THE CHALLENGES OF FORENSIC GENEALOGY: DIRTY DATA, ELECTRONIC EVIDENCE, AND PRIVACY CONCERNS

DIVYA RAMJEE[†] & KATELYN RINGROSE[††]
ACKNOWLEDGEMENTS[†††]

ABSTRACT

While forensic genealogy continues to gain popularity as a law enforcement tool for solving cold cases, discrepancies in testing and evidentiary standards, as well as ethical and privacy issues, continue to plague the practice. This Article examines the investigatory role of genetic information and the various methods by which genetic information can be collected, used, or shared, including by law enforcement. In the era of Big Data, we must understand the limitations posed by the reliability and accuracy of information included in private and publicly available genealogy databases and how those limitations compete with the desire to implement valid machine learning algorithms in the fields of criminology and law. Realizing that advancements in science often outpace regulatory legislation, this Article addresses ways in which private and publicly available genealogy services can safeguard genetic information, including associated identifying metadata. Furthermore, this Article sets forth policy recommendations that consider the importance of enhancing investigative techniques while ensuring appropriate evidentiary standards and Fourth Amendment protections.

TABLE OF CONTENTS

[†] Divya Ramjee is a doctoral candidate and adjunct professor at American University in Washington, D.C., in the Department of Justice, Law & Criminology. Her research interests include cybercrime and intellectual property crime, biotechnology, and the application of machine learning methodology in the field of criminology. *This Author is currently employed by the U.S. Department of Justice, and the views expressed in this Article are exclusively those of the Author and do not necessarily represent the views of the U.S. Department of Justice, its components, or the United States.*

[††] Katelyn Ringrose, JD, is the Christopher Wolf policy fellow at the Future of Privacy Forum—focusing on health and genetic privacy. Her publications have focused on the Fourth Amendment, health IoT-enabled devices, and more. *The views herein do not necessarily reflect those of Future of Privacy Forum supporters or board members.*

[†††] Thank you to Dr. Thomas Zeitzoff, Associate Professor at American University, and Dr. Tricia Bacon, Associate Professor at American University, for their invaluable advice and support.

INTRODUCTION

Recently, at-home genetic testing kits sold by direct-to-consumer (DTC) genetic testing services have become incredibly popular, with more than 26 million consumers having provided genetic material to DTC companies for genealogical purposes.[1] As the use of genealogy services grows, so too does law enforcement interest in accessing this genetic information for comparative purposes.[2] When used appropriately, genetic testing and analysis can be a powerful tool to exonerate the innocent, assist in the apprehension of criminals, and help identify remains.[3] However, concerns abound regarding the conceivable evidentiary value of genetic information processed by commercial entities and the potential privacy violations posed by law enforcement access to public or commercial genetic databases.[4]

Although genealogical genetic testing provides law enforcement with investigative advantages, there are still numerous scientific, legal, and policy issues that must be addressed.[5] Many consumers submitting their DNA to commercial databases or uploading their genetic profile to

[1] Antonio Regalado, *More Than 26 Million People Have Taken an At-Home Ancestry Test*, MIT TECH. REV. (Feb. 11, 2019), https://www.technologyreview.com/2019/02/11/103446/more-than-26-million-people-have-taken-an-at-home-ancestry-test/ (noting that these numbers do not take into account the fact that several DTCs and publicly available genealogy databases allow consumers to upload their genetic data if they have been genotyped by another company).

[2] Michael D. Edge & Graham Coop, *Attacks on Genetic Privacy Via Uploads to Genealogical Databases*, ELIFE (Jan. 7, 2020), https://elifesciences.org/articles/51810.

[3] Hannah Fry, *A DNA Match Brings Relief to Linda O'Keefe's Sister, Four Decades After Girl's Slaying*, L.A. TIMES (Feb. 22, 2019, 7:05 AM), https://www.latimes.com/local/lanow/la-me-linda-okeefe-cold-case-20190221-story.html; *see also Recent Cold Case Spotlights*, PROJECT: COLD CASE, https://www.projectcoldcase.org (last visited Nov. 10, 2020) (explaining that the technological advancements of DNA testing have led to a significant number of cold case arrests).

[4] Sara Debus-Sherrill & Michael B. Field, *Familial DNA Searching- an Emerging Forensic Investigative Tool*, 59 SCI. & JUST. 20, 21, 27 (2019).

[5] *Id.*

publicly available genealogy services do so to learn about their ancestry or medical history—and do not necessarily realize that companies, absent privacy commitments, can share these genetic data for alternate purposes with third parties, particularly law enforcement officials and agencies.[6]

This Article applies the Fourth Amendment, the "third-party doctrine," and the Electronic Communications Privacy Act (ECPA) and determines not only that ascertaining direct matches of genetic information held within privately- and publicly-held genetic databases necessitates warrants but also that familial DNA searching (FDS) poses numerous additional concerns that should be addressed through regulation. This Article encourages comprehensive policy conversations between multiple stakeholders, including DTC companies, publicly available genealogy services, law enforcement, privacy advocates, and legislators, regarding law enforcement's use of publicly available and private genetic databases.

In Part I, this Article examines genetic information at large, including the role of law enforcement databases, scientific issues with FDS, and regulatory protections offered by the main categories of genetic testing databases. Part II analyzes how scientific advancements outpace current regulatory schemes pertaining to the use of genetic information by law enforcement and how regulatory efforts currently treat sensitive personal information. Part III describes how DTC genetic services and publicly available genealogy services may incorporate stronger provisions within their privacy policies and practices to appropriately safeguard consumer data against unreasonable law enforcement access. This Article concludes with various policy recommendations that consider the importance of crime-solving while also ensuring evidentiary standards and Fourth Amendment protections related to law enforcement access to FDS results.

I. GENETIC INFORMATION: CODIS AND BEYOND

Data sharing provisions for DTC databases and publicly available genealogy databases are generally determined by a company's privacy policy, which outlines each company's respective compliance standards for various law enforcement data requests (generally a subpoena, warrant, or legal order, although some companies will also comply with more informal requests).[7] Publicly available genealogy databases contain genetic profiles willingly uploaded by consumers to search for possible ancestral links to other consumers already included in these databases. The open crowd-sourced nature of publicly available genealogy databases allows for access by almost any individual or entity, including law

[6] *See Your Privacy*, ANC., https://www.ancestry.com/cs/legal/privacystatement (last visited Nov. 10, 2020).

[7] *Id.*

enforcement personnel, absent any restrictions imposed by the public genealogy service.[8]

This expansion into publicly available genealogy databases increases the scope of DNA profile searches for law enforcement beyond their present capabilities.[9] However, it is important not to overstate current law enforcement use of such databases; leading DTCs have reported only a handful of law enforcement access requests for genetic data.[10] While some publicly available genealogy databases take a more cooperative approach to acquiescing to law enforcement requests, others are not transparent about their granting or denial of law enforcement access—it must be emphasized that the services that issue annual transparency reports regarding law enforcement access and require warrants to access genetic data are at the forefront of genetic privacy in this space.[11] While company-issued transparency reports and restrictions are an important part of safeguarding genetic information, more regulation is needed as law enforcement begins to consider the potential investigatory power of utilizing consumer datasets.

Studies funded by the U.S. Department of Justice (DOJ), the National Institute of Justice, and the National Criminal Justice Reference Service have noted that numerous jurisdictions across the United States have expressed a growing interest in the use of FDS to aid in criminal investigations.[12] To grasp the complexity and breadth of data sharing and data privacy issues currently at play, genetic testing data sets are best categorized under four main categories: (1) DTC genetic testing databases (including ancestral and genealogical tests, as well as health and wellness tests); (2) publicly available genealogy databases; (3) databases wholly intended for law enforcement officials; and (4) medical genetic testing databases (including regulated medical tests offered by DTCs, in addition to formal diagnostic testing offered only by licensed medical professionals).

Law enforcement-controlled databases do not require a warrant or subpoena for lawful access; however, to engage in database searches, agents need to comply with local laws, professional norms, quality assur-

[8] Matthias Gafni & Lisa M. Krieger, *Here's the 'Open-Source' Genealogy DNA Website that Helped Crack the Golden State Killer Case*, E. BAY TIMES, https://www.eastbaytimes.com/2018/04/26/ancestry-23andme-deny-assisting-law-enforcement-in-east-area-rapist-case/ (Sept. 21, 2018, 11:32 AM).

[9] Fry, *supra* note 3.

[10] Ancestry and 23andMe both report few law enforcement access requests for genetic data, with most requests the services receive being for user account information, like an individual's name or phone number to investigate credit card misuse or fraud. *See 23andMe Guide for Law Enforcement*, 23ANDME, https://bit.ly/3eC72Fk (last visited Nov. 10, 2020); *Ancestry Transparency Report July 2020*, ANC. (July 10, 2020), https://www.ancestry.com/cs/transparency.

[11] *See, e.g.*, *23andMe Guide for Law Enforcement*, *supra* note 10.

[12] Debus-Sherrill & Field, *supra* note 4, at 20, 27.

ance standards, efficacy standards, and other rules and regulations.[13] Law enforcement personnel routinely utilize the Federal Bureau of Investigation (FBI) Combined DNA Index System (CODIS), authorized and established by the 1994 DNA Identification Act.[14] CODIS maintains DNA profiles collected at three jurisdictional levels: National DNA Index System (NDIS), maintained by the FBI; State DNA Index Systems (SDIS), maintained by state-level forensic laboratories; and Local DNA Index Systems (LDIS), maintained by local-level forensic laboratories.[15]

This Article is restricted to an analysis of law enforcement's use of DTC and publicly available genealogy databases, as CODIS is governed by its own norms and medical genetic testing databases are controlled by insurers, physicians, and the federal government.[16] Many DTC companies are now moving into the medical genetic testing market, raising additional concerns regarding the necessity of increased safeguards to HIPAA protected health information.[17] There are a number of advantages in permitting law enforcement access to DTC and publicly available genealogy databases.[18] Statistics from the National Registry of Exonerations indicate that exonerations based on renewed DNA testing accounted for 21% of exonerations from 1989 through 2017, suggesting that the increased use of newer DNA technologies and data sets may aid in future exoneration efforts.[19] Additionally, the popularity of searching DTC data sets for investigative purposes has significantly increased since authorities in Sacramento, California, used publicly available genealogy database GEDMatch to apprehend Joseph DeAngelo in April 2018.[20] DeAngelo, better known as the Golden State Killer, is accused of twelve murders and forty-five rapes that took place from 1976 to 1986.[21] More recently, in December 2019, search of publicly available genealogy databases aided the DNA Doe Project, a nonprofit organization that partners

[13] JULIE E. SAMUELS ET AL., URB. INST., COLLECTING DNA AT ARREST: POLICIES, PRACTICES, AND IMPLICATIONS, FINAL TECHNICAL REPORT 3, 16 (2013).

[14] *See id.* at 3, 7; *see also* 34 U.S.C. § 12592 (2018) ("Index to Facilitate Law Enforcement Exchange of DNA Identification Information" which empowers the FBI director to establish an index of DNA identification records).

[15] SAMUELS ET AL., *supra* note 13, at 3.

[16] *See id.*

[17] *See* D.H. Kaye, *Please, Let's Bury the Junk: The CODIS Loci and the Revelation of Private Information*, 102 NW. U. L. REV. COLLOQUY 70, 71 (2007).

[18] Christi J. Guerrini et al., *Should Police Have Access to Genetic Genealogy Databases? Capturing the Golden State Killer and Other Criminals Using a Controversial New Forensic Technique*, PLOS BIOLOGY (Oct. 2, 2018), https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2006906.

[19] THE NAT'L REGISTRY OF EXONERATIONS, EXONERATIONS IN 2017 at 5 (2018).

[20] *See* Ryan Lillis et al., *'Open-Source' Genealogy Site Provided Missing DNA Link to East Area Rapist, Investigator Says*, SACRAMENTO BEE, https://www.sacbee.com/news/local/crime/article209987599.html (Apr. 28, 2018, 7:45 AM). The Authors of this Article are following unfolding events related to the apprehension of the Golden State Killer and intend to publish further updates on this particular topic.

[21] *Id.*; Guerrini et al., *supra* note 18.

with law enforcement, in identifying the remains of a missing person from Marion County, Ohio.[22]

Law enforcement access to DTC and publicly available genealogy databases could also potentially counterbalance issues of racial disparities associated with CODIS. Criminology scholars have long evidenced racial disparities in arrests and incarceration rates in the United States.[23] In an amicus brief, the Council for Responsible Genetics noted that, in 2010, Black Americans accounted for approximately 27% of adult arrests at a time when the adult Black population was only 12%.[24] These disparities ultimately translate into the disproportionate number of CODIS genetic profiles from racial minorities, particularly Black Americans and Latinos, as those who are arrested are required to provide DNA samples for CODIS inclusion.[25] These DNA specimens are collected after arrest or arraignment, and some states do not automatically expunge profiles upon failure to convict or case dismissal.[26] The Council noted that while many arrestees are Black and Latino, thereby populating DNA databases with genetic information from minority racial groups, they are often not convicted.[27]

While the potential to lower racial disparities is of overall net benefit, there are numerous negatives associated with law enforcement access to privately held genetic information. For example, consequences may exist for family members who unfairly become targets for supplementary investigation.[28] Researchers argue that this risk falls disproportionately on ethnic groups currently overrepresented in state and federal databases.[29] Thus, familial searches of these law enforcement databases risk exacerbating racial disparities by casting suspicion upon innocent com-

[22] *Success Stories: Marion County Jane Doe 1987*, DNA DOE PROJECT (July 1, 2019), https://dnadoeproject.org/case/marion-county-jane-doe-1987/.

[23] Casey T. Harris et al., *Are Blacks and Hispanics Disproportionately Incarcerated Relative to Their Arrests? Racial and Ethnic Disproportionality Between Arrest and Incarceration*, 1 RACE & SOC. PROBS. 187, 188–94 (2009); *see* Douglas A. Smith et al., *Equity and Discretionary Justice: The Influence of Race on Police Arrest Decisions*, 75 J. CRIM. L. & CRIMINOLOGY 234, 236 (1984).

[24] Brief of Council for Responsible Genetics as Amici Curiae Supporting Respondent at 5, Maryland v. King, 569 U.S. 435 (2013) (No. 12-207).

[25] *See* Harriet A. Washington, *Base Assumptions? Racial Aspects of US DNA Forensics*, *in* GENETIC SUSPECTS: GLOBAL GOVERNANCE OF FORENSIC DNA PROFILING AND DATABASING 63, 75–76 (Richard Hindmarsh & Barbara Prainsack eds., 2010); Jonathan Kahn, *Race, Genes, and Justice: A Call to Reform the Presentation of Forensic DNA Evidence in Criminal Trials*, 74 BROOK. L. REV. 325, 326–330, (2009).

[26] *See* Michael T. Risher, *Racial Disparities in Databanking of DNA Profiles*, ACLU N. CAL., https://www.aclunc.org/sites/default/files/racial_disparities_in_databanking_dna_profiles.pdf (last visited Nov. 11, 2020).

[27] *See* Robert D. Crutchfield et al., *Racial and Ethnic Disparity and Criminal Justice: How Much is too Much?,* 100 J. CRIM. L. & CRIMINOLOGY 903, 921 (2010); *see also* Michael Tonry, *Racial Politics, Racial Disparities, and the War on Crime*, 40 CRIME & DELINQ. 475, 475–80 (1994).

[28] Rori V. Rohlfs et al., *The Influence of Relatives on the Efficiency and Error Rate of Familial Searching*, PLOS ONE (Aug. 14, 2013), https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0070495.

[29] *Id.*

munity members and further straining community–police relations in minority communities.[30] Expanding the scope of law enforcement searches to include DTC and publicly available genealogy databases could provide a mechanism to counter the racial skew presented by CODIS-only searches.[31]

Researchers today believe that it is nearly possible to identify every white American through familial matching, with more matches accruing over time.[32] According to a recent study using MyHeritage's database, there is a 60% chance that one of the approximately 1.3 million DNA profiles currently in the database is a third cousin or closer relative of a person of European descent in the United States.[33] Increasing the number of white Americans available for law enforcement searches is potentially a positive method to counter the skewed representation of racial minorities within CODIS, though others rightly argue that simply increasing the sheer number of searchable profiles does little to end racial issues tied to biased policing and surveillance in minority communities.[34] Although increasing the available data set for potential suspects may lead to more arrests of European-descended suspects, racial minorities will not necessarily be arrested with any less frequency.[35] Nor does the use of DTC and publicly available genealogy data sets necessarily solve police–citizen interactions or the racial disparities that may exacerbate those interactions.

CODIS retains DNA profiles from convicted offenders as well as unknown profiles attributed to the putative perpetrator(s) of a crime.[36] As

[30] Erin Murphy, *Relative Doubt: Familial Searches of DNA Databases*, 109 MICH. L. REV. 291, 321–23 (2010); *see also* Becky Pettit & Bruce Western, *Mass Imprisonment and the Life Course: Race and Class Inequality in U.S. Incarceration*, 69 AM. SOCIO. REV. 151, 164–65 (2004); Robert D. Crutchfield et al., *Racial Disparity in Police Contacts*, 2 RACE & JUST. 179, 181–83 (2012).

[31] *See* Jennifer K. Wagner, *DNA, Racial Disparities, and Biases in Criminal Justice: Searching for Solutions*, 27 ALB. L.J. SCI. & TECH. 95, 123–25 (2017); *see also* Natalie Ram et al., *Genealogy Databases and the Future of Criminal Investigation*, 360 SCI. 1078, 1078 (2018).

[32] *See* Yaniv Erlich et al., *Identity Inference of Genomic Data Using Long-Range Familial Searches*, 362 SCI. 690, 690–694 (2018).

[33] *Id.* at 690 (also indicating a similar chance of familial identification at the same level for Americans of European descent when conducting searches with publicly available genealogy database GEDMatch. When the researchers attempted to determine the chances of a similar level relative in MyHeritage's database for Americans of sub-Saharan heritage, the percentage decreased to 40%).

[34] *See* Risher, *supra* note 26; Ram et al., *supra* note 31, at 1078–79; *see also* Jill S. Barnholtz-Sloan et al., *Examining Population Stratification via Individual Ancestry Estimates Versus Self-Reported Race*, 14 CANCER EPIDEMIOLOGY BIOMARKERS & PREVENTION 1545, 1546–50 (2005).

[35] Dorothy Roberts, *Collateral Consequences, Genetic Surveillance, and the New Biopolitics of Race*, 54 HOW. L.J. 567, 582–84 (2011); *see also* Jeffrey Rosen, *Genetic Surveillance for All*, SLATE (Mar. 17, 2009, 4:52 PM), http://beck2.med.harvard.edu/week10/Rosen%202009.pdf.

[36] *Frequently Asked Questions on CODIS and NDIS*, FED. BUREAU OF INVESTIGATION, https://www.fbi.gov/services/laboratory/biometric-analysis/codis/codis-and-ndis-fact-sheet (last visited Nov. 11, 2020) (providing that CODIS maintains these profiles in accordance with a number of privacy standards. For example, all CODIS computers are located in physically secure spaces, and all laboratory communications occur over private networks accessible to only criminal justice agencies approved by the FBI).

of September 2020, CODIS contains 19,500,814 DNA profiles, though only 2.67% of these profiles have actually assisted law enforcement investigations.[37] In comparison, according to a study from Massachusetts Institute of Technology, more than 26 million consumers have already taken at-home genetic testing kits available from DTC services.[38] Those consumers can port their genetic profiles into publicly available genealogy databases, like the platform GEDMatch, which houses around 1.3 million profiles.[39] Therefore, it is understandable why law enforcement agencies would be interested in accessing the larger pool of consumer genetic profiles held by DTCs and publicly available genealogy services.

A. Genealogy Testing: Differing Processing and Interpretation Standards

To reiterate, processed genetic information is stored and maintained within a number of different types of databases.[40] DTC databases retain the largest amount of individual genetic information, but publicly available genealogy databases, law enforcement databases, and medical databases also each house millions of DNA profiles that are accrued through different processes and regulated by different mechanisms.[41]

[37] *CODIS - NDIS Statistics*, FED. BUREAU OF INVESTIGATION, https://www.fbi.gov/services/laboratory/biometric-analysis/codis/ndis-statistics#Tables (last visited Nov. 11, 2020).

[38] Regalado, *supra* note 1 (noting that these numbers do not account for users that test with more than one consumer company).

[39] Jennifer Lynch, *Genetic Genealogy Company GEDmatch Acquired by Company with Ties to FBI & Law Enforcement—Why You Should Be Worried*, ELEC. FRONTIER FOUND. (Dec. 10, 2019), https://www.eff.org/deeplinks/2019/12/genetic-genealogy-company-gedmatch-acquired-company-ties-fbi-law-enforcement-why.

[40] *See* Regalado, *supra* note 1.

[41] *See infra* Table 1.

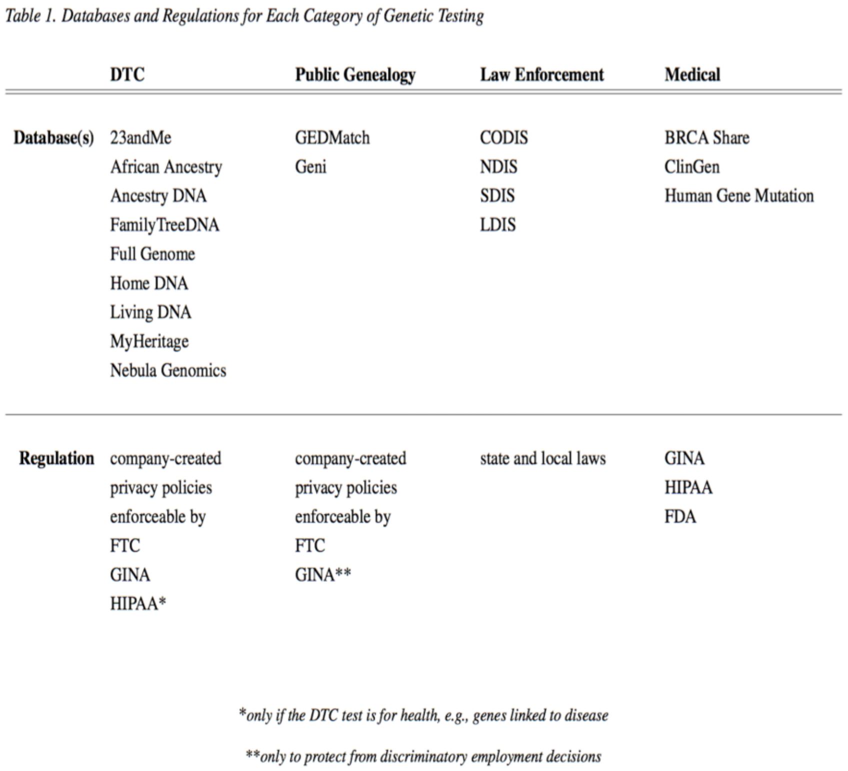

Table 1. Databases and Regulations for Each Category of Genetic Testing

	DTC	Public Genealogy	Law Enforcement	Medical
Database(s)	23andMe African Ancestry Ancestry DNA FamilyTreeDNA Full Genome Home DNA Living DNA MyHeritage Nebula Genomics	GEDMatch Geni	CODIS NDIS SDIS LDIS	BRCA Share ClinGen Human Gene Mutation
Regulation	company-created privacy policies enforceable by FTC GINA HIPAA*	company-created privacy policies enforceable by FTC GINA**	state and local laws	GINA HIPAA FDA

**only if the DTC test is for health, e.g., genes linked to disease*

***only to protect from discriminatory employment decisions*

TABLE 1. *Databases and Regulations for Each Category of Genetic Testing*[42]

Publicly available genealogy databases, as well as DTC databases that might lack staunch protections against unwarranted access, provide law enforcement with the unparalleled ability to conduct extensive familial searches on a large portion of the population.[43] However, some privacy advocates note that companies who provide genetic services to consumers or allow consumers to upload their profiles are largely unregulated, particularly regarding quality assurance and standards of sharing personally identifiable information with law enforcement and other third parties.[44] While other personally identifiable information (e.g., social security numbers, cookie identifiers, or passport numbers) can be reasonably stripped of identifying information (including metadata) and pseudonymized, presently, there are no accepted de-identification standards

[42] *See 23andMe Guide for Law Enforcement*, *supra* note 10; SAMUELS ET AL, *supra* note 13, at 3.

[43] *See* MICHAEL B. FIELD ET AL., ICF, STUDY OF FAMILIAL DNA SEARCHING POLICIES AND PRACTICES: CASE STUDY BRIEF SERIES 11–12 (2017); *see also* Ram et al., *supra* note 31, at 1078–79.

[44] *See* Ram et al., *supra* note 31, 1078–79.

for genetic information.[45] Researchers note that "[w]henever genetic samples are involved re-identification will be possible."[46]

The Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, which does not extend to genetic testing for genealogical purposes, does include a de-identification standard that allows appropriately de-identified data to be shared and sold.[47] De-identification requires the removal of eighteen specified identifiers or a determination that the risk of re-identification is minimal based on a risk-assessment certification process vetted by an expert.[48] Stripping genetic data of identifiers associated with phenotypic properties or disease traits may help in de-identifying the data but may lessen the research value and utility of the information.[49] Furthermore, removing all data that could possibly risk re-identification of genetic information would be an excessively exhausting process, both in terms of the scientific limitations and computing power.[50] Overall, it is arguable whether genetic information can ever truly be de-identified to an acceptable standard. If so, that data may be effectively stripped of its usefulness for the broader medical research community. Certainly, given the difficulty of stripping genetic data of identifiers, current law enforcement access to genetic information is not limited to de-identified data.

Pseudonymization, whereby particular genetic sequences are "tokenized" into coded descriptions of said sequences,[51] could potentially allow for genetic data sets to retain their viability for commercial and research uses while following de-identification protocols. Encryption of these tokenized data sets would provide an additional level of security, as the information is either inaccessible or unusable without a designated key.[52] Tokenization and other pseudonymization techniques often depend on patented approaches or trade secrets for removing identifying genetic information.[53] For instance, one company currently maintains a patent for a process that "strip[s] away phenotypic and identifying data from genomics records without fully severing the link between clinical and genomics data to facilitate efficient research."[54] However, these techniques

[45] Jeantine E. Lunshof et al., *From Genetic Privacy to Open Consent*, 9 NATURE REVS. GENETICS 406, 407 (2008).

[46] *Id.* at 410.

[47] AMANDA K. SARATA ET AL., CONG. RSCH. SERV., R44026, GENOMIC DATA AND PRIVACY: BACKGROUND AND RELEVANT LAW 7 (2015).

[48] *Id.*

[49] *See* Neil Versel, *IQvia Launches Genomics Technology Platform*, GENOMEWEB (Mar. 5, 2019), https://www.genomeweb.com/informatics/iqvia-launches-genomics-technology-platform.

[50] *See* Lunshof et al., *supra* note 45, at 409–10.

[51] Beáta Megyesi et al., *Learner Corpus Anonymization in the Age of GDPR: Insights from the Creation of a Learner Corpus of Swedish*, 36 NEALT PROC. SERIES 47, 54 (2018).

[52] *See* KRISTIN FINKLEA, CONG. RSCH. SERV., R44187, ENCRYPTION AND EVOLVING TECHNOLOGY: IMPLICATIONS FOR U.S. LAW ENFORCEMENT INVESTIGATIONS 1 (2016) (outlining how encryption has become a barrier to law enforcement, particularly in the mobile phone context).

[53] *See* Megyesi et al., *supra* note 51, at 54.

[54] Versel, *supra* note 49.

can be costly and time consuming, and the lack of privacy and security regulations for genetic information has not encouraged companies to invest in resources to maintain such standards.[55] While pseudonymization might be a privacy-preserving method for sharing genetic data with researchers, law enforcement interest in pseudonymous data would presumably be low—law enforcement is seeking to identify suspects on the basis of genetic information and de-identified data would have little to no utility.[56] Therefore, while privacy-preserving commercial techniques like de-identification and pseudonymization can yield benefits to consumers, at the moment consumers do not benefit from many privacy protections against law enforcement access.

Asking consumers to be aware of all privacy risks associated with using DTC and publicly available genealogy services is also a difficult undertaking. As the National Institutes of Health (NIH) reports, consumers use genetic testing services for myriad purposes, including determining familial trees, reconnecting with relatives, and learning about potential health risks.[57] The NIH advises that before submitting a DNA kit, consumers should understand how the company will handle their sample, how the company plans to safeguard genetic data, and whether and how that data will be used for secondary purposes (such as research, advertising, or law enforcement use).[58] It is arguable whether consumers can be knowledgeable on all of these risks associated with submitting their genetic material, especially given studies citing how rarely consumers read privacy policies.[59] Thus, the solution to genetic privacy cannot rest solely with educating consumers or asking companies to invest in privacy-enhancing technologies; true regulatory solutions must stem from government intervention and adherence to privacy-centric frameworks.

B. The Consequences of Lower Stringency Searches

DNA characterization, or DNA typing, refers to a number of methods used to determine and study an individual's genetic variations.[60] Forensic DNA typing can be considered a twofold process involving both DNA profiling and DNA matching.[61] DNA profiling focuses on unique

[55] Computational genomics, or the practice of interpreting genetic information from biological material, has increased in efficiency and cost but remains a costly and time-consuming endeavor. For more information, see Am. Soc'y of Hum. Genetics, *New File Type Improves Genomic Data Sharing While Maintaining Participant Privacy*, SCIENCEDAILY (Oct. 17, 2018), https://www.sciencedaily.com/releases/2018/10/181017141005.htm.

[56] *See* SIMSON L. GARFINKEL, NAT'L INST. OF STANDARDS & TECH., NISTIR 8053, DE-IDENTIFICATION OF PERSONAL INFORMATION 2, 36 (2015).

[57] *What Are the Benefits and Risks of Direct-to-Consumer Genetic Testing?*, MEDLINEPLUS, https://ghr.nlm.nih.gov/primer/dtcgenetictesting/dtcrisksbenefits (last updated Sept. 18, 2020).

[58] *See id.*

[59] Nili Steinfeld, *"I Agree to the Terms and Conditions": (How) do Users Read Privacy Policies Online? an Eye-Tracking Experiment*, 55 COMPUTS. HUM. BEHAV. 992, 998 (2016).

[60] COMM. ON DNA FORENSIC SCI., NAT'L RSCH. COUNCIL, THE EVALUATION OF FORENSIC DNA EVIDENCE 11 (1996) [hereinafter EVALUATION OF FORENSIC DNA].

[61] *Id.* at 15–20.

patterns at certain locations on a chromosome (loci) in a person's DNA.[62] These loci are characterized by the number of repeated genetic sequences, or short tandem repeats (STRs), that occur within each location.[63] An individual's DNA profile is built based on differences in STRs and the associated genetic variations (alleles).[64] DNA matching then uses an algorithm to determine the percentage probability that an individual's DNA profile matches the DNA profile of a forensic specimen.[65] When conducting DNA matching, different stringency requirements can dramatically affect the quantity and accuracy of match results.[66]

A high stringency CODIS search may only look for identical matches, while a moderate stringency search allows for matches that differ between alleles at a STR locus.[67] The FBI initially implemented moderate stringency searches to assist in searching forensic DNA profiles that are partially degraded from the crime scene, contain DNA from more than one individual, or vary in result due to the use of different DNA typing kits.[68] Partial matching may occur as an unintended product of a moderate stringency search.[69] When a CODIS search is conducted on a sample, search results may show that a candidate offender profile is not an exact match to a crime scene sample but that there are a number of shared alleles that indicate the candidate sample may be a biological relative to the source of the forensic sample.[70] While partial matching may suggest a relative to a DNA profile and provide investigative leads, moderate stringency searches exhibit low efficiency in identifying true relatives in CODIS.[71] Furthermore, relaxing stringency levels to omit certain loci for matching also increases the number of false positive results.[72]

Unlike partial matching, FDS is a purposeful database search to identify potential biological relatives of a forensic DNA profile.[73] Furthermore, familial matching involves law enforcement searches strictly based upon relatedness in more indirect and aimless capacities. The FBI has stated that it does not engage in familial searches of CODIS.[74] However, the FBI has employed FDS for comparisons of CODIS DNA pro-

[62] *Id.* at 63.

[63] *Id.* at 23.

[64] *See Frequently Asked Questions on CODIS and NDIS*, *supra* note 36.

[65] *See* EVALUATION OF FORENSIC DNA, *supra* note 60, at 127–28.

[66] Murphy, *supra* note 30, at 297.

[67] Debus-Sherrill & Field, *supra* note 4, at 20–21.

[68] *Frequently Asked Questions on CODIS and NDIS*, *supra* note 36.

[69] FIELD ET AL., *supra* note 43, at 1.

[70] Rohlfs et al., *supra* note 28.

[71] *See id.*

[72] *SWGDAM Recommendations to the FBI Director on the "Interim Plan for the Release of Information in the Event of a 'Partial Match' at NDIS"*, FED. BUREAU OF INVESTIGATION, https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/oct2009/standard_guidlines/swgdam.html (last visited Nov. 11, 2020).

[73] In states that allow FDS, such searches are to only occur after a routine search has been conducted.

[74] *Frequently Asked Questions on CODIS and NDIS*, *supra* note 36.

files to those in DTC FamilyTreeDNA's database,[75] a move harshly criticized by privacy advocates.[76] This ability to perform familial searches differentiates genetic material from other sensitive biometric information, such as fingerprints, facial templates, blood type, and iris scans.[77]

The DOJ has noted that its own laboratories do not generally analyze single nucleotide polymorphisms (SNPs) during forensic DNA casework, and thus outsources genetic material to vendor laboratories that perform forensic genetic genealogical DNA analysis (FGG).[78] FGG involves the examination of more than half a million SNPs, instead of STRs in traditional DNA typing.[79] SNPs span the human genome and may potentially be used to compare shared blocks of DNA between a forensic sample and possible relatives.[80] Law enforcement then uploads the FGG profile to a publicly available genealogy database, or a DTC database if permitted access, where the profile is compared via automation against consumer profiles.[81] Finally, a proprietary computer algorithm then evaluates potential familial relationships between the FGG profile and consumer profiles.[82] The investigative use of FGG searching involves different technologies, genetic markers, and algorithms than used by CODIS, and therefore, these searches are restricted to consumer databases and are not uploaded, searched, or retained by any CODIS index.[83]

Some privacy advocates note that familial matching poses numerous issues regarding individual consent; notice; access, control, and transparency; as well as protections against unreasonable search and seizure.[84] While state and local laws and norms dictate FDS procedures for

[75] *See* E-mail from Bennett Greenspan, President, My Family Tree DNA, to Stephen S. Kramer, Att'y, Fed. Bureau of Investigation (Feb. 3, 2019, 09:33 AM) [hereinafter Greenspan E-mail] (available at https://www.documentcloud.org/documents/6746361-Family-Tree-DNA-FBI-Emails.html#document/p39/a547716).

[76] *See, e.g.*, Matthew Haag, *FamilyTreeDNA Admits to Sharing Genetic Data with F.B.I.*, N.Y. TIMES (Feb. 4, 2019), https://www.nytimes.com/2019/02/04/business/family-tree-dna-fbi.html.

[77] CLIVE E. BOWMAN & PETER GRINDROD, MATHEMATICAL INST., UNIV. OF OXFORD, KINSHIP, FAMILIAL SEARCHING AND BIOMETRICS 2, 7 (2019), https://www.researchgate.net/publication/333982366.

[78] In appropriate cases, the DOJ will outsource biological material to vendor laboratories that perform FGG, and these contracts are to be reviewed by legal counsel to ensure inclusion of appropriate language requiring privacy and security controls for handling biological samples, FGG or SNP-based genetic profiles, and other information and data both submitted to, and generated by, those vendors. U.S. DEP'T OF JUST., INTERIM POLICY: FORENSIC GENETIC GENEALOGICAL DNA ANALYSIS AND SEARCHING 3 (2019) [hereinafter DOJ INTERIM POLICY].

[79] *Id.*

[80] The interim policy states that recombination, or reshuffling, of the human genome is expected with each generation, and it is thus possible to use "predicted levels" of recombination to "make inferences regarding potential familial relationships." Issues with these assumptions will be explained later in this Article. *Id.*

[81] *Id.*

[82] *Id.* at 3–4.

[83] *Id.*

[84] FUTURE OF PRIV. F., PRIVACY BEST PRACTICES FOR CONSUMER GENETIC TESTING SERVICES 1 (2018) [hereinafter PRIVACY BEST PRACTICES] (highlighting consumer control over personal genetic data).

SDIS and LDIS,[85] law enforcement can readily utilize public genealogy sites and commercial databases that permit such access.[86] FDS of these data sets risks implicating innocent persons who are unaware that a family member's DNA profile is part of a genealogical database.[87]

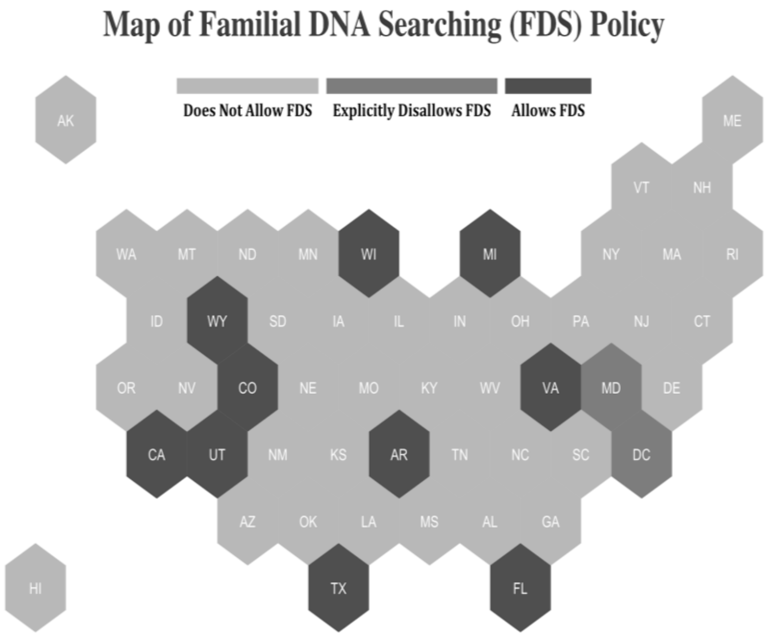

Figure 1. Map of U.S. states that indicates which states [1] lack a formal policy preventing FDS but do not engage in FDS (Does Not Allow FDS), [2] have a formal policy prohibiting FDS (Explicitly Disallows FDS), and [3] allow for FDS (Allows FDS). Source: FBI, 2019

FIGURE 1. *Map of Familial DNA Searching (FDS) Policy*[88]

While consumers may consent to uploading their genetic profiles to sites like GEDMatch, which now require consumers to opt-in for law enforcement access to their respective data, their family members have not consented to their now indirect inclusion in these databases.[89] Individuals who are unaware that they could be matched, based on a rela-

[85] *See infra* Figure 1. States vary in their approaches to FDS, and some have unique processes through which FDS may be conducted. For example, California does not categorically ban the use of FDS but instead requires investigators to garner approval through an interdisciplinary committee prior to engaging in an FDS search. *See CODIS - NDIS Statistics*, *supra* note 37.

[86] Jason Tashea, *Genealogy Sites Give Law Enforcement a New DNA Sleuthing Tool, but the Battle Over Privacy Looms*, ABAJOURNAL (Nov 1, 2019, 4:20 AM), https://www.abajournal.com/magazine/article/family-tree-genealogy-sites-arm-law-enforcement-with-a-new-branch-of-dna-sleuthing-but-the-battle-over-privacy-looms.

[87] *See We Are Updating Our Terms of Service and Privacy Statement Regarding Law Enforcement Matching Preferences*, FAMILYTREEDNA, [hereinafter *Updating Our Terms*] https://mailchi.mp/familytreedna/updates-to-our-terms-of-service-and-privacy-policy-march19?e=dfef197239 (last visited Nov. 11, 2020).

[88] *See CODIS-NDIS Statistics*, supra note 37.

[89] *See GEDmatch.Com Terms of Service and Privacy Policy*, GEDMATCH, https://www.gedmatch.com/tos.htm (last updated Dec. 9, 2019).

tive's familial DNA profile in a database, lack the right to expunge their own or their family member's genetic record from a commercial or publicly available database.[90] Additionally, these searches involve very little transparency. For example, FamilyTreeDNA consumers were unaware of the company's decision to enter into an agreement with the FBI, and that agreement neither accounted for consumer choice nor clearly indicated how law enforcement could routinely access consumer data.[91] Transparency does have its limitations; even if individuals are aware that law enforcement access occurs, those same individuals may be unaware that a family member has shared genetic information.

In addition to these issues concerning meaningful consent and transparency, the lack of enforced quality standards in publicly available genealogy databases and across DTC services means that the majority, if not all, of the profiles stored in DTC databases would not qualify for inclusion in forensic DNA databases managed by law enforcement.[92] Therefore, law enforcement reliance on the search results of such databases is precarious.

C. The Issues with Dirty Data and Algorithmic Bias

DTCs, as commercial entities, are not legally mandated to follow the FBI's standards for quality assurance of rapid DNA analysis.[93] Rather, the reliability of the information provided by the DTCs is assumed, despite the great variation in the types of genetic testing kits and methodologies conducted by each DTC company.[94] Additionally, and unfortunately, there is no open-access information from these companies to verify that similar genetic tests follow the same methodology and CODIS core loci for testing and analysis.[95]

Using simulated data, University of California, Berkeley researchers demonstrated that the algorithmic protocols used by the state of California, in identifying the Golden State killer, have a high probability of around 80%–99% of identifying a familial match of first-degree relatives.[96] While California's methodology effectively matches first-degree

[90] *See* Jocelyn Kaiser, *New Federal Rules Limit Police Searches of Family Tree DNA Databases*, SCI. MAG. (Sept. 25, 2019, 1:55 PM), https://www.sciencemag.org/news/2019/09/new-federal-rules-limit-police-searches-family-tree-dna-databases.

[91] John Verdi & Carson Martinez, *FamilyTreeDNA Agreement with FBI Creates Privacy Risks*, FUTURE OF PRIV. F. (Feb. 6, 2019), https://fpf.org/2019/02/06/familytreedna-agreement-with-fbi-creates-privacy-risks.

[92] *See* SCI. WORKING GRP. ON DNA ANALYSIS METHODS, SWGDAM JULY 2019 REPORT 2 (2019).

[93] *Cf.* John M. Butler, *U.S. Initiatives to Strengthen Forensic Science & International Standards in Forensic DNA*, 18 FORENSIC SCI. INT'L: GENETICS 4, 15 (2015) (FBI quality assurance standards apply to forensic DNA laboratories but not to DTC commercial entities).

[94] Genevieve Rajewski, *Pulling Back the Curtain on DNA Ancestry Tests*, TUFTSNOW (Jan. 26, 2018), https://now.tufts.edu/articles/pulling-back-curtain-dna-ancestry-tests.

[95] *See id.*

[96] Rohlfs et al., *supra* note 28. For unrelated individuals, the same methodologies evidenced a low probability that an unrelated person in the database is identified as a first-degree relative. *Id.*

familial relationships, the reliability of identifying more distant sharing relatives (half-siblings, first cousins, half-first cousins, or second cousins) has a substantial probability of error; as allele sharing decreases, uncertainty increases.[97] Additionally, though there is little risk of falsely identifying an unrelated individual as a first-degree relative, there is a substantial risk of a false positive for a relative more distant than first-degree.[98] Furthermore, another study found that the number of false positives generated increased as the likelihood of a first-degree relationship decreased, suggesting that it would be difficult to identify true first-degree relatives in large databases using FDS techniques.[99]

As with the studies above, much of the research verifying the likelihood rates and effectiveness of FDS relies upon simulated data.[100] Simulation-based assessments demonstrate the utility of using what is known as the kinship index in testing the relationship between pairs of individuals when using FDS.[101] The kinship index that drives FDS algorithms relies on a likelihood ratio calculated by the probability of having certain genotypes if individuals are related as claimed compared to the probability of having those certain genotypes if the individuals are unrelated.[102] A combined relationship index multiplies the kinship index at each STR loci.[103]

Internal FDS research from 23andMe uses a reference data set of 14,393 individuals who have self-reported their ancestry.[104] 23andMe uses patented statistical modeling to discover variants that are typically associated with phenotypic traits and curated model sets based on re-

[97] *See, e.g.*, Steven P. Myers et al., *Searching for First-Degree Familial Relationships in California's Offender DNA Database: Validation of a Likelihood Ratio-Based Approach*, 5 FORENSIC SCI. INT'L: GENETICS 493, 493 (2011).

[98] *Id.* For example, there is a 3%–18% probability that a first cousin will be identified as a full sibling. Rohlfs et al., *supra* note 28.

[99] Thomas M. Reid et al., *Use of Sibling Pairs to Determine the Familial Searching Efficiency of Forensic Databases*, 2 FORENSIC SCI. INT'L: GENETICS 340, 340, 342 (2008) ("[T]he number of false positives generated prior to finding a true match was inversely related to the likelihood of sibship suggesting that many true siblings would not be easily found in a large forensic database via familial searching techniques.").

[100] *See, e.g.*, Tacha Hicks et al., *Use of DNA Profiles for Investigation Using a Simulated National DNA Database: Part II. Statistical and Ethical Considerations on Familial Searching*, 4 FORENSIC SCI. INT'L: GENETICS 316, 317 (2010).

[101] David J. Balding et al., *Decision-Making in Familial Database Searching: KI Alone or Not Alone*?, 7 FORENSIC SCI. INT'L: GENETICS 52, 52–53 (2013).

[102] Jianye Ge & Bruce Budowle, *Kinship Index Variations Among Populations and Thresholds for Familial Searching*, PLOS ONE, May 16, 2012, at 1, 2; *see also* Daniel Kling & Andreas Tillmar, *Forensic Genealogy—A Comparison of Methods to Infer Distant Relationships Based on Dense SNP Data*, 42 FORENSIC SCI. INT'L: GENETICS 113, 115 (2019).

[103] Kristen Lewis O'Connor, *Interpretation of DNA Typing Results for Kinship Analysis*, NAT'L INST. OF STANDARDS & TECH. (Jan. 25, 2011), https://strbase.nist.gov/pub_pres/OConnor_USCIS_interpretation%20of%20DNA.pdf.

[104] *Ancestry Composition: 23andMe's State-of-the-Art Geographic Ancestry Analysis*, 23ANDME, https://www.23andme.com/ancestry-composition-guide-pre-v5 (last visited Nov. 12, 2020).

search available from published scientific studies.[105] There is no indication as to the type of odds ratios, i.e., measures of association between explanatory determinations and outcome determinations, that DTCs like 23andMe employ to identify levels of kinship.[106] Furthermore, the value of utilizing a kinship index, as shown in simulation-based experiments with controlled data classification, is dramatically reduced when training and testing algorithms on unreliable data, i.e., self-reported data.[107]

An algorithm is only as reliable as the data on which it is constructed,[108] and relying on unregulated and "unclean" data is problematic, particularly when considering applications in law enforcement processes[109] like familial searching. There is no denying that large data sets contain "dirty" data that reduces the integrity of the data set, including missing data, miscoding, human error, and statistical inaccuracies.[110] Using dirty data sets then institutes varying levels of systematic bias (e.g., systematic miscoding) and random bias (e.g., "missing at random" data) in modelling calculations.[111] The biggest challenge for handling dirty data lies in correcting for these issues, as failure to do so may deliver incorrect and unpredictable results, especially as the likelihood of a familial relationship decreases.[112]

Algorithmic bias depends on how algorithms are trained, validated, and ultimately deployed, and acknowledging this bias is crucial to recognizing and accounting for data that may skew results compared to the true representation of the population.[113] The trustworthiness of an algorithm's results depends on the quality and cleanliness of the data.[114] Important data integrity characteristics include accuracy, completeness, consistency, uniformity, and validity of the data; DTC genetic testing

[105] *Estimating the Likelihood for a Trait*, 23andMe, https://customercare.23andme.com/hc/en-us/articles/221705208-Estimating-the-Likelihood-for-a-Trait (last visited Nov. 12, 2020).

[106] *See, e.g.*, *id.*

[107] *See* Rashida Richardson et al., *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 N.Y.U. L. Rev. Online 15, 41 (2019); Yanyao Shen & Sujay Sanghavi, *Learning with Bad Training Data via Iterative Trimmed Loss Minimization* 1 (The Univ. of Tex. at Austin, 2019), https://arxiv.org/pdf/1810.11874.pdf; *see also* Zhixin Qi et al., *Impacts of Dirty Data: an Experimental Evaluation* 5–11 (Harbin Inst. of Tech., 2018), https://arxiv.org/pdf/1803.06071.pdf.

[108] *See* Betsy Anne Williams et al., *How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications*, 8 J. Info. Pol'y 78, 109 (2018).

[109] Karen Hao, *Police Across the US are Training Crime-Predicting AIs on Falsified Data*, MIT Tech. Rev. (Feb 13, 2019), https://www.technologyreview.com/s/612957/predictive-policing-algorithms-ai-crime-dirty-data.

[110] Jesmeen M.Z.H. et al., *A Survey on Cleaning Dirty Data Using Machine Learning Paradigm for Big Data Analytics*, 10 Indonesian J. Elec. Eng'g & Comput. Sci. 1234, 1236 (2018).

[111] Williams et al., *supra* note 108, at 80.

[112] *Id.* at 109.

[113] *See id.*; *see also* Sydney J. Freedberg Jr., *Pentagon's AI Problem is 'Dirty' Data: Lt. Gen. Shanahan*, Breaking Def. (Nov. 13, 2019, 9:52 AM), https://breakingdefense.com/2019/11/exclusive-pentagons-ai-problem-is-dirty-data-lt-gen-shanahan.

[114] Thomas C. Redman, *Can Your Data Be Trusted?*, Harv. Bus. Rev. (Oct. 29, 2015), https://hbr.org/2015/10/can-your-data-be-trusted.

services may have issues in some, if not all, of the five criteria.[115] Additionally, marketing of the DTC companies promotes a measure of accuracy that is not well established or explained to consumers for ancestry, health, and forensic purposes.[116]

When considering explainable machine learning algorithms for familial database searching, the modelling process is best divided into four steps: prior probability, prior odds, posterior odds, and posterior probability.[117] Prior probability is determined using nongenetic information to assign a weight to nongenetic data.[118] Prior odds then uses the calculated prior probability to determine the odds of a familial relationship.[119] For example, a father and son (based on public records) would have a prior probability of 0.5 and a prior odds ratio of 1.[120] Following genetic testing, the posterior odds use the prior odds and the kinship index or combined relationship index to determine a weighted odds ratio for the familial identification.[121] Finally, the posterior probability then calculates the probability of a familial match in layman's terms, e.g., 99% familial match.[122]

Creating such an algorithm with a high level of accuracy requires a high level of data integrity when determining prior and posterior probabilities and odds.[123] DTC companies rely on self-reported information in

[115] *See* Sarah Valentine, *Impoverished Algorithms: Misguided Governments, Flawed Technologies, and Social Control*, 46 FORDHAM URB. L.J. 364, 387–89 (2019); *see also* Muhammad Raza, *Data Integrity vs Data Quality: An Introduction*, BMC BLOGS (July 3, 2018), https://www.bmc.com/blogs/data-integrity-vs-data-quality/.

[116] Christopher F. C. Jordens et al., *Direct-to-Consumer Personal Genome Testing: The Problem is not Ignorance–it is Market Failure*, 9 AM. J. BIOETHICS 13, 13 (2009).

[117] For more in-depth information on explainable machine learning and AI, see Deirdre K. Mulligan & Kenneth A. Bamberger, *Procurement as Policy: Administrative Process for Machine Learning*, 34 BERKELEY TECH. L.J. 773, 781–857 (2019); Ribana Roscher et al., *Explainable Machine Learning for Scientific Insights and Discoveries*, 8 IEEE ACCESS 42200, 42200–05 (2020); Umair Saeed et al., *Application of Machine Learning Algorithms in Crime Classification and Classification Rule Mining*, 4 RSCH. J. RECENT SCIS. 106, 106–14 (2015); Wojciech Samek et al., *Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models*, ITU J.: ICT DISCOVERIES (SPECIAL ISSUE NO. 1), Oct. 13, 2017, at 1, 2.

[118] *See* Bruce Budowle et al., *Use of Prior Odds for Missing Persons Identifications*, INVESTIGATIVE GENETICS (2011), https://investigativegenetics.biomedcentral.com/track/pdf/10.1186/2041-2223-2-15; KLAAS SLOOTEN & RONALD MEESTER, FORENSIC IDENTIFICATION: DATABASE LIKELIHOOD RATIOS AND FAMILIAL DNA SEARCHING 2 (2012), https://arxiv.org/pdf/1201.4261.pdf.

[119] SLOOTEN & MEESTER, *supra* note 118, at 17–18.

[120] Prior odds are calculated with the prior probability as the prior probability divided by 1 minus the prior probability.

[121] AMANDA SOZER ET AL., AABB, GUIDELINES FOR MASS FATALITY DNA IDENTIFICATION OPERATIONS 27 (2010), http://www.aabb.org/programs/disasterresponse/Documents/aabbdnamassfatalityguidelines.pdf; *see also* Kim Gin et al., *The 2018 California Wildfires: Integration of Rapid DNA to Dramatically Accelerate Victim Identification*, 65 J. FORENSIC SCI. 791, 794 (2020).

[122] The posterior probability of a relationship can be calculated in two ways: (1) *(posterior odds/(1 + posterior odds)) X 100*, or (2) *(CRI × prior probability/((CRI X prior probability) + (1 - prior probability)) X 100*.

[123] *See* Tal Grossman & Alan Lapedes, *Use of Bad Training Data for Better Predictions*, *in* ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 343, 344 (J.D. Cowan et al. eds., 1993).

the considerations of their priors, with no indication as to what prior probabilities are used based on the nongenetic information provided as well as the flexibility and inference strategies to account for the errors or "nois[e]" in self-reported data.[124] Additionally, the kinship index and combined relationship index used by FDS algorithms are easily skewed based on the population-related stratifications in the training and testing data.[125] Furthermore, incorrect flexibility and predictive measures can ignore the existence of genetic mutations in the general population as well.[126] DTCs are not required to publicize their error rates, and unfortunately, when error rates are not monitored or publicized, laboratories cannot take preventative measures to ensure accuracy in results.[127]

Perhaps the greatest issue in handling biases is that some of these FDS methodologies focus on static modeling.[128] The current approach using the likelihood ratio described above is sensitive to both population frequencies and fluctuations in genetic maps (positions of the SNPs and STRs).[129] Considering the level of noise in these DTC databases, familial database searching is better served by utilizing a more Bayesian approach that allows for flexibility in the probabilistic determinations for kinship analysis;[130] using inflexible models results in skewed probabilities that falsely identify individuals as familial matches.[131] Laboratories should consider allowing for flexibility in the minimum likelihood ratios for determining genealogical relationships, as well as using Baysian methodologies for creating an algorithm that accurately accounts for the noise that exists not only in DTC databases but in the general population.[132]

[124] *Id.*; *see also* Jason Compton, *Data Quality: The Risks of Dirty Data and AI*, FORBES (Mar. 27, 2019, 1:21 PM), https://www.forbes.com/sites/intelai/2019/03/27/the-risks-of-dirty-data-and-ai/#21c535792dc7.

[125] *See* Ge & Budowle, *supra* note 102, at 1.

[126] *See* Elizabeth D. Schifano et al., *SNP Set Association Analysis for Familial Data*, 36 GENETIC EPIDEMIOLOGY 797, 797–98 (2012); Sajad Mirzaei & Yufeng Wu, *RENT+: An Improved Method for Inferring Local Genealogical Trees from Haplotypes with Recombination*, 33 BIOINFORMATICS 1021, 1021–22 (2017).

[127] *See* Ate Kloosterman et al., *Error Rates in Forensic DNA Analysis: Definition, Numbers, Impact and Communication*, 12 FORENSIC SCI. INT'L: GENETICS 77, 84–85 (2014).

[128] *See* SOZER ET AL., *supra* note 121, at 16, 26.

[129] *See id.* at 15–16; Gin et al., *supra* note 121, at 794, 797–98.

[130] Jukka Corander et al., *Bayesian Analysis of Genetic Differentiation Between Populations*, 163 GENETICS 367, 367, 372 (2003); *see* Lingfei Wang et al., *High-Dimensional Bayesian Network Inference from Systems Genetics Data Using Genetic Node Ordering*, FRONTIERS GENETICS, Dec. 20, 2019, at 2; Lingfei Wang & Tom Michoel, Controlling False Discoveries in Bayesian Gene Networks with Lasso Regression P-Values 1–2 (Mar. 27, 2018) (unpublished manuscript), https://www.biorxiv.org/content/10.1101/288217v1.full; Siddharth Samsi et al., *Large-Scale Bayesian Kinship Analysis*, IEEE XPLORE (Nov. 29, 2018), https://ieeexplore.ieee.org/document/8547549.

[131] *See* Jens Hainmueller & Chad Hazlett, *Kernel Regularized Least Squares: Reducing Misspecification Bias with a Flexible and Interpretable Machine Learning Approach*, 22 POL. ANALYSIS 143, 144 (2014).

[132] *See* Rajmud Somorjai et al., *A Data-Driven, Flexible Machine Learning Strategy for the Classification of Biomedical Data*, *in* ARTIFICIAL INTELLIGENCE METHODS & TOOLS FOR SYSTEMS BIOLOGY 67, 67–69 (Werner Dubitzky & Francisco Azuaje eds., 2004); *see also* Joseph P. Hoffbeck & David A. Landgrebe, *Covariance Matrix Estimation and Classification with Limited Training*

Aside from systematic and random biases that exist in DTC data sets, and in the machine learning models themselves, the lack of transparency regarding FDS algorithms further complicates accurate familial matching.[133] Opaqueness of the FDS process prevents individuals from pinpointing errors in the algorithm that may also contribute to incorrect results.[134] Many FDS software providers on the market, as well as the DTC companies and publicly available genealogy databases, hold that their algorithms are proprietary and thus, do not share how they determine familial matching decisions.[135]

23andMe alone holds over twenty-three patents that range from a design patent protecting the look and functionality of their website to their network-based data processing system.[136] 23andMe sued Ancestry in 2018 for infringing on what they alleged were their "art methods" for determining relationships from parents to children, either patrilineal or matrilineal.[137] Alternatively, commercial market success has no impact on law enforcement databases, and the practices and standards of those databases are subject to greater transparency and accountability.[138] Without any type of third-party check, faulty algorithms could continue to be used on dirty data, providing incorrect results to consumers and law enforcement officials conducting FDS. Furthermore, the rise of whole-genome and whole-exome sequencing demands that data sharing and algorithmic accountability in FDS be given more attention by regulatory agencies.[139] Finally, the reliability of the forensic science and statistical power of FDS requires much more comprehensive study prior to adoption as a law enforcement practice.

II. ADVANCES IN SCIENCE OUTPACE REGULATORY SCHEMES

Alongside issues regarding the evidentiary value of genetic information, as well as the robustness of the data sets and algorithms used for familial searching, numerous issues pertaining to government policy must be addressed. Current policy measures lack adequate guidance for law enforcement and do not resolve the privacy risks posed by FDS.

Data, 18 IEEE TRANSACTIONS PATTERN ANALYSIS & MACHINE INTEL. 763, 764–67 (1996); Chris Seiffert et al., *RUSBoost: Improving Classification Performance when Training Data is Skewed*, IEEE XPLORE (Jan. 29, 2009), https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4761297.

[133] *See* Murphy, *supra* note 30, at 294–323.

[134] *See id.*

[135] *See* Birgit Verbeure et al., *Analysing DNA Patents in Relation with Diagnostic Genetic Testing*, 14 EUR. J. HUM. GENETICS 26, 29–30 (2006).

[136] *Patents Assigned to 23andMe, Inc.*, JUSTIA PATENTS, https://patents.justia.com/assignee/23andme-inc (last visited Nov. 13, 2020).

[137] Steve Brachmann, *23AndMe Sues Ancestry.com Over DNA Genetic Testing Kits*, IPWATCHDOG (May 15, 2018), https://www.ipwatchdog.com/2018/05/15/23andme-sues-ancestry-com-dna-genetic-testing-kits/id=97269/.

[138] Lyria Bennett Moses & Janet Chan, *Algorithmic Prediction in Policing: Assumptions, Evaluation, and Accountability*, 28 POLICING & SOC'Y 806, 817–18 (2018).

[139] Robert Cook-Deegan et al., *The Next Controversy in Genetic Testing: Clinical Data as Trade Secrets?*, 21 EUR. J. HUM. GENETICS 585, 586 (2013).

Proposed regulatory measures should take into account existing legal precedent and advances in scientific standards and endeavor to incorporate feedback from a variety of stakeholders.

A. Current Policy Measures

Under the Abandonment Doctrine, some legal theorists argue that consumers abandon their DNA when submitting specimens to DTCs for analysis, and therefore, law enforcement may investigate these abandoned samples as routine.[140] However, other legal experts claim that individuals hold privacy interests in their DNA and that those interests do not lapse when specimens are submitted for genetic testing.[141] Furthermore, they posit that the theory of abandonment does not justify wide-scale genetic surveillance and necessitates the development of standards regarding FDS and consumer privacy.[142]

In September 2019, the DOJ released an interim policy, titled *Interim Policy on Forensic Genetic Genealogical DNA Analysis and Searching*, affecting how federally-funded law enforcement agencies access commercial DNA databases.[143] The policy outlines the importance of "developing practices that protect reasonable interests in privacy, while allowing law enforcement to make effective use of [familial searching] to help identify violent criminals, exonerate innocent suspects, and ensure the fair and impartial administration of justice to all Americans."[144] The policy further outlines that law enforcement will not arrest a suspect absent probable cause based on a familial link but that law enforcement may interact and "extract" genetic samples from potential relatives.[145] Effectively, the policy endorses the practice of interacting with and testing potential matches via standard specimen collection methods on the mere basis of a potential familial link.[146]

The policy does put forth four important caveats regarding law enforcement access to commercial or public genetic information[147]:

> (1) Law enforcement officers may search through genetic profiles from commercial databases if they first try and fail to identify a suspect through a search of CODIS.[148]

[140] Elizabeth E. Joh, *Reclaiming "Abandoned" DNA: The Fourth Amendment and Genetic Privacy*, 100 NW. U. L. REV. 857, 859 (2006).

[141] *Id.* at 871–73; David H. Kaye, *The Genealogy Detectives: A Constitutional Analysis of "Familial Searching"*, 50 AM. CRIM. L. REV. 109, 134–35 (2013).

[142] Joh, *supra* note 140, at 884.

[143] DOJ INTERIM POLICY, *supra* note 78, at 1–2.

[144] *Id.* at 1.

[145] *See id.* at 3–4.

[146] *See id.* at 3.

[147] *Id.*

[148] *Id.* at 3–4.

> (2) Law enforcement officers must identify themselves prior to searching consumer genetic databases, limiting the potential for officers to pose as civilians while uploading offender DNA.[149]
>
> (3) Law enforcement officers and investigators must obtain prosecutorial approval to conduct a search, tempering the potential for abuse by rogue investigators.[150]
>
> (4) Law enforcement officials cannot conduct searches to determine a suspect's genetic predisposition for diseases, other medical conditions, or psychological traits.[151]

According to the DOJ, the DOJ will "continue to assess its investigative tools and techniques to ensure that its policies and practices properly reflect its law enforcement mission and its commitment to respect individual privacy and civil liberties."[152] While this reform is progress in regulating FDS, it still neglects important civil liberties and inadequately addresses issues of informed consent and meaningful notice.[153]

The DOJ policy additionally notes that law enforcement should only use services that provide explicit notice to consumers, but studies demonstrate that few people read or understand privacy policies prior to signing up for a service.[154] Although consumers may perceive genealogical testing as innocuous, DTC companies essentially allow law enforcement to conduct searches that may identify a relative who is a non-user. Research conducted at Stanford University revealed that users of DTCs, prior to the highly public Golden State Killer apprehension, believed that their information was anonymized to disallow such law enforcement access.[155] After the study concluded, in a separate article, the researchers noted that "overall the participants lacked an understanding of both individual and societal level risks to privacy that commercial genetic databases pose."[156]

Familial matching may cast suspicion over an entire family, potentially invading the privacy of all family members simply because certain

[149] *Id.* at 6.

[150] *Id.* at 1.

[151] *Id.* at 1–2.

[152] *Id.* at 1.

[153] *See* Norman P. Lewis et al., *DTC Genetic Testing Companies Fail Transparency Prescriptions*, 30 NEW GENETICS & SOC'Y 291, 292 (2011).

[154] Uri Benoliel & Shmuel I. Becher, *The Duty to Read the Unreadable*, 60 B.C. L. Rev. 2255, 2257 (2019).

[155] Guerrini et al., *supra* note 18.

[156] Jen King, *"It's Not Personal" – DNA, Privacy, and Direct to Consumer Genetic Testing*, CTR. INTERNET & SOC'Y BLOG (Nov. 7, 2019, 3:55 PM), http://cyberlaw.stanford.edu/blog/2019/11/%E2%80%9Cit%E2%80%99s-not-personal%E2%80%9D-%E2%80%94-dna-privacy-and-direct-consumer-genetic-testing.

relatives are perceived as possible suspects.[157] As the DOJ policy permits, allowing law enforcement to acquire reference samples from potential biological relatives can become problematic as the individuals genetically associated with a suspect are completely innocent of the investigated crime.[158] As illustrated above,[159] forty states and the District of Columbia understand the drawbacks of using familial matching for crime-solving and thus do not permit familial matching through CODIS, recognizing that conducting FDS in commercial DTC databases essentially provides an unregulated backdoor to bypass federal and state restrictions on familial matching.[160]

Finally, the DOJ policy is unclear when it comes to the Department's stance on third-party accountability measures. In its current state, the policy does not detail how law enforcement officers and prosecutors will verify that an appropriate and quality-assured CODIS search occurred.[161] Although the interim policy only allows searches of DNA databases following a substantial threat to public safety, or the commission or attempt of a violent crime, without further guidance, the definition of what constitutes a public threat is in the hands of law enforcement.[162] As this is merely an interim policy, the DOJ may address this issue and others with its final policy on forensic genetic genealogy in 2020.[163]

B. Legal Issues with Law Enforcement Use of Consumer Genetic Databases

A number of legal theories conflict with unfettered law enforcement access to personal information.[164] However, these theories are rarely applied to genetic information due to the quick-moving nature of the technology and its ability to rapidly outpace legal norms. Existing and potential legal issues regarding law enforcement access are best broken into four main categories. The first is the issue of individual privacy rights and how law enforcement access to information may violate Fourth Amendment protections against unlawful searches and seizures.[165] The second is the issue of the third-party doctrine, whereby individuals enjoy

[157] Rohlfs et al., *supra* note 28.

[158] *Id.*

[159] *See supra* Figure 1.

[160] *See* SARA DEBUS-SHERRILL & MICHAEL B. FIELD, ICF, UNDERSTANDING FAMILIAL DNA SEARCHING: POLICIES, PROCEDURES, AND POTENTIAL IMPACT, SUMMARY OVERVIEW 3–5 (2017).

[161] Katelyn Ringrose, *DOJ Doesn't Go Far Enough to Limit Searches of Consumer DNA Services*, HILL (Oct. 4, 2019, 11:00 AM), https://thehill.com/opinion/technology/463835-doj-doesnt-go-far-enough-to-limit-searches-of-consumer-dna-services.

[162] *Id.*

[163] *Id.*

[164] DEBUS-SHERRILL & FIELD, *supra* note 160, at 6.

[165] KATELYN RINGROSE & ALISSA GUTIERREZ, FUTURE OF PRIV. F., CONSUMER GENETIC TESTING COMPANIES & THE ROLE OF TRANSPARENCY REPORTS IN REVEALING GOVERNMENT REQUESTS FOR USER DATA 8–9 (2020); *see also* Katelyn N. Ringrose, *A Cautionary Note: Genealogy Companies Need to Stop Giving Warrantless DNA Clues to Law Enforcement*, 124 PENN ST. L. REV. STATIM 302, 314 (2019).

greater protections over their communications and information when that information is not shared with others, like publicly available genealogy and DTC companies.[166] Third, there is the issue of admissibility and whether genetic profiles processed by nonaccredited laboratories should serve as reasonable proof.[167] Fourth and finally, there is the issue of electronic evidence standards.[168] Genetic material, once processed, becomes an electronic profile.[169] Norms around the destruction of genetic material vary; some DTCs allow consumers to request the destruction of their submitted material while others spontaneously destroy the material after a certain time period.[170] However, there is no application of these destruction protocols for electronic records.

C. Privacy Rights Conflict with Unwarranted Access

According to studies undertaken by the National Academy of Sciences as well as the National Human Genome Research Institute (NHGRI), genetic privacy has long been regarded as a healthcare issue under HIPAA.[171] Under HIPAA safeguards, genetic data is protected if it is maintained by a HIPAA-covered healthcare provider, health plan, or healthcare clearinghouse.[172] DTC companies do not fall under the respective umbrellas of healthcare provider, health plan, or healthcare clearinghouse.[173] The Genetic Information Nondiscrimination Act (GINA), a federal statute that prohibits employers, employment agencies, labor unions, and joint labor-management committees from disclosing genetic information, also dictates the importance of privacy in the genetic arena.[174] However, laws regulating genetic privacy have largely followed health information and not genetic information for purposes of identification.

[166] *See* Greenspan E-mail, *supra* note 75; *see also* FamilyTreeDNA, *Dr. Barbara Rae-Venter and Gene by Gene Join Forces to Shape the Future of Investigative Genetic Genealogy*, CISION PR NEWSWIRE (Sept. 27, 2019, 9:00 AM), https://www.prnewswire.com/news-releases/dr-barbara-rae-venter-and-gene-by-gene-join-forces-to-shape-the-future-of-investigative-genetic-genealogy-300926689.html.

[167] *See* COMM. ON DNA TECH. IN FORENSIC SCI., NAT'L RSCH. COUNCIL, DNA TECHNOLOGY IN FORENSIC SCIENCE 132 (1992).

[168] *See* RINGROSE & GUTIERREZ, *supra* note 165, at 6.

[169] *See id.* at 4.

[170] Kevin Loria, *How to Delete Your Data from 23andMe, Ancestry, and Other Sites*, CONSUMER REPS. (Jan. 29, 2019), https://www.consumerreports.org/health-privacy/how-to-delete-genetic-data-from-23andme-ancrestry-other-sites/.

[171] Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-19, 110 Stat. 1936.

[172] *See* OFF. FOR CIV. RTS., U.S. DEP'T OF HEALTH & HUM. SERVS., OCR PRIVACY BRIEF: SUMMARY OF THE HIPAA PRIVACY RULE 2 (2003).

[173] 42 U.S.C. § 1320d (2018).

[174] *See* 29 C.F.R. § 1635.8(b)(6) (2020) (outlining an exception to GINA allowing employers and training programs that conduct DNA analysis for law enforcement purposes as a forensic laboratory or for purposes of human remains identification to request or require genetic information from their employees, but only when it is used for analysis of DNA identification markers for quality control to detect sample contamination).

According to the NHGRI, a handful of legal cases have lent to and shaped genetic privacy laws pertaining to genetic data as personally identifiable information in the United States.[175] These cases include *Olmstead v. United States*,[176] *Katz v. United States*,[177] *Berger v. New York*,[178] and *United States v. Miller*.[179] Two cases stand out as most relevant in the field of law enforcement access to commercial genetic databases: *Maryland v. King*[180] and *Carpenter v. United States*.[181] Both stood before the U.S. Supreme Court within the past decade and helped establish the boundaries of the Fourth Amendment.[182] While these cases all dealt with the scope of the Fourth Amendment and the limits of privacy when communications and information are held by third parties, each of their holdings lay another brick in the path to genetic privacy in the United States.

1. *Maryland v. King*

In 2013, in a contentious 5–4 decision, the U.S. Supreme Court in *King* ruled that collecting an arrestee's DNA via buccal swab is "a legitimate police booking procedure that is reasonable under the Fourth Amendment."[183] Therefore, genetic information may be taken from arrestees to create CODIS profiles. However, this holding is limited to the taking and retention of genetic material of arrested individuals, and the Court noted that the "expectations of privacy of an individual taken into police custody 'necessarily are of a diminished scope'" when compared with the rights enjoyed by others.[184] The Court also noted that a DNA swab could only be taken following an "arrest supported by probable cause to hold for a serious offense."[185] Furthermore, the Court emphasized that a "sample may not be [added to] a database before [an] individual is arraigned."[186]

With this ruling, the Court classified buccal swabbing as a search under the Fourth Amendment but diminished the need for a warrant given that the arrestee was already in police custody for a serious offense

[175] *See Privacy in Genomics*, NAT'L HUM. GENOME RSCH. INST., https://www.genome.gov/about-genomics/policy-issues/Privacy (last visited Nov. 13, 2020).

[176] 277 U.S. 438 (1928).

[177] 389 U.S. 347 (1967).

[178] 388 U.S. 41 (1967).

[179] 425 U.S. 435 (1976).

[180] 569 U.S. 435 (2013).

[181] 138 S. Ct. 2206 (2018).

[182] *King*, 569 U.S. at 465–66; *Carpenter*, 138 S. Ct. at 2262 (Gorsuch, J., dissenting) (the majority opinion explores the issue of the third-party doctrine in relation to cell-site location information, and Justice Gorsuch touches briefly on genetic privacy in his dissent: "Can [the government] secure your DNA from 23andMe without a warrant or probable cause? *Smith* and *Miller* say yes it can—at least without running afoul of *Katz*. But that result strikes most lawyers and judges today—me included—as pretty unlikely.").

[183] *King*, 569 U.S. at 465–66.

[184] *Id.* at 462 (quoting Bell v. Wolfish, 441 U.S. 520, 557 (1979)).

[185] *Id.* at 465.

[186] *Id.* at 443.

and probable cause precipitated the arrest.[187] Furthermore, the Court noted that, in terms of the processing of DNA, the detainee's DNA "loci came from noncoding DNA parts that do not reveal an arrestee's genetic traits and are unlikely to reveal any private medical information."[188] In contrast, individuals who submit a buccal swab to DTC testing services reveal substantial information about their medical and genealogical traits.[189] In fact, due to evolving technology since that case was decided, consumer services are increasingly marketed to test genomes for diseases.[190] While taking a swab from an arrestee is legal under the Supreme Court ruling, the Court acknowledges that arrestees have a lowered expectation of privacy.[191] Consumers have no such lowered expectation of privacy.[192]

According to privacy researchers at the Center for Internet and Society at Stanford Law, uploading an unidentified suspect's DNA to a commercial website is qualitatively different from taking a buccal swab from an arrestee.[193] First, an arrestee must be identified and formally accused of a crime prior to being subject to a swab.[194] Second, while an arrestee's ability to assert their reasonable expectation of privacy is limited, the ability of members of the general population is not.[195] Finally, arrestees whose DNA is collected may have that DNA uploaded to CODIS, whereby they can expect adherence to certain privacy and safety procedures.[196] For example, information as to their health and genetic predispositions are left out of a CODIS analysis, whereby such information is a common feature of commercial websites.[197]

2. *Carpenter v. United States*

The Supreme Court noted that *Carpenter* raised two important issues: first, whether a person has an expectation of privacy in their physical location and movements; and second, whether the person has a reasonable expectation of privacy in information voluntarily turned over to third parties.[198] The Court answered affirmatively to both questions.[199] *Carpenter* largely addressed issues related to third-party data retention

[187] Ringrose, *supra* note 165, at 312.
[188] *Id.* at 313 (internal quotations omitted).
[189] *Id.*
[190] *See* ANC., https://www.ancestry.com/health (last visited Nov. 13, 2020) (information regarding AncestryHealth).
[191] *King*, 569 U.S. at 463.
[192] Ringrose, *supra* note 165, at 313.
[193] *Id.*; *see also* King, *supra* note 156.
[194] Ringrose, *supra* note 165, at 313.
[195] *King*, 569 U.S. at 463.
[196] Ringrose, *supra* note 165, at 313.
[197] *Id.*
[198] *See* Carpenter v. United States, 138 S. Ct. 2206, 2217 (2018).
[199] *See id.*; *see also* Ringrose, *supra* note 165, at 316.

and whether consumers hold a reasonable expectation of privacy in information held by commercial entities and sided with consumers.[200]

Privacy experts would argue that if the *Carpenter* Court held that there is a reasonable expectation of privacy in one's location, even when a commercial cell phone company holds that location, it is reasonable for individuals to hold a similar, if not stronger, expectation in the privacy of their genetic information when a DTC company holds that information.[201] Some privacy advocates argue that individuals hold a significant expectation of privacy in their DNA, even if they voluntarily turn it over for commercial purposes, and that there is a violation when it comes to familial matching an individual's DNA who has not consented to a commercial use of their genetic information.[202] While both individuals legally deserve similar protections, privacy experts should agree that it is violative to run familial searches against individuals who simply had the ill luck of being genetically related to someone who subscribed to a consumer genetic-testing service.[203]

3. Genetic Privacy and the Third-Party Doctrine

The cases outlined above help create the third-party doctrine, an evolving notion under the Fourth Amendment pertaining to an individual's reasonable expectation of privacy in information turned over to others.[204] The Fourth Amendment contains two clauses that can be at odds.[205] The first clause requires that all searches and seizures be reasonable, and the second requires that all warrants meet certain minimum requirements, particularly when it comes to sufficiently describing the things and information being seized.[206]

In one of the earliest Fourth Amendment cases, the Supreme Court held that anything "exposed" on the outside of a piece of mail is not entitled to Fourth Amendment protections.[207] Similarly, under the "plain view" doctrine, Justice Brandeis noted that using a searchlight to view cases of liquor on the deck of a ship was not a Fourth Amendment search.[208] In the case of a letter or the contents of a briefcase, it might be reasonable to assume that an individual loses their expectation of privacy when they give their physical belongings to another. However, in the

[200] *See* Ringrose, *supra* note 165, at 318.

[201] *Id.* at 316.

[202] *Id.* at 318.

[203] *Id.*

[204] Orin Kerr & Greg Nojeim, *The Data Question: Should the Third-Party Records Doctrine Be Revisited?*, ABA J. (Aug. 1, 2012, 9:20 AM), http://www.abajournal.com/magazine/article/the_data_question_should_the_third-party_records_doctrine_be_revisited/.

[205] *See* Scott E. Sundby, *A Return to Fourth Amendment Basics: Undoing the Mischief of* Camara *and* Terry, 72 MINN. L. REV. 383, 383–84 (1988).

[206] *See id.*

[207] *Ex parte* Jackson, 96 U.S. 727, 735–36 (1877).

[208] *See* United States v. Lee, 274 U.S. 559, 563 (1927).

Internet age, letters and communications are essentially always in the hands of others (including internet service providers). Justice Sotomayor best articulated this notion:

> More fundamentally, it may be necessary to reconsider the premise that an individual has no reasonable expectation of privacy in information voluntarily disclosed to third parties. This approach is ill suited to the digital age, in which people reveal a great deal of information about themselves to third parties in the course of carrying out mundane tasks.[209]

In the genetic privacy context, all processing necessarily requires the use of technology. When processing and storage takes place in the hands of a third party, the context looks remarkably similar to precise geolocation information, for which the Supreme Court ruled that warrants are necessary to gather cell site location information.[210] According to a report penned by the Congressional Research Service, while "it appears unlikely that Congress would be willing to completely eliminate the third-party doctrine," Congress may yet be inclined to "engage in a subject-by-subject approach, in which Congress limits the third-party doctrine in certain areas."[211] Genetic privacy is one such area where Congress could intervene to limit the ability of law enforcement to access genetic information held by third parties—that intervention would need to be guided by appropriate scientific knowledge and an acknowledgement of existing legal precedent.

D. Admissibility Standards

In the trial context, the standards set in both *Daubert v. Merrell Dow Pharmaceuticals, Inc.*[212] and *Frye v. United States*[213] establish admissibility requirements for scientific evidence in a case.[214] The federal court system exclusively follows the *Daubert* standard while state courts vary between following the *Daubert* and *Frye* standards.[215] Three different cases establish the *Daubert* standard.[216] In *Daubert*, the primary case decided in 1993, the U.S. Supreme Court held that the Federal Rules of Evidence superseded the *Frye* standard for admissibility of expert evi-

[209] United States v. Jones, 565 U.S. 400, 417 (2012) (Sotomayor, J., concurring) (citations omitted).

[210] Carpenter v. United States, 138 S. Ct. 2206, 2223 (2018).

[211] RICHARD M. THOMPSON II, CONG. RSCH. SERV., R43586, THE FOURTH AMENDMENT THIRD-PARTY DOCTRINE 26 (2014).

[212] Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579 (1993).

[213] Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).

[214] *Daubert*, 509 U.S. at 597–98*; Frye*, 293 F. at 1014.

[215] Emily Pincow & Alexis Kellert, *The Battle of the Experts*, AM. BAR ASS'N (Nov. 26, 2018), https://www.americanbar.org/groups/litigation/committees/mass-torts/practice/2018/the-battle-of-the-experts/.

[216] *See Daubert*, 509 U.S. at 579–98; Kumho Tire Co. v. Carmichael, 526 U.S. 137, 137–48 (1999); Gen. Elec. Co. v. Joiner, 522 U.S. 136, 136–47 (1997).

dence in federal courts.[217] *Frye* established that an "expert opinion" must be based on a scientific technique that is generally accepted as reliable in the relevant scientific community, a standard now known as the *Frye* standard.[218] The other two cases that contribute to the *Daubert* standard are *General Electric Co. v. Joiner*[219] and *Kumho Tire Co. v. Carmichael*.[220] In the former case, the Supreme Court emphasized the importance of scientific methodology and provided courts the power to exclude expert testimony if there is a significant gap between the data available and the conclusions put forth by an expert.[221] In the latter case, the Supreme Court expanded a judge's gatekeeper role to nonscientific expert testimony.[222]

The current standard of admissibility for DNA evidence follows the American Bar Association (ABA) Standard 5.1-Admissibility of DNA evidence, wherein "expert testimony concerning DNA evidence, including statistical estimates, should be admissible if based on a valid scientific theory, a valid technique implementing that theory, and testing and interpretation properly applying that theory and technique."[223] This standard implies that admissibility of DNA evidence from experts, such as DTCs, should follow the presently valid and generally accepted technique that requires laboratory testing equivalent to that outlined in the FBI's quality assurance[224] guidelines.[225] Additionally, the DOJ interim policy also states that FGG "involves different DNA technologies, genetic markers, algorithms, and databases from those used by CODIS," which are the established valid techniques for DNA typing and profiling.[226]

Since there are no requirements for commercial databases to maintain DNA results from procedures that follow the FBI's quality assurance

[217] *Daubert*, 509 U.S. at 588–89.

[218] *Frye*, 293 F. at 1014.

[219] 522 U.S. 136 (1997).

[220] 526 U.S. 137 (1999).

[221] *Gen. Elec. Co.*, 522 U.S. at 146–47.

[222] *Kumho Tire Co.*, 526 U.S. at 147–48.

[223] ABA Standards for Criminal Justice: Admissibility of DNA Evidence Standard 16-5.1, in ABA Standards for Criminal Justice: DNA Evidence (3d ed. 2007).

[224] FED. BUREAU OF INVESTIGATION, QUALITY ASSURANCE STANDARDS FOR FORENSIC DNA TESTING LABORATORIES 1 (2009) [hereinafter QUALITY ASSURANCE STANDARDS]. For information from the FBI and other agencies on stringency and quality assurance standards, see Letter from Alice R. Isenberg, Section Chief, FBI Lab'y, to Anthony Onorato, Chair, FBI Lab'y (Apr. 16, 2014) (on file with author); SCI. WORKING GRP. ON DNA ANALYSIS METHODS, INTERPRETATION GUIDELINES FOR MITOCHONDRIAL DNA ANALYSIS BY FORENSIC DNA TESTING LABORATORIES 2–3 (2019); SCI. WORKING GRP. ON DNA ANALYSIS METHODS, INTERPRETATION GUIDELINES FOR AUTOSOMAL STR TYPING BY FORENSIC DNA TESTING LABORATORIES (2017); FED. BUREAU OF INVESTIGATION, ADDENDUM TO THE QUALITY ASSURANCE STANDARDS AUDIT FOR DNA DATABASING LABORATORIES PERFORMING RAPID DNA ANALYSIS AND MODIFIED RAPID DNA ANALYSIS USING A RAPID DNA INSTRUMENT (2014); FED. BUREAU OF INVESTIGATION, THE FBI QUALITY ASSURANCE STANDARDS AUDIT FOR DNA DATABASING LABORATORIES (2011).

[225] QUALITY ASSURANCE STANDARDS, *supra* note 224, at 1.

[226] DOJ INTERIM POLICY, *supra* note 78, at 3–4.

guidelines, and the scientific standards behind their testing methodologies are not subject to validity testing, such DNA evidence does not pass either the *Frye* or *Daubert* standard for admissibility.[227] According to a study conducted by the National Institute for Standards and Technology (NIST), accredited forensic laboratories are legally mandated to adhere to FBI standards for quality[228] assurance[229] as well as follow requirements indicated by the *Daubert* admissibility standard.[230] This means that if a laboratory cannot provide a probability statistic for a DNA profile under FBI standards, it must report the results as uninterpretable or inconclusive.[231]

Even if the *Daubert* evidentiary standard is met in some way, privacy issues regarding lawful access of DTC genetic information still require attention. Privacy proponents argue that warrantless access to DTC genetic databases should be prohibited and that lawful access should only occur after obtaining a warrant based on probable cause.[232] While law enforcement databases contain genetic profiles of individuals who were arraigned based on probable cause following their arrests, consumer databases do not facially require a law enforcement standard of proof, unless that standard is required and articulated under their privacy policy.[233] Furthermore, authorities are not required to obtain a warrant to submit an access request or a genetic profile to a consumer entity, nor do they need to follow the other stringent requirements imposed by CODIS.[234] At the moment, there is no comprehensive federal privacy law in the United States to govern the use and processing of genetic information outside the healthcare scheme.[235]

E. The Use of Genetic Data as Electronic Evidence

The genetic material that consumers initially send to DTC services are often destroyed within a period of two to ten years, and some services allow consumers to request the destruction of their material immediate-

[227] The use of a subpoena by law enforcement to compel a DTC or publicly available genetic database to conduct a new FDS, instead of mere access to a historical result, is not only beyond the scope of a lawful search but would also not produce a result that would meet the admissibility standards discussed; aside from legal process, the claim of receiving a significantly valid result is effectively a moot point as issues with the accuracy of FDS algorithms and data integrity persist.

[228] QUALITY ASSURANCE STANDARDS, *supra* note 224, at 1.

[229] *Request for Public Comment on Federal Bureau of Investigation (FBI) National Quality Assurance Standards*, AM. ACAD. OF FORENSIC SCIS. (Aug. 24, 2017), https://news.aafs.org/asb-news/request-for-public-comment-on-federal-bureau-of-investigation-fbi-national-quality-assurance-standards/.

[230] John M. Butler, *U.S. Initiatives to Strengthen Forensic Science & International Standards in Forensic DNA*, 18 FORENSIC SCI. INT'L: GENETICS 4, 15–16 (2015).

[231] QUALITY ASSURANCE STANDARDS, *supra* note 224, at 1–2.

[232] Ringrose, *supra* note 165, at 329.

[233] *See id.* at 326–27.

[234] *Id.* at 329.

[235] *See, e.g.*, *2019 Consumer Data Privacy Legislation*, NAT'L CONF. OF STATE LEGISLATURES (Jan. 3, 2020), https://www.ncsl.org/research/telecommunications-and-information-technology/consumer-data-privacy.aspx.

ly.[236] While this genetic material may be destroyed, a user's genetic profile may be housed for longer.[237] Publicly available genealogy databases allow users to port their data from DTC services onto their own servers and then over to the public database.[238] This means that public databases may never be privy to the genetic material and only handle the data. Furthermore, genetic profiles, not samples, are compared against one another,[239] essentially leading to the comparison of two electronic data sets for familial matching.[240] Various legal issues prevail regarding how the law treats genetic information and how the data itself differs from the material that can be retested.[241] Genetic information can be subject to many errors, including processing errors, interpretation errors, and human errors.[242] As a form of electronic evidence, different rules apply to genetic information than to genetic material and other biological evidence.[243]

An interesting issue regarding genetic information and law enforcement access to commercial databases is whether genetic information should be construed as an electronic communication. If an individual sent their genetic profile in a machine- or human-readable format to a friend over email, would that fall under the tenants of a communication? In that case, it would be possible for law enforcement to access that conversation and the accompanying genetic information under the ECPA[244] or, alternatively, if the information were stored, through the Stored Communications Act (SCA).[245]

The ECPA and SCA largely apply to communications in transit, communications in remote or home storage, and unopened emails and messages.[246] These communications require different warrant or subpoena standards dependent on where a communication is stored and whether or when it is accessed.[247] Neither statutory regime allows, or should allow, for unwarranted access to genetic information for purposes of famil-

[236] *See* Eric Ravenscraft, *How to Protect Your DNA Data Before and After an at-Home Test*, N.Y. TIMES (June 12, 2019), https://www.nytimes.com/2019/06/12/smarter-living/how-to-protect-your-dna-data.html; Rachele Hendricks-Sturrup & Katelyn Ringrose, *Letter to the Editor: Advice on Genetic Testing*, N.Y. TIMES (Oct. 31, 2019), https://www.nytimes.com/2019/10/31/opinion/letters/automakers-emissions.html.

[237] Ravenscraft, *supra* note 236.

[238] Nila Bala, *Criminal Suspects Deserve Genetic Privacy, Too*, SLATE (Mar. 18, 2019, 7:30 AM), https://slate.com/technology/2019/03/genetic-genealogy-law-enforcement-suspects-dna-privacy-gedmatch.html.

[239] *See* DEBUS-SHERRILL & FIELD, *supra* note 160, at 1–2.

[240] *See id.*

[241] *See* Aziza Ahmed, *Ethical Concerns of DNA Databases Used for Crime Control*, PETRIE-FLOM CTR. (Jan. 14, 2019), https://blog.petrieflom.law.harvard.edu/2019/01/14/ethical-concerns-of-dna-databases-used-for-crime-control/.

[242] *Id.*

[243] *See, e.g.*, H.B. 57, 2019 Leg., Gen. Sess. (Utah 2019).

[244] 18 U.S.C. § 2516 (2018) (originally enacted as Title III of the Omnibus Crime Control and Safe Streets Act of 1968, 18 U.S.C. §§ 2510–2520).

[245] *Id.* § 2701.

[246] *See id.* §§ 2510, 2701.

[247] *See id.* § 2518.

ial matching.[248] Wiretapping cases, and federal laws on the subject, have largely concerned bank records, pen registers, and forms of communication that can be overheard or intercepted in traditional means (i.e., listening in on a suspect's phone conversations).[249] The ECPA broadened the definition of "interception" by adding the words "or other acquisition" so that the SCA is no longer limited to interception of communications that are heard.[250] However, genetic information, as well as requesting companies and entities to find family members based on a genetic signature, is a category far outside this scope.[251]

A user of genetic testing services may receive their genetic information from a service, and that information may be considered a communication.[252] A proper analogue here is law enforcement issuing a warrant to a social media platform asking for information regarding a particular user. However, that analogue ends when law enforcement requests that a company run a familial search to find relatives of a particular suspect. At that point, a warrant searching an entire database for potential family members would be overly broad.

Under the Fourth Amendment, warrants must be (1) justified by probable cause and (2) must sufficiently describe the place to be searched and the persons or things to be seized.[253] In the familial searching context, law enforcement may find evidence in a genetic database that contains information from millions of people but does not equate to conducting searches with sufficient degree of particularity.[254] Furthermore, such searches implicate the issues of consent and notice relating to family members who may not know that they are subject to a search.[255] Finally, given the arguments addressed above about the efficacy of familial searching, the individuals identified as particular family members may not be closely related to the suspect, or related at all.[256] This further frustrates the notion that such a warrant would be, in any way, particular. By issuing a warrant alongside a genetic profile and expecting DTCs or publicly available genealogy databases to find family members of the individual to whom that profile belongs, law enforcement is, in effect, forcing private companies to fulfill an investigatory role.

[248] *See id.* §§ 2510–2523, 2701–2712.

[249] *See* United States v. Kahn, 415 U.S. 143, 147 (1974) (involving interception of phone conversations); United States v. Turner, 781 F.3d 374, 383 (8th Cir. 2015) (involving affidavits for issuing wiretaps using pen registers); United States v. London, 66 F.3d 1227, 1231 (1st Cir. 1995) (involving federal agents obtaining bank records); 18 U.S.C. § 3123 (repeatedly mentions pen registers).

[250] 18 U.S.C. § 2510 (amended in 1986 to include "or other" following "aural"); *see also id.* § 3127.

[251] *See id.* §§ 2510–2523, 3121–3127.

[252] *See* GENETICS HOME REFERENCE, DEP'T OF HEALTH & HUM. SERVS., HELP ME UNDERSTAND GENETICS: DIRECT-TO-CONSUMER GENETIC TESTING 4 (2020).

[253] U.S. CONST. amend. IV.

[254] *See* Rohlfs et al., *supra* note 28.

[255] *See* Kaiser, *supra* note 90.

[256] *See* M.Z.H. et al., *supra* note 110, at 1234.

ECPA, HIPAA, and GINA are largely considered the three most important federal statutes pertaining to the privacy and security of genetic information.[257] The ECPA applies to genetic privacy to a much lesser degree than the GINA or HIPAA, but ECPA does address law enforcement access to information.[258] To secure an ECPA interception legal order, a DOJ official must approve the application for the court order authorizing the interception of wire or oral communications.[259] The procedure is only available where there is probable cause to believe that the wiretap or electronic eavesdropping will produce evidence of a federal crime.[260] Recently, AncestryDNA denied a law enforcement warrant for improper service, and a legal scholar noted that "[i]f statistical probability standing alone is sufficient to define probable cause, then probable cause is going to be virtually meaningless in an era of big data."[261] Warrants for genetic information should be based on probable cause and direct matches—not the statistical probability of finding a familial match amongst millions of users.

III. Policy Recommendations

Methodological- and principle-driven investigatory work for law enforcement should not focus on whether two individuals are related, but rather on whether there is genetic and nongenetic information to support the claimed relationship. Therefore, instead of treating DTC and publicly available genealogy databases as open forums, whereby genetic matches and relatives may be found, investigators should rely on proof outside of a mere genetic link. Policy in this arena should disallow overbroad familial matching on publicly available genealogy and DTC databases, as such searches lack evidentiary backing, and address the need for warrants to search for direct hits. There are numerous ways to honor genetic privacy,

[257] *See* 18 U.S.C. § 2511 (2018); 42 U.S.C. § 1320d-9 (2018); 29 C.F.R. §1635.8(a) (2020) (all three statues have some bearing on the practices of consumer genetic testing services, with ECPA providing safeguards for communications, HIPAA providing privacy and security protections for PHI held by covered entities, and GINA prohibiting some discriminatory uses of genetic information).

[258] *See* 18 U.S.C. § 2516 (discussing law enforcement access to information but not directly discussing genetic information).

[259] *Id.*

[260] *See id.* § 2511(2)(i).

> It shall not be unlawful under this chapter [18 U.S.C. §§ 2510 et seq.] for a person acting under color of law to intercept the wire or electronic communications of a computer trespasser transmitted to, through, or from the protected computer, if . . . (I) the owner or operator of the protected computer authorizes the interception of the computer trespasser's communications on the protected computer; (II) the person acting under color of law is lawfully engaged in an investigation; (III) the person acting under color of law has reasonable grounds to believe that the contents of the computer trespasser's communications will be relevant to the investigation; and (IV) such interception does not acquire communications other than those transmitted to or from the computer trespasser.

Id.

[261] Peter Aldhous, *A Court Tried to Force Ancestry.com to Open up its DNA Database to Police. The Company Said No.*, BuzzFeed News (Feb. 3, 2020, 7:11 PM), https://www.buzzfeednews.com/article/peteraldhous/ancestry-dna-database-search-warrant.

including changes adopted by the companies within their privacy policies regarding how they intend to collect and process genetic data. Ultimately, however, the responsibility rests with the government to dictate the acceptable perimeters of genetic searches.

A. Understanding DTC Privacy Policies

Presently, besides recommendations within the DOJ's interim guidance, DTC companies and publicly available genealogy databases solely regulate law enforcement's access to consumer profiles.[262] There are no enforcement mechanisms to ensure that the companies voluntarily enter into privacy agreements with consumers.[263] Aside from bringing suit with the Federal Trade Commission (FTC), there are not many oversight mechanisms other than consumer reporting to ensure that companies uphold those promises.[264]

When examining the privacy policies for the nine DTC companies listed previously, this Article considers three critical privacy best practices: transparency, use and onward transfer, and consent.[265]

> (1) Transparency: Three of the nine DTC companies issue annual transparency reports that discuss the number of times and ways law enforcement access and search their respective DNA databases.[266]
>
> (2) Use and Onward Transfer: Eight of the nine DTC companies require a warrant, subpoena, or other legal order to provide law enforcement access to their respective DNA databases.[267]
>
> (3) Consent: None of the DTC companies provide a mechanism for gaining consumer consent to the sharing of genetic information with law enforcement.[268] However, the public database GEDmatch that contains uploaded DNA from consumers of the major DTCs recently created an opt-in process for users to consent to sharing their genetic information with law en-

[262] *See* Lindsey Van Ness, *DNA Databases Are Boon to Police but Menace to Privacy, Critics Say*, STATELINE (Feb. 20, 2020), https://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2020/02/20/dna-databases-are-boon-to-police-but-menace-to-privacy-critics-say.

[263] *See* FED. TRADE COMM'N, A BRIEF OVERVIEW OF THE FEDERAL TRADE COMMISSION'S INVESTIGATIVE, LAW ENFORCEMENT, AND RULEMAKING AUTHORITY 3 (2019).

[264] *See* Elisa Jillson, *Selling Genetic Testing Kits? Read on.*, FED. TRADE COMM'N: BUS. BLOG (Mar. 21, 2019, 11:35 AM), https://www.ftc.gov/news-events/blogs/business-blog/2019/03/selling-genetic-testing-kits-read.

[265] *See* PRIVACY BEST PRACTICES, *supra* note 84, at 3–7.

[266] *Id.*

[267] *See supra* Table 1.

[268] *See* PRIVACY BEST PRACTICES, *supra* note 84, at 3–7.

> forcement.[269] Only 185,000 or 14.2% of the database's 1.3 million total users opted-in to sharing their genetic information.[270]

According to recent reports, consumers are unlikely to read or understand privacy policies for many of the services they use.[271] This also holds true of genetic testing services, and for this reason, consumer-friendly language is critically necessary.[272] There is a clear need for concise privacy policies that are readily understood by consumers and enforcement mechanisms to hold companies to their promises. According to researchers, "the average readability level of these [online privacy policies] is comparable to the usual score of articles in academic journals, which typically do not target the general public."[273] Therefore, a blanket privacy policy is insufficient for consumers to understand the privacy rights they relinquish when providing their DNA. As this Article mentions above, DNA is no longer purely individualistic when utilizing FDS techniques.[274] Therefore, privacy policies should not be restricted solely to direct consumers—but should also extend to relatives of consumers who wish to inquire about the service's approach to law enforcement requests.

B. Regulatory Recommendations

Insofar as regulatory protections, there are a few ways for genetic privacy to be addressed and protected within the United States. The first route is through federal or state legislative efforts to enact comprehensive privacy protections. The states referenced above that have banned or provided regulatory guidance on FDS have contributed sectoral laws[275] on the use of genetic information, but a comprehensive bill[276] would address broader data privacy concerns.[277] This Article argues that while federal or state proposals may lead to more privacy protections, agencies themselves are best positioned to enact swift change.

[269] Kashmir Hill & Heather Murphy, *Your DNA Profile Is Private? A Florida Judge Just Said Otherwise*, N.Y. TIMES (Nov. 5, 2019), https://www.nytimes.com/2019/11/05/business/dna-database-search-warrant.html.

[270] *Id.*

[271] *See* Aaron Smith, *Half of Online Americans Don't Know What a Privacy Policy is*, PEW RSCH CTR. (Dec. 4, 2014), https://www.pewresearch.org/fact-tank/2014/12/04/half-of-americans-dont-know-what-a-privacy-policy-is/.

[272] *See* Guerrini et al., *supra* note 18.

[273] Benoliel & Becher, *supra* note 154, at 2256.

[274] *Updating Our Terms*, *supra* note 87.

[275] *See, e.g.*, S.B. 980, 2019-2020 Leg., Reg. Sess. (Cal. 2020) (if passed, establishes obligations for DTC genetic testing companies and others that collect or process genetic information).

[276] *See* John Verdi & Katelyn Ringrose, *California SB 980 Would Codify Many of FPF's Best Practices for Consumer Genetic Testing Services, but Key Differences Remain*, FUTURE OF PRIV. F. (July 24, 2020), https://fpf.org/2020/07/24/california-sb-980-would-codify-many-of-fpfs-best-practices-for-consumer-genetic-testing-services-but-key-differences-remain/ (explaining how California's proposed SB 980 is an example of a potential comprehensive state bill).

[277] *See* John Verdi & Carson Martinez, *FPF Perspective: Limit Law Enforcement Access to Genetic Datasets*, FUTURE OF PRIV. F. (Oct. 12, 2018), https://fpf.org/2018/10/12/fpf-perspective-limit-law-enforcement-access-to-genetic-datasets/.

Legislators continue to introduce federal and state privacy laws, with over a dozen federal comprehensive privacy bills or draft bills announced over the past several years and countless state legislation on the horizon.[278] A federal privacy bill can address some of the issues associated with law enforcement use of DTC and publicly available genealogy databases by citing genetic information as a category of sensitive information, which would subsequently enjoy additional protections against nonconsensual data use and processing.[279] Most federal privacy bills would vest rulemaking power with the FTC, and in this scenario, the FTC would be the policy authority for DTCs.[280] Prior to vesting the FTC with investigation and enforcement power in the arena of genetic privacy, legislators should consider the importance of issues around law enforcement access and potentially include funding for greater research around ways to protect consumers of DTC and publicly available genealogy services.

While most proposed federal privacy legislation deals with consumer privacy and does not necessarily regulate law enforcement access, these bills could designate genetic information as particularly sensitive, thus deserving greater protections in the processing context. An enforcement agency could then extend those protections through later rulemaking to include provisions related to government access. Some state bills already examine and attempt to address law enforcement use of biometric technologies.[281] For example, the Washington Privacy Act, which failed to pass in 2020, would have regulated law enforcement use of facial recognition technologies.[282]

While the future of federal privacy regulation remains unclear, states are at the forefront of privacy legislation.[283] Although a patchwork of state laws governing consumer privacy will make compliance difficult, some states are positioned to enforce standards regarding law en-

[278] *See 2019 Consumer Data Privacy Legislation*, *supra* note 235; Stacey Gray et al., *A New U.S. Model for Privacy? Comparing the Washington Privacy Act to GDPR, CCPA, and More*, FUTURE OF PRIV. F. (Feb. 12, 2020), https://fpf.org/2020/02/12/a-new-model-for-privacy-in-a-new-era-evaluating-the-washington-privacy-act/ (explaining that CCPA and WPA are both state comprehensive privacy schemes that would address consumer privacy harms, but neither place strong protections against law enforcement access to consumer genetic data).

[279] *See* Stacey Gray, *Long Overdue: Comprehensive Federal Privacy Law*, FUTURE OF PRIV. F. (Nov. 15, 2018), https://fpf.org/2018/11/15/fpf-comments-on-a-national-baseline-consumer-privacy-law/.

[280] *See* Cameron F. Kerry, *Game on: What to Make of Senate Privacy Bills and Hearing*, BROOKINGS (Dec. 3, 2019), https://www.brookings.edu/blog/techtank/2019/12/03/game-on-what-to-make-of-senate-privacy-bills-and-hearing/.

[281] *See 2019 Consumer Data Privacy Legislation*, *supra* note 235.

[282] Pollyanna Sanderson et al., *It's Raining Privacy Bills: An Overview of the Washington State Privacy Act and Other Introduced Bills*, FUTURE OF PRIV. F. (Jan. 13, 2020), https://fpf.org/2020/01/13/its-raining-privacy-bills-an-overview-of-the-washington-state-privacy-act-and-other-introduced-bills/.

[283] Gray et al., *supra* note 278.

forcement use of DTCs and publicly available genealogy databases.[284] In addition to potential state privacy laws, local police departments may institute their own regulations regarding use and access of DTCs. While the current DOJ interim policy sets standards for the use of DTCs, localities may choose the implementation of these standards, such as disallowing use altogether and setting stringent standards on use.[285] Furthermore, though local policies may do some of the work of a comprehensive regulatory scheme, differences between local regulations run the risk of creating a miscellany of privacy protections across the nation for the use of public genealogy and DTC genetic databases.

The question remains whether privacy legislation is best positioned to solve issues pertaining to law enforcement access to DTC genetic data. Privacy advocates largely agree that some regulation must exist to protect genetic privacy.[286] Some would argue that this is a question best left to the courts, though awaiting a potential court ruling can be an excessively lengthy process while technology continues to rapidly evolve.[287] DTC genetic companies continue to proliferate different standards regarding how, as well as when, they will share information with law enforcement, necessitating the creation of a federal standard.[288]

The DOJ is well-positioned to set enforceable rules through policy measures, regulating all federally funded law enforcement. The DOJ's recent interim policy did restrict law enforcement's use of DTCs but only in a limited capacity.[289] The interim policy did not address issues regarding efficacy and viability of DTC data, questions of transparency, or the issue of a warrant requirement. Law enforcement access to DTC databases does not follow the stringency that accompanies searches of law enforcement databases, nor do all DTC databases require a warrant prior to allowing law enforcement access to genetic information.[290] In essence, consumers of DTC genetic tests have significantly lessened genetic privacy rights, even when compared with individuals who were arrested based on probable cause.[291]

[284] *See, e.g.*, H.B. 1189, 2020 Leg., Reg. Sess. (Fla. 2020); S.B. 980, 2019-2020 Leg., Reg. Sess. (Cal. 2020).

[285] *See* Van Ness, *supra* note 262.

[286] *See* Katelyn Ringrose, *FPF and Privacy Analytics Identify "A Practical Path Toward Genetic Privacy"*, FUTURE OF PRIV. F. (Apr. 23, 2020), https://fpf.org/2020/04/23/fpf-and-privacy-analytics-identify-a-practical-path-toward-genetic-privacy/ (arguing that there needs to be a practical way of ensuring genetic privacy, either through technical means or organizational controls).

[287] *See* Lyria Bennett Moses, *Recurring Dilemmas: The Law's Race to Keep up with Technological Change*, 7 U. ILL. J.L. TECH. & POL'Y 239, 285 (2007).

[288] *See* Bala, *supra* note 238.

[289] *See* DOJ INTERIM POLICY, *supra* note 78, at 1.

[290] *See* Claire Abrahamson, *Guilt by Genetic Association: The Fourth Amendment and the Search of Private Genetic Databases by Law Enforcement*, 87 FORDHAM L. REV. 2539, 2543 n.24 (2019); *see also supra* Table 1.

[291] *See* Bala, *supra* note 238.

A 2015 task force report on 21st century policing, sponsored by the Office of the President and directed to law enforcement on the topic of Fourth Amendment searches, stated that law enforcement agents should be required to seek consent before a search and explained that individuals have the right to refuse consent when there is no warrant or probable cause.[292] The report also stated that officers should ideally obtain written acknowledgement that they sought consent to a search.[293] When revising their interim policy for 2020, the DOJ should address critical lapses in consumer privacy by simplifying the process for notice and consent, disallowing familial matching, and clarifying the need for a warrant prior to engaging in genetic searches.

This Article agrees that genetic databases pose incredible benefits when it comes to crime solving, providing solace for victims' families, and aiding in exonerating innocent suspects. However, those benefits should not and do not outweigh the potential harms associated with unwarranted and unregulated law enforcement access. Privacy-centric guidelines, as well as ethical and evidentiary-bound processes, must be enacted to ensure the maximizing of benefits and the minimizing of privacy harms associated with law enforcement access. A warrant standard for accessing genetic information is paramount, and the DOJ is well-positioned to require such a standard. Furthermore, more empirical research is necessary to understand the quality and efficacy of consumer databases prior to their use by law enforcement. Familial matching poses incredible privacy risks, far beyond those posed by ascertaining direct hits, including harms to civil liberties, lack of notice and choice for family members, and lack of transparency. It is best practice that law enforcement cease utilizing commercial and publicly available genealogy databases until, and if, these concerns are addressed.

Perhaps the most significant policy recommendation is to re-establish the National Commission on Forensic Science (NCFS). In 2013, the DOJ and NIST established the NCFS as a Federal Advisory Committee.[294] The NCFS was intended to promote scientific validity and improve the coordination of forensic science.[295] The NCFS was composed of federal, state, and local forensic science service providers; research scientists and academics; law enforcement officials; prosecutors, defense attorneys, and judges; and other stakeholders from across the country.[296] However, the NCFS was discontinued by the DOJ in 2017, with the intention of maintaining its activities in-house.[297] External ac-

[292] OFF. OF CMTY. ORIENTED POLICING SERVS., FINAL REPORT OF THE PRESIDENT'S TASK FORCE ON 21ST CENTURY POLICING 27 (2015).
[293] *Id.*
[294] NAT'L COMM'N ON FORENSIC SCI., NAT'L INST. OF STANDARDS & TECH., REFLECTING BACK – LOOKING TOWARD THE FUTURE 1 (2017) [hereinafter REFLECTING BACK].
[295] *See id.*
[296] *Id.*
[297] *See id.* at 2.

countability, effective information on how searches are conducted, and reliable regulation is incredibly important in the genetic evidence arena. Without an independent authority to hold law enforcement accountable to a strict set of regulations and norms, Fourth Amendment violations are bound to occur.

While moving forward with recommendations on the issue of law enforcement use of DTCs, myriad stakeholder concerns, including new quality assurance standards, must be addressed.[298] Cognizable stakeholders include law enforcement, privacy advocates, industry experts from DTC companies, and experts in electronic evidence standards. Concerns will differ amongst each stakeholder group, and a confluence of views is needed to reach consensus on transparent and thorough protocols for law enforcement use of DTC databases. As discussed above, there is a clear need for a collaborative and independent commission to provide timely advice and policy recommendations for the rapidly evolving technology of familial matching in the field of forensic science.[299]

CONCLUSION

With regulatory gaps clearly needing to be addressed, the road to robust genetic privacy in the United States is a long one. There are numerous regulatory mechanisms that could allow for appropriate crime-solving mechanisms while ensuring strong privacy safeguards. Though regulation may come in the form of a federal or state privacy bill and could be aided by an internal shift in DTC genetic testing services' practices, the likely next step in ensuring protections is a federal law enforcement policy outlining the privacy concerns of utilizing DTC databases and the importance of adhering to a warrant requirement for direct searches. While such a policy will be helpful in the near term, legislative regulation is necessary to create lasting protections. When crafting federal or state-level regulations, the legislature must consider the evidentiary value of DTC genetic data to understand the benefits and risks associated with the use of commercially generated profiles in a criminal justice setting. Finally, attention must be paid to issues of scientific concern—including concerns around the quality of testing laboratories and the efficacy of certain data analysis methods.

[298] *See About us*, SCI. WORKING GRP. ON DNA ANALYSIS METHODS, https://www.swgdam.org/about-us (last visited Nov. 13, 2020).

[299] *See, e.g.*, REFLECTING BACK, *supra* note 294, at 1–2.