A principled approach to cross-sector genomic data access

Smith, Marcus; Miller, Seumas

DOI
10.1111/bioe.12919

Publication date
2021

Document Version
Final published version

Published in
Bioethics

Citation (APA)
Smith, M., & Miller, S. (2021). A principled approach to cross-sector genomic data access. *Bioethics*, *35*(8), 779-786. https://doi.org/10.1111/bioe.12919

Received: 5 October 2020 | Revised: 1 April 2021 | Accepted: 30 May 2021

DOI: 10.1111/bioe.12919

bioethics WILEY

ORIGINAL ARTICLE

A principled approach to cross-sector genomic data access

Marcus Smith[1] | Seumas Miller[2,3,4]

[1]Center for Law and Justice, Charles Sturt University, Canberra, Australia

[2]Australian Graduate School of Policing and Security, Charles Sturt University, Canberra, Australia

[3]Delft University of Technology, The Hague, The Netherlands

[4]University of Oxford, Oxford, UK

Correspondence
Seumas Miller, Australian Graduate School of Policing and Security, Charles Sturt University, Level 1, 10-12 Brisbane Avenue, Barton ACT 2600
Email: semiller@csu.edu.au

Abstract

Genomic data is growing in importance as scientific knowledge and technology develop, and in availability, as direct-to-consumer genomic health testing and recreational genealogy services become more widely utilized. Access to genomic data needs to be considered in light of individual privacy. Cross-sector use of genomic health and ancestry data, and by law enforcement in particular, raises ethical questions and requires appropriate regulation. This article discusses the significance of genomic data and focuses on investigative genetic genealogy, namely the use of genomic health and ancestry data to advance law enforcement, as an example of cross-sector use. An ethical framework is developed that contributes to a more principled approach to genomic data access.

KEYWORDS

direct-to-consumer genomic testing, DNA evidence, genomic data, investigative genetic genealogy, privacy, recreational genealogy

1 | INTRODUCTION

Genomic data is central to advancements in medical research and promises further unprecedented individual and public health benefits, including improved diagnosis and treatment of some of the most serious afflictions, such as cancer, heart disease, and inherited disorders. Genomics refers to the holistic study of the entire human genome, in contrast to genetics, which refers to the study of single genes and how related traits or conditions are passed on to subsequent generations.[1]

Genomic data contains fundamental health and ancestry information about human beings, and its scope continues to expand with the ongoing development of scientific knowledge and advances in technology. The number and size of genomic databases established by public and commercial organizations are also increasing. Commercial health and ancestry testing,[2] marketed directly to consumers, has also grown significantly, and the number of individuals included in these databases is now of the order of tens of millions. The security of genomic information and the circumstances under which it may be accessed is an associated developing issue.

In recent years, access to genomic health, and particularly ancestry, data by law enforcement agencies (known as investigative genetic genealogy, IGG) has highlighted the fact that this information may be used in ways that were neither consented to nor originally intended.[3] Depending on the circumstances under which it is obtained, there may

[1]National Human Genome Research Institute (NHGRI). (2019). *A brief guide to genomics*. Retrieved from https://www.genome.gov/about-genomics/fact-sheets/A-Brief-Guide-to-Genomics (last accessed June 1, 2021).

[Correction added on 15 September 2021, after online publication: Corresponding author address has been changed in this version.]

[2]We define 'commercial health and ancestry testing' as that conducted by a private company for profit, as opposed to that undertaken by a government health or law enforcement agency.

[3]The IGG technique has been used recently in relation to commercial databases, for example the 'Golden State Killer' case, discussed later in the article, where it was used to apprehend and prosecute a serial killer in the United States. This is in contrast to other examples, such as the use of the national DNA database created in Sweden for health purposes (known as the PKU Registry) in forensic investigations: Mendelsohn, T. (2016, Jun 7). Sweden's national DNA database could be released to private firms. *Ars Technica*. Retrieved from https://arstechnica.com/tech-policy/2016/07/sweden-national-dna-database-private-firms/ (last accessed June 1, 2021).

be some justification for its use in relation to serious crimes such as murder or assault. However, this type of access undermines the existing law and procedures established in liberal democracies relating to the collection of genomic data from suspects that have been established over the past 20 years, raising questions regarding whether law enforcement agencies are conducting their investigations in accordance with the law.

This article discusses the use of genomic health and ancestry data by law enforcement agencies (which we term cross-sector use) as an example to inform ethical principles for genomic data access. The first part of the article considers genomics in healthcare, including the benefits accruing from the increased availability of genomic data in recent years, and the implications and issues associated with the use of genomic data collected for this purpose. The second part focuses on the use of genomic data in law enforcement investigations and on cases where genomic health and ancestry data has been used, as an example of cross-sector use. The third part explores the issue further, outlining a principled approach to genomic data access that draws on ethical principles of security, privacy and joint rights to argue that cross-sector genomic data should only be used in law enforcement investigations in certain circumstances.

2 | GENOMIC DATA IN HEALTHCARE

Since the early 1990s, genomics has become increasingly important in understanding and treating health conditions, raising a broad range of ethical questions about how it, and the information it generates, should be used. Interest in genomics grew rapidly during the completion of the Human Genome Project (between 1990 and 2003), undertaken by the United States Department of Energy and the National Institutes of Health with the aim of locating and sequencing all human genes.[4]

Genomics is becoming increasingly sophisticated and can now provide predictive health screening capable of identifying a predisposition to specific diseases later in life. It can inform lifestyle choices to improve health outcomes, and facilitate ancestry screening to determine which ethnic population or global region a person descends from and identify genetic relatives. The *All of Us* initiative in the United States, which plans to sequence the genomes of a million Americans, and the *100,000 Genomes Project* in the U.K. are early steps toward population-wide databases that will further expand the importance of genomic data and inform medical research.[5] Potential public health benefits from genomics include new or improved interventions to prevent and treat disease, and a better understanding of populations' predispositions to specific diseases, leading to improved public health planning.

Genomic data is sensitive and can reveal information about a person's health, susceptibility to disease, ethnic background, paternity, and relationship to others. It can potentially be integrated with other population and economic data to inform public health interventions, targeting and expenditure. The Human Genome Organisation's *Imagined futures* document outlines likely future issues for genomics associated with data security, privacy and trust.[6] Considerations associated with storing genomic data in repositories include consent to data inclusion, how data can be used, the threats of human error and hacking, and the interoperability of different storage formats. More recently, the Global Alliance for Genomics and Health has engaged in the development of standards and frameworks for international genomic data sharing,[7] and the Nuffield Council on Bioethics has identified scientific developments related to crime and security as a key issue on the horizon for the field.[8]

Previous genomic research has predominantly focused on specific technologies, such as gene editing,[9] rather than on the vast amount of data generated, which is now emerging as an issue of equivalent importance. While there are existing ethical guidelines and legislation relating to the use of genomic data in clinical practice, there are significant gaps in relation to its cross-sector use, such as the use of data in a law enforcement investigation when it had initially been created for health or ancestry purposes. New approaches to regulation and consent may be needed to address the rapid expansion in genomic data, and the ways in which it is being used, particularly given the ease with which it can now be generated.[10]

The disruption of traditional healthcare by the biotechnology sector, the financial incentive to commercialize new technology, and public interest in genomics have led to the emergence of direct-to-consumer genomics companies that offer mail-order testing for health conditions and ancestry.[11] These include 23andMe, Genomic Diagnostics, Ancestry.com and FamilyTreeDNA.[12] GEDmatch enables users to upload data produced by other companies to search for potential unknown genetic relatives. As the cost of the associated technology has decreased, so has the cost of services, which have become increasingly popular. Consumers receive testing equipment in the mail, carry out their own cheek swab, and return the swab to the company: no medical practitioner, nurse, or other health

[4]See NHGRI, op. cit. supra note 1.

[5]Feero, F., Wicklund, C. A., & Veenstra, D. (2018). Precision medicine, genome sequencing, and improved population health. *Journal of the American Medical Association*, *319*, 1979–1980.

[6]Capps, B., Chadwick, R., Chalmers, D. R. C., Clarke, A., Clayton, E., Liu, E., & Winslett, M. (2013). *Imagined futures: Capturing the benefits of genome sequencing for society*. London, UK: HUGO Committee on Ethics, Law and Society.

[7]Global Alliance for Linked Genomics and Health (GALGH). (2019). *Enabling responsible linked genomic data sharing for the benefit of human health*. Retrieved from https://www.ga4gh.org (last accessed June 1, 2021).

[8]Nuffield Council on Bioethics. (2019). *Horizon scanning workshops*. Retrieved from https://nuffieldbioethics.org/future-work/horizon-scanning-workshops (last accessed June 1, 2021).

[9]Gyngell, C., & Savulescu, J. (2015). The medical case for gene editing. *Ethics in Biology, Engineering and Medicine*, *6*, 57–66.

[10]Kaye, J., Whitley, E. A., Lund, D., Morrison, M., Teare, H., & Melham, K. (2015). Dynamic consent: A patient interface for twenty-first century research networks. *European Journal of Human Genetics*, *23*, 141.

[11]Commercial ancestry databases are also referred to as recreational genealogical databases.

[12]See e.g. https://www.23andme.com; https://www.ancestry.com; https://www.familytreedna.com; https://www.gedmatch.com (last accessed June 1, 2021).

professional is involved. The company's laboratory undertakes the testing and provides the results by email, discarding the biological sample but retaining the genomic data. It has been reported that by 2019, more than 15 million people had submitted their genomic data to Ancestry.com, and more than 10 million people had submitted it to 23andMe, along with several million more to other companies.[13]

23andMe has been offering health and ancestry tests since 2007. Between 2013 and 2017, the company varied its services as the Federal Drug Administration (FDA) increased regulatory requirements, and since 2017 it has provided 'genetic risk tests' (as opposed to 'diagnostic tests') for a range of conditions, including Alzheimer's disease, Parkinson's disease and celiac disease (FDA, 2017). The company's privacy statement states that they share information, including genomic information,[14] with third parties, including not only aggregate information about the prevalence of genetic traits in their customers, but also genomic data about individuals, as required by 'laws, regulations, judicial or other government subpoenas, warrants, or orders'.[15] There is scope for consumers to opt out of using these services if the potential for their data to be used in a law enforcement investigation concerns them.

As analytic capabilities and artificial intelligence techniques develop, all forms of data are increasingly being collected by governments and the private sector. With the advent of smartphones, metadata collection by governments has been a prominent issue over the past decade;[16] facial recognition databases have been proposed by governments in recent years, incorporating images from passports and driver's licences;[17] and social media companies have been criticized for their collection of web-browsing data for advertising purposes.[18] The capacity to integrate genomic data with other data types, such as metadata, financial data and biometrics, adds to these concerns, particularly given its significance: that is, genomic data is unlike most other kinds of sensitive personal information, as an individual's genome is a reliable life-long identifier. Moreover, the genome of a person is constitutive of that person's individual-specific (biological) identity.[19]

The establishment of comprehensive, integrated government databases of the personal information of citizens has the potential to create a power imbalance in liberal democracies, as is increasingly evident from the establishment of social credit systems in authoritarian states such as China.[20] Individual rights to privacy and autonomy, appropriate regulation, and democratic accountability must be carefully considered when developing law and policy in relation to genomic and other databases—genomic and other databases rise to new safety and security concerns, such as the possibility of identity theft, even as they strengthen privacy protection by reducing unauthorized access to private information.[21]

In light of these potential developments, data security in healthcare is becoming a far more significant issue for governments and communities as a whole, as greater volumes of data are generated and used in novel ways. Instances of large-scale data breaches involving national institutions, governments and businesses are becoming common. Whether data is held in the public or the private sector, genomic databases are a rich data source that hold valuable information relating to the health and associated vulnerabilities of populations, are vital to medical research and population health, and can be used to identify individuals.

3 | GENOMIC DATA IN LAW ENFORCEMENT

Genomic data, referred to in law enforcement as DNA evidence, has proven highly effective in the identification of offenders and the prosecution of serious crimes such as murder and sexual assault, as well as in exonerating innocent persons. It has been used for this purpose since the DNA profiling technique was first discovered in the late 1980s. The technique has traditionally involved comparing a DNA profile from a biological sample at a crime scene with one obtained from a suspect or a database of known offenders, in order to establish a link between an individual and the crime. A DNA profile is created by analysing repetitive sequences of DNA in the genome. It is essentially a set of numbers that contains only identifying information, rather than an indication of health conditions or physical traits.[22]

Genomic databases, referred to as DNA databases in law enforcement, are well established, alongside other biometrics such as fingerprint databases and, most recently, facial recognition databases. Genomic databases for this purpose were established in the late 1990s and early 2000s.[23] The national database in the United States includes over 18 million individuals, and that in the U.K., over

[13]Regalado, A. (2019, Feb 11). More than 26 million people have taken an at-home ancestry test. *MIT Technology Review*. Retrieved from https://www.technologyreview.com/2019/02/11/103446/more-than-26-million-people-have-taken-an-at-home-ancestry-test/ (last accessed June 1, 2021).

[14]23andMe Privacy Policy, section 2(b)(ii). Retrieved from https://www.23andme.com/en-int/about/privacy/ (last accessed June 1, 2021).

[15]Ibid: section 4(e).

[16]Walsh, P., & Miller, S. (2016). Rethinking 'five eyes' security intelligence policies and practices post 9/11/post-snowden. *Intelligence and National Security*, *31*(3), 345–368; Miller, S. (2018). Machine learning, ethics and law. *Australian Journal of Information Systems*, *22*, 1–13.

[17]Smith, M., & Miller, S. (2021). The ethical application of biometric facial recognition technology. *AI and Society*, 36.

[18]Chen, B. (2018, May 16). Google's file on me was huge. Here's why it wasn't as creepy as my Facebook data. *New York Times*. Retrieved from https://www.nytimes.com/2018/05/16/technology/personaltech/google-personal-data-facebook.html (last accessed June 1, 2021).

[19]It should be noted that this is a developing issue as there are complexities in analysing and linking genomic data. While this is routinely done in relation to forensic DNA profiling, the vast amount of data included in the genome make the linking of large genomic datasets a challenging proposition, although one that is becoming more feasible as new technology develops. See, for example, Mittelstadt, B. D., & Floridi, L. (2015). The ethics of big data: Current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics*, *22*, 303–341.

[20]Qiang, X. (2019). The road to digital unfreedom: President Xi's surveillance state. *Journal of Democracy*, *30*, 53–67.

[21]Henschke, A. (2017). *Ethics in an age of surveillance*. Cambridge, U.K.: Cambridge University Press.

[22]Smith, M. (2015). *DNA evidence in the Australian Legal System*. New York, NY: LexisNexis.

[23]Ibid.

5 million individuals.[24] These databases generally only include DNA profiles from convicted offenders. The inclusion of suspects in these databases was considered in the U.K. case, *S and Marper v United Kingdom*,[25] with the European Court of Human Rights subsequently ruling that in light of Article 8 of the European Convention on Human Rights:[26]

> …the blanket and indiscriminate nature of the powers of retention of the fingerprints, cellular samples and DNA profiles of persons suspected but not convicted of offences, as applied in the case of the present applicants, fails to strike a fair balance between the competing public and private interests and that the respondent State has overstepped any acceptable margin of appreciation in this regard.[27, 28]

There have been proposals in some countries to establish universal forensic DNA databases (including a country's entire population) for law enforcement purposes, to improve the investigation and prosecution of crime. This is a controversial proposal, both ethically and in terms of popular support—many would object to a national database of DNA profiles, with individuals included irrespective of whether they have been convicted of committing a crime, as an affront to their individual privacy and autonomy.[29] However, it may occur incidentally at some point in time, if national genomic databases are established for health purposes and law enforcement agencies are able to access this data for investigations. A similar phenomenon is taking place with biometric facial recognition databases. By drawing on repositories of driver's licence and passport images, a near universal biometric facial recognition database of adult populations can be established, as some governments have sought to do by amending the regulations that govern access to existing image repositories.[30] The prospect of a similar development (a universal database) in genomics is highlighted by genomic testing becoming more broadly used, most recently being proposed to enhance immigration procedures in the United States and other developed countries to adjudicate familial claims that have implications for citizenship,[31] and in the national health genomic databases described above, such as the *All of Us* database in the United States.

The expansion of genomic datasets (which are potentially available for law enforcement access) can be contrasted with the existing legislative requirements that govern how genomic data can be used by law enforcement in criminal investigations. In legal systems around the world, legislation and case law govern how DNA evidence can be used in law enforcement investigations. Forensic procedure legislation and evidence law govern the circumstances in which forensic samples may lawfully be obtained and retained, the conditions under which the national DNA database can be searched, and when evidence obtained through a forensic procedure may be admitted at trial.[32] However, provisions exist in evidence or criminal procedure law in most jurisdictions to allow evidence that has been obtained improperly, or not according to legal process, for example, if the desirability of admitting the evidence outweighs the undesirability of admitting evidence that has been obtained in the way in which it was. This means that if a court deems evidence to be so important that it would be unjust for it not to be used, it may allow the use of that evidence at trial even if investigators obtained it illegally. In the case of scientific evidence, such as genomic data, the court will also be concerned that it can be relied upon. This includes determining whether the expert presenting the evidence has the appropriate knowledge, skill, experience, training, or education; whether the evidence is based on reliable scientific principles and methods; and whether it has been tested, subjected to peer review, and is generally accepted in the scientific community.[33]

It is in this context that the issue of cross-sector use of genomic health and ancestry data by law enforcement in criminal investigations arises. A contemporary option for law enforcement agencies conducting an investigation, if they do not obtain a match for a suspect's DNA profile on their national database, is to search a commercial genomic database.[34] Law enforcement agents are effectively searching for a potential common ancestor via whom they can identify their suspect. This method is of much broader scope than the one-to-one matching against a database of convicted offenders (regulated by legislation) that they would use when searching a traditional DNA database. IGG enables searching of up to fourth cousins (potentially 100 people) of the single donor that submitted their genomic data to a health or ancestry testing company.[35] Given that more than 26 million people,[36] mostly in the United States, have submitted their genomic data for testing to one of these companies,

[24]Federal Bureau of Investigation (FBI). (2020). *NDIS statistics*. Retrieved from https://www.fbi.gov/services/laboratory/biometric-analysis/codis/ndis-statistics (last accessed June 1, 2021); United Kingdom Government Statistics (UKGS). (2020). *National DNA database statistics*. Retrieved from https://www.gov.uk/government/statistics/national-dna-database-statistics (last accessed June 1, 2021).

[25][2008] ECHR 1581.

[26]Council of Europe, *European Convention for the Protection of Human Rights and Fundamental Freedoms*, as amended by Protocols Nos. 11 and 14, 4 November 1950.

[27][2008] ECHR 1581, 119.

[28]In contrast to the European *Marper* case, the U.S. Supreme Court explicitly permitted the legality of arrestee DNA collection in *Maryland v. King*, 569 U.S. 435 (2013).

[29]Smith, M. (2018). Universal forensic DNA databases: Balancing the costs and benefits. *Alternative Law Journal*, *43*, 131–135.

[30]Mann, M., & Smith, M. (2017). Automated facial recognition technology: Recent developments and approaches to oversight. *University of New South Wales Law Journal*, *40*, 121–145.

[31]Lee, C., & Voigt, T. (2020). DNA testing for family reunification and the limits of biological truth. *Science, Technology, & Human Values*, *45*, 430–454.

[32]For example in the United States, the DNA Fingerprint Act of 2005 allows an arrestee's profile to be uploaded to the federal database at the time of arrest. If the arrestee is not subsequently charged with an offence, the burden lies with the arrestee to file a court order stating that the charges have been dismissed.

[33]See e.g. in the United States, Federal Rules of Evidence, rule 702; *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993).

[34]Phillips, C. (2018). The Golden State Killer investigation and the nascent field of forensic genealogy. *Forensic Science International: Genetics*, *36*, 186–188.

[35]Ibid.

[36]See Regalado, op. cit. supra note 12.

multiplying that figure by 100 provides an indication of the potential scope of this investigative technique.

There is a detailed process that law enforcement agents would need to undertake next to identify their suspect, if they found an individual on, for example, Ancestry.com, that they believed was a second cousin of the suspect. The process indicates that a significant number of people would need to be investigated and ruled out. Investigators would hypothesize that there is a common set of great-grandparents, and using government records of births, deaths and marriages would construct a family tree of three generations up to that point. They would then construct four family trees of the great grandparents and narrow down the list of grandparents, parents, great uncles and aunts, uncles and aunts, siblings, and first and second cousins on the basis that some may be deceased, live overseas, or could be excluded based on other data such as age or eyewitness reports. Finally, they would need to reduce this to a small number of individuals that they would then subject to overt or covert action, possibly including covertly obtaining a sample of their biological material so as to compare their DNA profile with the crime scene sample.[37]

This technique therefore involves the use of genomic data provided not for the purposes of a law enforcement investigation, but by a member of the public to obtain information about their personal health and genetic ancestry, who did not anticipate its use for this purpose. It raises significant questions regarding privacy and autonomy. It has already been used as a last resort in identifying offenders who have committed serious offences, and given the public benefit in arresting a murderer and preventing future serious crimes from being committed, its use in this context would be less controversial. However, given that there is no legislative backing for the use of this technique (although the release of data for law enforcement purposes may be mentioned in the fine-print terms and conditions), it would be concerning if it came to be used for minor offences where the public benefit is far more limited.

Evidence obtained as a result of this technique has been admitted at trial on a discretionary basis in the United States, owing to its probative value in a significant case. It was used in the identification and conviction of former police officer Joseph DeAngelo, who has since been convicted of 13 murders over a 12-year period in the 1970s and 1980s, which was popularly referred to as the 'Golden State Killer' case in the United States.[38] It has been reported that law enforcement used the GEDmatch site to identify Deangelo as one of a number of potential suspects. After identifying a distant relative of their suspect, they traced a family tree back to the 1880s before undertaking surveillance on a number of people and finally arresting DeAngelo after obtaining DNA from his rubbish and confirming a match with the crime scene DNA profile. It has been reported that law enforcement has since used GEDmatch to search for unknown suspects in more than 100 investigations, leading to other arrests.[39]

There are a number of potential problems with this type of activity. As has already been noted, it is unregulated by existing forensic procedure legislation and amounts to a fishing expedition rather than a targeted, proportionate (and legal) law enforcement investigation. It places a large number of genetic relatives under suspicion, from whom law enforcement may deem it necessary to covertly take biological samples. As noted in the first part of this article, companies such as GEDmatch and 23andMe now state in their privacy policy that genomic data may be released to law enforcement; however, the implications extend beyond the individual that submitted their genomic data to their genetic relatives.[40]

4 | ETHICAL ANALYSIS

The expanding use of genomic data that has been described above raises a number of pressing ethical concerns. Fundamental moral principles must continue to be valued in liberal democracies, notwithstanding the benefits to individual and public health, and community safety that the unrestrained use of this data may afford.[41] The cross-sector use of genomic data can be understood from the perspectives of individual privacy, autonomy, public safety, and democratic accountability in various domains. These domains include law enforcement, public health, medical research, and private sector commercialization. Central to the ethical, legal and policy issues associated with genomic data is the tension that exists between the legitimate collection of information by law enforcement, health and other government agencies, as well as commercial service provision, on the one hand, and individual rights to privacy and autonomy on the other. In a criminal law and national security context, the threat of terrorism over the past 20 years has resulted in ever greater powers for law enforcement and intelligence agencies to collect evidence and conduct surveillance in order to prevent, detect and disrupt these activities, and these have extended to other forms of crime.[42]

It is sometimes assumed that the relationship between, for instance, autonomy and security is a zero-sum relationship and that, therefore, any increase in security that decreases someone's autonomy will necessarily lead to an overall loss in autonomy. This assumption is false; or, at least, it is often false. For instance, if the police have access to the DNA of all persons with a record of having

[37]Scudder, N., McNevin, D., Kelty, S. F., Funk, C., Walsh, S. J., & Robertson, J. (2019). Policy and regulatory implications of the new frontier of forensic genomics: Direct-to-consumer genetic data and genealogy records. *Current Issues in Criminal Justice*, *31*, 194–216.

[38]Gold, R. (2019). From swabs to handcuffs: How commercial DNA services can expose you to criminal charges. *California Western Law Review*, *55*, 491–519.

[39]DeLisi, M. (2018). Forensic epidemiology harnessing the power of public DNA sources to capture career criminals. *Forensic Science International*, *291*, 20–21.

[40]Murphy, E. (2018). Law and policy oversight of familial searches in recreational genealogy databases. *Forensic Science International*, *292*, 5–9.

[41]Kaye, J., Whitley, E. A., Lund, D., Morrison, M., Teare, H., & Melham, K. (2015). Dynamic consent: A patient interface for twenty-first century research networks. *European Journal of Human Genetics*, *23*, 141–146.

[42]Miller, S. (2009). *Terrorism and counter-terrorism: Ethics and liberal democracy*. Oxford, UK: Blackwell.

committed serious crimes, then, given that the number of such persons is small but they commit a large percentage of serious crimes, their loss of autonomy in respect of control over their DNA may be more than offset not only by an overall reduction in harm, but also by an overall increase in autonomy. This is because many persons will enjoy an increase in their autonomy, namely those persons who would have been future victims of crime had the offenders in question not been incarcerated for their past crimes, or deterred from future crimes, as a result of criminal investigators' access to the DNA of these offenders. Here it is important to note that serious crimes such as grievous bodily harm, rape and domestic violence are in large part attacks on autonomy. An analogous point concerning an assumed zero-sum relationship can be made in respect of privacy and security, especially when it is taken into account that infringements of privacy can often be mitigated, such as, in the case of law enforcement's use of big-data analytics, by processes of anonymization of data prior to the point of identification of suspects. That said, increases in law enforcement powers, including increased cross-sector genomic data access, have the potential to unacceptably compromise autonomy, privacy, and other liberal democratic principles.

Public safety and security are fundamental values in liberal democracies, as in other polities, including many authoritarian ones. However, liberal democracies are also committed to democracy and individual privacy and autonomy, and, therefore, to democratic accountability.[43] Accordingly, fundamental ethical principles must continue to be valued, notwithstanding the benefits to community safety that access to commercial genomic databases, such as 23andMe or Ancestry.com, can provide by enabling law enforcement to detect and convict perpetrators of serious crimes. While debates will continue between proponents of security, on the one hand, and defenders of privacy, on the other, there is often a lack of clarity in relation to the values or principles allegedly in conflict—these principles and the relationships between them will now be discussed.

The notion of privacy has proved difficult to explicate adequately. Nevertheless, there are a number of general points that can be made. First, privacy is a right that people have in relation to other persons and organizations with respect to: (a) the possession of information (including genomic data) about themselves by other persons and by organizations, for example personal health, familial and identity information stored in genomic databases; or (b) the observation/perceiving of themselves—including of their movements, relationships and so on—by other persons, for example via law enforcement having access to their genomic data that facilitates linkage with a particular location based on an analysis of biological material deposited at that site.[44] Genomic data is therefore implicated in both informational and observational concerns.

Second, the right to privacy is closely related to the more fundamental moral value of autonomy. Roughly speaking, the notion of privacy delimits an informational and observational 'space', namely the private sphere. This informational space includes genomic data; specifically, the data constituting a person's genome that is particular to that person and, relatedly, a person's DNA profile. However, the right to autonomy consists of a right to decide what to think and do, and the right to control the private sphere. So the right to privacy consists of the right to exclude organizations and other individuals (the right to autonomy) both from personal information, such as genomic data, and from observation and monitoring of where that person is, or has been. Naturally, the right to privacy is not absolute; it can be overridden. Moreover, its precise boundaries are unclear: a person does not have a right not to be observed in a public space, but, arguably, has a right for law enforcement agencies not to have access to their genomic data, although this right can be overridden under certain circumstances, namely if they have been convicted of a serious crime (their DNA profile will then be included in a forensic database). For instance, this right might be overridden if an individual is reasonably expected of being involved in a crime, and police have a warrant, approval from a judicial officer, legislative authority etc., and then only for the purpose of identifying persons who have committed a specific crime. If persons have committed a serious crime, such as murder or assault, in the past, it would be morally acceptable to utilize the retention of their genomic data (*as it relates to identity, not health conditions*) by including it in a database and matching against samples obtained from crime scenes. This is a specific and targeted measure to improve public safety, and even then the data can only be used in such a way that has been legislated for by a democratically accountable government. As discussed above, there are already millions of individuals in countries such as Australia, the U.K. and the United States included in forensic DNA databases of this type.

Third, a degree of privacy is necessary in order for people to pursue their personal projects, whatever those projects might be. For one thing, reflection is necessary for planning, and requires a degree of freedom from distracting intrusions, including intrusive surveillance, of others. For another, knowledge of someone else's health status, familial relationships or genomic identity can lead to that information and any associated vulnerabilities being exploited, or otherwise compromised. *Autonomy*—including the exercise of autonomy in the public sphere—requires a measure of privacy.

Thus far we have considered the rights of a *single* individual. However, it is important to consider the implications of the infringement, indeed violation, of the privacy of groups of people and, ultimately, of the whole citizenry by the state (and/or by other powerful institutional actors, such as corporations). Such violations on a large scale can lead to a power imbalance between the state and the citizenry and, thereby, undermine liberal democracy itself.

Accordingly, while it is morally acceptable to access genomic data for necessary circumscribed purposes, such as the provision of healthcare or medical research, or, with the consent of the relevant individuals, for ancestry testing, it would not be acceptable to collect this data in an indiscriminate manner without consent and with no legal authority, to investigate crime. However, the DNA profiles of

[43]Miller, S. (1997). Privacy and the internet. *Australian Computer Journal*, *29*(1), 12–16; Miller, S., & Gordon, I. (2014). *Investigative ethics* Ch. 10 op. cit.

[44]Ibid.

convicted offenders on forensic DNA databases are, and arguably ought to be, available for law enforcement purposes, for example to assist in the investigation of serious crimes. The issue that then arises is the determination of the point on the spectrum at which privacy and security considerations are appropriately balanced.

In light of the above analysis of privacy, we are entitled to conclude that some form of it is a constitutive human good. As such, infringements of privacy ought to be avoided. That said, as mentioned above, privacy can reasonably be overridden by security considerations under some circumstances, such as when lives are at risk. After all, the right to life is, in general, a weightier moral right than the right to privacy. Thus, utilizing genomic data in a forensic DNA database or from a suspect to investigate a serious crime such as a murder, if conducted under warrant or legislative provisions, is surely ethically justified. On the other hand, intrusive access to the genomic data of individuals, collected for another purpose, where those individuals have not had any contact with the criminal justice system, and the data was obtained without any legal authority, particularly in relation to relatively minor offences such as theft, is far less likely to be justified. Moreover, given the importance of, so to speak, the aggregate privacy of the citizenry, relatively small-scale threats to public safety are unlikely to be of sufficient weight to justify substantial infringements of privacy, for example unregulated access to the genomic relationships of millions of people by law enforcement agencies. Furthermore, regulation and associated accountability mechanisms need to be in place to ensure that, for instance, a genomic database created for a legitimate purpose, for example health or ancestry testing with the express consent of the individuals involved, is not accessed, except with the appropriate legal authority and in relation to the investigation of serious crimes.

Here we need again to stress the particular significance of genomic data but now elaborate on the reasons for this. Genomic data, and DNA profiles in particular, are (in effect, namely for our purposes here and, therefore, issues of gene-editing aside) unchanging and unalterable; therefore, they are a reliable life-long identifier. This means that they have greater utility for law enforcement than do other forms of personal data. However, it also means that there is much more at stake in terms of an individual's privacy and autonomy should this genomic data be provided to law enforcement or other agencies (including private sector ones). Moreover, the genome of a person is constitutive of that person's individual-specific (biological) identity. Accordingly, the threshold for the infringement of an individual's right to control access to their genomic data is higher than it is for most other personal information. And there is a further point here. For the genome of a person is not only constitutive of that person's individual-specific (biological) identity, that same genome is *in part* constitutive of the individual-specific (biological) identity of the person's relatives (to a decreasing extent depending on the degree of relatedness; for example a sibling is more related than a second cousin). Accordingly, there is a species of joint right to control genomic data in play here, and not merely an exclusively individual right.

Joint rights are rights that attach to individual persons but do so jointly.[45] Thus, roughly speaking, two or more agents have the right to some good if they each have a right to that good, no-one else has a right to that good, and if the individual right of one of these persons to the good is dependent on the individual rights of the others to the good.[46] The right to control one's genome data needs to be regarded, we suggest, as a (qualified)[47] joint right; that is, as a right jointly held with the individual's relatives. If these rights are, as we are suggesting, joint rights, then it follows that an individual may not have an exclusive individual right to provide his or her genomic data to direct-to-consumer genetic testing providers, or to law enforcement. Of course, when it comes to serious crimes, the consent of an individual regarding access to his or her genomic data is not necessarily required, for example if the individual is a past offender and hence his or her genomic data in the form of a DNA profile is held in a law enforcement database. However, in cases where identifying the person who has committed a crime relies on the genomic data of relatives known to be innocent, and the relatives in question have a joint right to the data in question, then it may be that *all* of these relatives need to have consented to the collection of the genomic data in question.[48] For in voluntarily providing their DNA to law enforcement, a person is, in effect, providing law enforcement with the partially overlapping DNA data of their relatives. But presumably a person does not have a moral right to decide to provide law enforcement with another person's DNA data. Accordingly, it seems that a person, A, does not have a moral right to *unilaterally* provide law enforcement with his or her own data, namely A's DNA data, given that in doing so A is providing to law enforcement the partially overlapping DNA data of A's relatives, B, C, D etc. Rather, A, B, C, D etc. have an (admittedly qualified) joint moral right to the DNA data in question, and, therefore, the right (being a joint right) has to be exercised jointly; that is, perhaps all (or most) have to agree. Naturally, as is the case with individual moral rights, joint moral rights can be overridden. For instance, A's individual right to know whether he is vulnerable to a hereditary disease might justify his providing his genomic data to health authorities and doing so without the consent of any of his relatives. Again, the joint moral right of a group of persons to refuse to provide law enforcement with the DNA data in a murder investigation, for instance, may well be overridden by their collective moral responsibility to assist the police.

[45]Miller, S. (1999). Collective rights. *Public Affairs Quarterly*, *1*(4), 331–346; Miller, S. (2001). *Social action: A teleological account*. Ch. 7. Cambridge, UK: Cambridge University Press; Miller, S. (2003). Institutions, collective goods and individual rights. *Protosociology*, *18*, 184–207; Miller, S. (2010). *The moral foundations of social Institutions: A philosophical study*. Ch. 2. Cambridge, UK: Cambridge University Press. [Correction added on 15 September 2021, after online publication: In-text citation for footnote 45 has been updated.]

[46]It is a qualified joint right given that the genomic data of any one of the persons is not identical to the genome data of the other persons, that is, the sets of genomic data are overlapping

[47]It is a qualified joint right given that the genomic data of any one of the persons is not identical to the genome data of the other persons; that is, the sets of genomic data are overlapping

[48]This consent issue adds to other problems that exist with direct-to-consumer genetic testing, such as the accuracy of the tests and the fact that the results are not provided in a clinical setting by a healthcare professional.

5 | CONCLUSION

We have described the cross-sector access of genomic data, collected for health and ancestry purposes, by law enforcement for criminal investigation purposes. It is likely that these practices, which have been documented in the United States, are also being undertaken in other liberal democracies, such as Australia and the U.K., although there is not currently any publically available data to support this. In light of these developments, we have outlined the relevant ethical principles and identified a number of actual or potential problems that arise.

The issues in this area cannot be framed in terms of a simple weighing of, let alone trade-off between, individual privacy rights versus the community's interest in public safety. The issues are far more ethically complex, and we conclude with three general points.

First, law enforcement access to and searching of the genomic data of citizens, held by private companies and created for specific purposes, without legislative oversight or regulation, and the utilization of this data in investigations, infringes privacy rights, has the potential to create a power imbalance between governments and citizens, and risks undermining important principles hitherto taken to be constitutive of the liberal democratic state, such as that an individual has the right to freedom from state interference absent prior evidence of violation by that individual of its laws, subject to transparent and appropriately justified exceptions.

Second, as part of the introduction of laws to regulate this activity, if these laws are deemed to be justified, the cross-sector use of genomic data in this way must be clearly and demonstrably justified in terms of efficiency and effectiveness in law enforcement investigations, and its use circumscribed accordingly, rather than by general appeal to community security or safety.

Finally, in so far as the use of genomic data created for health or ancestry purposes can be justified for the investigation of serious crimes, and privacy and other concerns mitigated, it is imperative that this use be subject to accountability mechanisms to guard against misuse. Moreover, the citizenry should be well informed about these systems and should have consented to the use of these systems for the specific, justified purposes in question: they should be publically debated, backed by legislation, and their operation subject to judicial review.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Marcus Smith https://orcid.org/0000-0001-9810-979X

AUTHOR BIOGRAPHIES

Dr Marcus Smith is Senior Lecturer in Law at Charles Sturt University and Adjunct Professor of Law at the University of Canberra. He holds a PhD in law from the Australian National University. His recent books include *Technology law* (Cambridge University Press, 2021), *Biometrics, crime and security* (Routledge, 2018) and *DNA evidence in the Australian legal system* (LexisNexis, 2016).

Professor Seumas Miller holds research positions at Charles Sturt University, TU Delft and the University of Oxford. He is the Principal Investigator on a European Research Council Advanced Grant. His recent authored books include *Institutional corruption* (Cambridge University Press, 2017) and *Dual use science and technology, ethics and weapons of mass destruction* (Springer, 2018).

How to cite this article: Smith M., & Miller S. (2021). A principled approach to cross-sector genomic data access. *Bioethics*, 35:779–786. https://doi.org/10.1111/bioe.12919