profile and, subsequently, the judiciary's perception of the privacy implications of a CODIS profile and the appropriateness of the "fingerprint analogy" (e.g., [7,8]).

There was never a consensus among scientists that ncDNA was deserving of the "provocative term" coined by Ohno (9). While the diversity of non-protein-coding regions remained poorly understood for decades, at least four hypotheses explained the maintenance of these seemingly nonfunctional regions of the genome. The "selectionist hypothesis" posited that these regions regulate gene expression (10). The "neutralist hypothesis" posited these regions have no function but are transmitted passively as relics of evolutionary processes (10). The "intragenomic selectionist hypothesis" posited that non-protein-coding regions actively promote their own transmission and accumulate because of their elevated reproduction rate relative to protein-coding regions (10).[1] The "nucleotypic hypothesis" posited that these regions act to maintain structural integrity of the genome (10). When Ohno himself first used the term "junk DNA" to refer to all ncDNA, he had explicitly stated, "Certain untranscribable and/or untranslatable DNA base sequences appear to be useful…" (1, p. 367) Sydney Brennar, a molecular biologist, had distinguished "junk" from "garbage," explaining that, while garbage is worthless, used up, and thrown away, junk is of potential value and stored for unspecified future use (e.g., [9]). The characterization of ncDNA as "junk DNA" ultimately had the effect of "repel[ling] mainstream researchers" from studying it (11, p. 1246). It was not until the 1990s that scientists gave increasing attention to "junk DNA"—along with increased attention to every aspect of the genome spurred by the Human Genome Project—and began to appreciate the diversity of ncDNA "not [as] a single midden heap…but [as] a complex mix of different types of DNA, many of which are vital…" (12, p. 608). Table 1 provides a summary of non-protein-coding elements of the genome.

[1]This hypothesis was accompanied by the coinage of another unfortunate phrase, "selfish DNA."

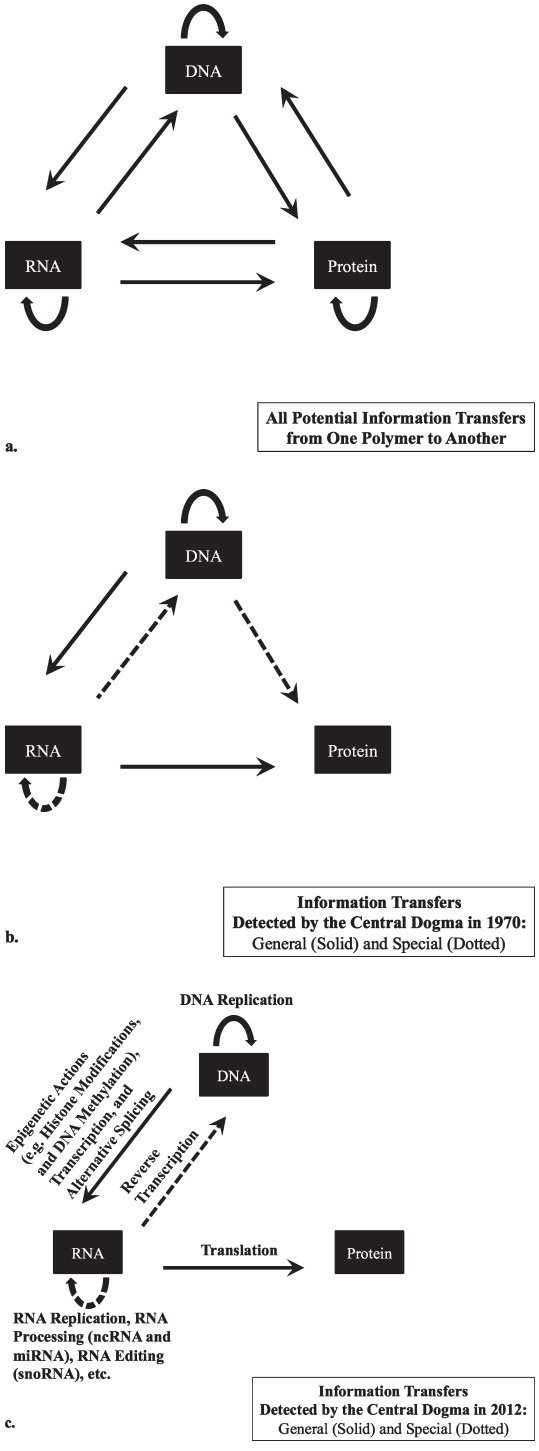

FIG. 1—*Information transfers and the central dogma. Panel (a) shows all possible transfers (adapted from Crick's figure 1 [15]). Panel (b) shows detected transfers in 1970 (adapted from Crick's figure 3 [15]). Panel (c) shows detected transfers in 2011 with notations summarizing some of the complexities of gene expression (adapted from Mattick's figure 1 (18) and Slack's figure 1 [19]).*

TABLE 1—*Summary of non-protein-coding genomic elements.*

Non-Protein-Coding Genomic Element			Brief Description
Transcription regulatory elements			Molecular elements considered typical of gene structure, such as promoters, enhancers, and intronic splicing signals (21)
Introns			Segments of DNA located within genes that interrupt or separate exons from one another
5′ and 3′ untranslated regions		UTRs	Transcribed DNA sequences preceding (5′ UTR) and following (3′ UTR) coding sequences containing regulatory elements, such as binding sites for microRNAs (miRNAs), and polyadenylation signals (22)
RNA-specifying genes	MicroRNAs	miRNAs	Destabilize or inhibit the translation of targeted mRNAs; 19–25 nucleotides in length (23)
	Transfer RNAs	tRNAs	Facilitate translation by transporting specific amino acids to the ribosome; *c.* 80 nucleotides in length (23)
	Ribosomal RNAs	rRNAs	Facilitate the movement of tRNAs along the mRNA during translation; 4 types (18S, 28S, 5.8S, and 5S) (23)
	Spliceosomal RNAs	snRNAs	Facilitate the processing of pre-mRNAs (i.e., help splice introns that are not self-splicing); 5 types (U1, U2, U4, U5, and U6) (23)
	Small Nucleolar RNAs	snoRNAs	Facilitate posttranscription modifications of rRNAs, tRNAs, and snRNAs; 2 types (H/ACA box and C/D box) (23)
	Piwi-Interacting RNAs	piRNAs	Protect the integrity of the genome in germline cells during spermatogenesis; 26–34 nucleotides in length (23)
	RNAse P/MRP genes		Process tRNA and rRNA precursors (23)
	Long noncoding RNAs	lncRNAs	*c.* 200+ nucleotides in length, such as *XIST*, which silences an X chromosome during X-inactivation (23)
Repeat elements	Satellite DNA		DNA sequences often near centromeres and telomeres α-satellite or alphoid DNA, a 171-bp sequence that is repeated in tandem and clustered at the centromeres of all chromosomes. Repeat size of satellite DNA may be between 2 and 2000 bp and the size of the repeat array may be greater than 1000 bp (10,21)
	Minisatellites or Variable Number Tandem Repeats	VNTRs	Repeat units of 10–200 bp clustered into repeat arrays of 10–100 units Found near the telomeres (the terminal ends of chromosomes), but are also distributed across the chromosomes (10,21)
	Microsatellites or Short Tandem Repeats	STRs	Repeat units of 2–5 bp arranged in arrays of 10–100 units (10,21)
	Short Interspersed Nucleotide Elements	SINEs	*c.* 1,500,000 copies of SINEs present in the genome account for more than 10% of the genome (10,24)
	Long Interspersed Nucleotide Elements	LINEs	*c.* 850,000 copies of LINEs present in the genome, account for roughly one-fifth of the genome (10,24)
	Retrovirus-like Elements		*c.* 450,000 copies present in the genome (24)
	Transposons		*c.* 300,000 copies present in the genome (24)
Pseudogenes			Exhibit similarity to genes but lack introns and promoters and contain poly-A tails. Most pseudogenes have lost the ability to be transcribed (10,21,25)

The "-omic revolutions" that are dramatically and rapidly changing our understanding of the genome have not called into question the central dogma per se (as shown in Fig. 1), although they have certainly nuanced it by stressing the importance of noncoding function and have also challenged the conceptualization and definition of a "gene" (e.g., [13]). The components and physical boundaries of genes are no longer clear and discrete. Genes are more than just exons stitched together during transcription and subsequently translated into proteins. For example, in different contexts different combinations of exons may be used rather than all of them. Accordingly, the definition of a gene has been broadened to encompass not only the exonic sequence but also introns and intronic splicing sites, as well as promoters, enhancers, and other cis- and trans-regulatory elements (i.e., the factors located close to and far from the exons, respectively) that contribute to known phenotypes or functions. With the term "gene" increasingly being used to specify not only DNA sequences that encode proteins but also DNA sequences that do not encode protein but do specify RNA transcripts with known function, the term may be increasingly confusing to nonscientists and may be of diminishing operational value to scientists (see, e.g., [13]). With this in mind, we can leave the "junk" vernacular behind and refocus our attention to the current understanding of the human genome's structure and function and, specifically, how the standard and recommended CODIS markers (14,15) are characterized within this context.

Armed with the scientific and technological advances of the last 40–50 years, scientists in 2012 are able to better appreciate the complexities of the informational transfers articulated in 1958 as "the central dogma." (Coincidentally, despite over-generalizations and an array of distinct ideas attributed to it, "the central dogma"—as clarified by Francis Crick in 1970—did *not* stipulate that information transfer was only and always transferred from DNA to RNA to protein, did *not* stipulate that RNA lacked function aside from encoding proteins and "sa[id] nothing about control mechanisms" or gene expression [16, p. 562]). The diverse origins, characteristics, functions, and evolution of non-protein-coding regions of the genome are given increasing attention as scientists move beyond a simple Mendelian (one gene-one trait or disease) model and seek a more holistic understanding of human inheritance. This increased appreciation for non-protein-coding regions of the genome does not, however, inherently give rise to increased significance of the diverse array of particular types of non-protein-coding regions.

Recent court opinions have asserted the markers in the standard CODIS profile were characterized as "junk DNA, because

'they are thought not to reveal anything about trait coding'" (e.g., [17, p. 5]). However, the 13 standard CODIS loci were attributed (indeed, burdened) with the label "junk DNA" because they are all microsatellites, and hence non-*protein*-coding. Indeed, the phrase "trait coding" itself reflects a dearth of genetic literacy among the legal profession. That those 13 specific loci—as well as the recent recommendation of 11 additional loci—were chosen for inclusion in a panel designed for identification purposes with an emphasis placed on a lack of association with known phenotypes (14,15) is an entirely separate issue from the loci being non-protein-coding elements. Accordingly, it is appropriate to encourage the discontinued characterization that CODIS loci are "junk DNA" (see also [18]). It is also appropriate to warn nonscientists that to imply the CODIS loci are each or collectively involved in gene expression and are now important for a wide array of traits and conditions of biomedical relevance is unfounded (19).

Selection of loci used for identification purposes is not a permanent, unalterable decision. Rather, it is possible for the forensic science community to revisit such decisions periodically and substitute markers in the event statistical associations, causal relationships, or predictive value for biomedically relevant phenotypes become known. Selection of markers for identity should be directed by the inherent usefulness of each marker to discriminate individuals and the experimental ease of amplification, rather than the negative qualitative value of the marker in detecting phenotype. Moreover, the arbitrariness of marker selection must be kept in mind—which phenotypes are considered "sensitive" or "medically relevant" are themselves subjective determinations and not universally agreed. Normative arguments surrounding the use of genetic information for molecular photofitting or phenotyping or the storage and unrestricted analysis of DNA samples can and must be kept separate from questioning whether it is scientifically possible to select a set of markers that are of value restricted to identification purposes. The scientific community, should it choose to do so, can relegate the "junk DNA" phraseology to the history books and forge ahead to a more nuanced understanding of genomics and the central dogma.

Acknowledgments

This work was funded by Grant No. P50HG004487-05 and Grant No. K99HG006446 from the National Human Genome Research Institute (NHGRI). The content is solely the author's and does not necessarily represent the official views of the NHGRI or the University of Pennsylvania.

References

1. Ohno S. So much 'junk' DNA in our genome. Brookhaven Symp Biol 1972;23:366–70.
2. Brosius J, Gould SJ. "On Genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA". Proc Natl Acad Sci U S A 1992;89(22):10706–10.
3. Gibbs WW. The unseen genome: gems among the junk. Sci Am 2003;289(5):46–53.
4. *People v. Buza*, San Francisco Co. Super. Ct. SCN 207818 (First App. Dist, Ct. of App. CA, Aug. 4, 2011) at 22 (citing Gibbs' "Gems among the Junk" Sci Amer 2003 and explaining that the "quantity and nature" of information decipherable from the CODIS profile "will undoubtedly increase").
5. *United States v. Mitchell*, 652 F.3d 387 (3rd Cir (Pa) 2011), Dissenting opinion at 19 (stating "…it is little comfort that only so-called 'junk DNA' is used to compile a suspect's DNA profile" and noting "'with advances in technology, junk DNA may reveal far more extensive genetic information.' United States v. Kriesel, 379 F.3d 941, 947 (9th Cir. 2007).")
6. *State v. Abernathy*, No. 3599-9-11 (Vt. Super. Ct. June 1, 2012).
7. *United States v. Shavlovsky*, 2012 WL 652672 (E.D. Cal. 2012).
8. *United States v. Kriesel*, 508 F.3d 941, 947-948 (9th Cir. 2007).
9. Rabinow P. The anthropology of reason. Anthropol Today 1992;8(5):7–10, 7–8, paraphrasing Syndey Brenner.
10. Graur D, WH L. Fundamentals of molecular evolution, 2nd edn. Sunderland, MA: Sinauer, 2000;14, 274–5, 386–7, 392-4.
11. Makalowski W. Genomics. Not junk after all. Science 2003;300 (5623):1246–7.
12. Nowak R. Mining treasures from 'Junk DNA'. Science 1994;263 (5147):608–10.
13. Gingeras TR. Origin of phenotypes: genes and transcripts. Genome Res 2007;17:682–90.
14. Hares DR. Expanding the CODIS core loci in the United States. Forensic Sci Int Genet 2012;61(1):e52–e54.
15. Ge J, Eisenberg A, Budowle B. Developing criteria and data to determine best options for expanding the core CODIS loci. Investig Genet 2012;3:1. Doi: 10.1186/2041-2223-3-1.
16. Crick F. Central dogma of molecular biology. Nature 1970;2278 (5258):561–3.
17. *People v. Buza*, San Francisco Co. Super. Ct. SCN 207818 (First App. Dist, Ct. of App. CA, Aug. 4, 2011) at 5 (quoting Haskell v. Brown, 677 F.Supp.2d 1187, 1190 (N.D.Cal. 2009)).
18. Kaye DH. Please let's bury the junk: the CODIS loci and the revelation of private information. NW Univ Law Rev 2007;102:70–81.
19. Katsanis SH, Wagner JK. Characterization of the standard and recommended CODIS markers. J Forensic Sci 2012; doi: 10.1111/j.1556-4029.2012.02253.x [Epub ahead of print].
20. Mattick JS. Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. BioEssays 2003;25:930–9.
21. Slack FJ. Regulatory RNAs and the demise of 'junk' DNA. Genome Biol 2006;7(9):328. Doi: 10.1186/gb-2006-7-9-328.
22. Maroni G. Molecular and genetic analysis of human traits. Malden, MA: Blackwell Science, 2001;57, 58–61, 61–65, 68, 82.
23. Neilson JR, Sandberg R. Heterogeneity in mammalian RNA 3′ end formation. Exp Cell Res 2010;316(8):1357–64.
24. Wright MW, Bruford EA. Naming 'junk': human non-protein coding RNA (ncRNA) gene nomenclature. Hum Genomics 2011;5(2):90–8.
25. Jobling MA, Hurles ME, Tyler-Smith C. Human evolutionary genetics: origins, peoples & disease. New York, NY: Garland Publishing, 2004;31.
26. Pink RC, Wicks K, Caley DP, Punch EK, Jacobs L, Carter DRF. Pseudogenes: pseudo-functional or key regulators in health and disease? RNA 2011;17:792–8.

Jennifer K. Wagner,[1] J.D., Ph.D.
[1]Center for the Integration of Genetic Healthcare Technologies, University of Pennsylvania, Philadelphia, PA, 19104.
E-mail: jennifer.wagner@uphs.upenn.edu

CERTIFICATE OF SERVICE

Case Name: **Elizabeth Aida Haskell, et al. v. Kamala Harris, et al.** No. **10-15152**

I hereby certify that on September 12, 2012, I electronically filed the following documents with the Clerk of the Court by using the CM/ECF system:

28(j) LETTER

I certify that **all** participants in the case are registered CM/ECF users and that service will be accomplished by the CM/ECF system.

I declare under penalty of perjury under the laws of the State of California the foregoing is true and correct and that this declaration was executed on September 12, 2012, at San Francisco, California.

Susan Chiang	s/ Susan Chiang
Declarant	Signature

20637391.doc