A sociotechnical approach to genomic data privacy: a comparative analysis

Jacob S. Sherkow *

Abstract

- The sharing of genomic data across international borders presents significant privacy law challenges.
- Secure computing environments on smartphones allow the storing and processing of sensitive data without the underlying data being shared with processors.
- A novel technology, described here, to process genomic data within a secure computing environment seems to comport with European Union and US privacy laws, despite their differing aims and rules.
- This technology suggests there may be technological solutions to privacy law fragmentation across jurisdictions, so long as data subjects socially trust the technology and have control over their data.

Introduction

Of all of the types of personal data, human genetic data is one of the most sensitive, constituting both a 'unique identifier and a person's book of life'.[1] Accordingly, the sharing of genetic data—and especially genomic data, the complete sequence of person's genetic makeup—remains legally restricted in various ways.[2] Advances in genomic sequencing and secure computing environments, however, have opened avenues towards, and increased demand for, easy and accessible genomic data-sharing.[3] One approach in particular—a novel system, described here, that sequesters genomic data within a secure computing environment—seems to be an especially viable approach due to the recent availability of secure computing environments on smartphones.[4] Whether such an approach satisfies international data privacy laws, or whether users would adopt the technology if further developed, requires careful analysis of the technical features of computationally secure genomic data as set against current privacy regimes. This 'sociotechnical approach'—marrying technological advancements with user trust in technology—suggests a path forward: a way to safely, securely, and legally share genomic data. This analysis—part of a larger grant project funded by the National Institutes of Health to develop a sociotechnical approach to genomic data-sharing—may be further used as a representative case study for sociotechnical solutions to various data privacy issues in secure computing environments.

In recent years, the demand for and use of genomic data have dramatically increased.[5] The rise of ever cheaper and more accurate genomic sequencing technologies has led to a proliferation of direct-to-consumer genetics companies marketing their services for a variety of purposes, both consequential and frivolous.[6] This

* Jacob S. Sherkow, College of Law, University of Illinois, Champaign, Illinois 61820, United States; Carle Illinois College of Medicine, University of Illinois, Urbana, Illinois 61801, United States; European Union Center, Urbana, Illinois 61801, United States; Carl R. Woese Institute for Genomic Biology, Urbana, Illinois 61801, United States. Email: jsherkow@illinois.edu. My thanks to Carl Gunter, Timo Minssen, Marcelo Corrales Compagnucci, Kathleen Liddell, and Mateo Aboy for their valuable insights during discussions about this work.

1 Ellen Wright Clayton and others, 'The Law of Genetic Privacy: Applications, Implications, and Limitations' (2019) 6 J Law Biosci 1, 2.

2 Regulation (EU) 2016/679 of the European Parliament and of the Council of April 27, 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (GDPR) [2016] OJ L119/1, art 9(1); 42 USC s 2000ff-5(b) (2022).

3 Ali Amr and others, 'Controlling My Genome with My Smartphone: First Clinical Experiences of the PROMISE System' (2022) 111 Clinical Res Cardiology 638.

4 'Secure Enclave' (Apple Platform Security, 7 May 2024) <https://support.apple.com/guide/security/secure-enclave-sec59b0b31ff/web> [https://perma.cc/N4UU-VAMA]; 'Trustee TEE' (Android Open Source Project, 29 April 2024) <https://source.android.com/docs/security/features/trusty> [https://perma.cc/D5AX-QYH5].

5 Eric D Green and others, 'Strategic Vision for Improving Human Health at The Forefront of Genomics' (2020) 586 Nature 683.

6 Jacob S Sherkow, 'The Myth of DNA Trade Secrecy' (2024) 75 U Cal LJ 1047, 1061–62.

is to say nothing of the dramatic expansion of genomic sequencing for medical purposes.[7] Yet, from a privacy perspective, sharing genomic data remains particularly fraught because individuals often seek to analyse their genomic data with third-party services that have varying levels of privacy protection and legal compliance.[8] This opens users to serious cybersecurity risks, such as the attack that befell the recently bankrupt direct-to-consumer genomic sequencing company 23andMe, where, in 2023, hackers gained access to the genomic data of over seven million individuals.[9] Given the size, complexity, and sensitivity of genomic data, securely transferring genomic information for analysis remains challenging.[10]

This article describes a novel solution to this problem, as developed under a grant from the National Institutes of Health (No. R01HG012249-01): the processing of genomic data within a secure computing environment. Secure computing environments comprise a separate and isolated computing environment—both hardware and software—from a more general computing environment.[11] This can include, for example, separate computer chips, computer memory that can only be accessed by trusted chips, or software that requires certain chips to encrypt and decrypt data.[12] The architecture of such environments makes it extraordinarily difficult for data to 'leak out' into the general computing environment, where it runs the risk of being copied elsewhere or stolen.[13] Once a rare and high-tech enterprise, secure computing environments are now widely available on smartphones such as Apple's iPhone and a number of Android-based smartphones.[14] For genomic data, this presents the possibility of sequestering access to the underlying genomic data through a 'host application' running within the secured computing environment, then bringing any third-party interpretive services to the data rather than the other way around, as is practised now. This approach heralds the possibility of making full use of the third-party genomic analysis ecosystem with much fewer risks than current genomic business models present.

Whether such an approach satisfies international privacy laws, though, requires detailed analysis. In the European Union (EU), the General Data Protection Regulation (GDPR) generally divides rights and responsibilities among data 'subjects', 'controllers', and 'processors'.[15] The processing of genomic data within a secure computing environment, as described here, would likely render the host application a 'controller' under Article 24 of the GDPR, thereby requiring certain security measures such as 'data minimization', the limiting of data available to be processed.[16] At the same time, the nature of secure computing environments—that is, the sequestering of data away from the general computing environment—appears largely to satisfy the GDPR's strictures on data processors.[17] This makes the approach described here likely compliant with the GDPR.

US privacy law, in contrast, is badly patchwork, with little in the way of generally applicable national regulations governing the computation of genomic data. S*ubnational* laws, however—namely those particular to US states—have recently filled this void. A number of states have now enacted general data privacy laws, some of which parallel the GDPR. Some of these include specific provisions pertaining to genetic data. And yet other states, even without a general data privacy statute, have laws that apply to the processing and transfer of genomic information. In these cases, as with the GDPR, it appears that the processing of genomic data within a secure computing environment satisfies their legal strictures.

While this analysis is specific to genomic data and the technology described here, it does suggest some broader lessons about processing sensitive data within secure computing environments. First, despite the differences in international data privacy regimes, secure computing environments may be a satisfactory *technical* solution to data privacy's *jurisdictional* problem: secure computing environments satisfy many of the technical

7 Robin Z Hayeems and others, 'Clinical Utility of Genomic Sequencing: A Measurement Toolkit' (2020) 5 NPJ Genomic Med 56.

8 Julianne M O'Daniel and others, 'A Survey of Current Practices for Genomic Sequencing Test Interpretation and Reporting Processes in US Laboratories' (2017) 19 Genetics Med 575.

9 Mack DeGeurin, 'Hackers Got Nearly 7 million People's Data from 23andMe. The Firm Blamed Users in "Very Dumb" Move' *The Guardian* (London, 15 February 2024) <https://www.theguardian.com/technology/2024/feb/15/23andme-hack-data-genetic-data-selling-response> [https://perma.cc/RRY9-W5EN]; Sara Gerke and others, 'Bankruptcy, Genetic Information, and Privacy–Selling Personal Information' (2025) 392 New England J Med 937.

10 Ronald Pulivarti and others, 'Cybersecurity of Genomic Data' (National Institute of Standards and Technology (US) 20 December 2023) <https://nvlpubs.nist.gov/nistpubs/ir/2023/NIST.IR.8432.pdf> [https://perma.cc/T3UD-BJWB].

11 André Brandão, João S. Resende and Rolando Martins, 'Hardening Cryptographic Operations through the Use of Secure Enclaves' (2021) 108 Computers & Security 102327.

12 ibid.

13 ibid.

14 'Secure Enclave' (n 4); 'Trustee TEE' (n 4).

15 GDPR (n 2), arts 12, 24, 28.

16 ibid art 25.

17 ibid art 28(3)(e), (f); Lauren Biernacki and others, 'Sequestered Encryption: A Hardware Technique for Comprehensive Data Privacy' (2022) 2022 IEEE Int'l Symp Secure and Private Execution Environment Design 73.

requirements imposed upon data controllers and processors, including data minimization, pseudonymization, and rights of access. Secondly, social trust remains paramount to adopting secure computing environment technology for processing sensitive data. Users must trust the security of the computing environment and the limitations of the host application for adoption to be widespread enough that the technology can meet its potential. Thirdly, the technology described here suggests the importance of a zone of control around sensitive data to take advantage of the legal benefits of the technology. Compliance is achieved here precisely because the host application technically controls what data, if any, processors can take with them outside the secure computing environment.

The 'Privacy and the sharing of genomic data' section of this article explains the legal and technical challenges arising from the sharing of genomic data and describes a new sociotechnical approach—namely, the processing of genomic data within a secure computing environment. The 'Sharing computationally secure genomic data in the EU and USA' section analyses the legal implications of the technology under privacy regimes in the EU and the USA, including a comparison of the two jurisdictions. Th 'Lessons for sharing sensitive data through secure computing environments' section explores some broader lessons about international data privacy law, given the use case of the technology described here. The article concludes with suggestions for future work in the area.

Privacy and the sharing of genomic data

The usual approach to sharing genomic data

The human genome is the sum total of DNA in a human cell, the code for what genetically constitutes each individual.[18] The genome can be decoded by sequencing its underlying DNA, arranging the order of its chemical bases—adenosine (A), cytosine (C), thymine (T), and guanine (G)—that make up a strand of DNA's double helix.[19] There are a variety of reasons to sequence human genomes and obtain genomic data. Most saliently, clinical genomic sequencing seeks to diagnose diseases or to understand the genetic basis for disease aetiology.[20] But human genomic data can also be used to conduct scientific research; investigate crimes; determine an individual's family relations, ethnic ancestry, or geographic area of origin; to satisfy a philosophical 'right to know'; and, for artistic purposes—such as art projects using visualizations of genomic sequence data—or purely for entertainment.[21]

The means to obtain genomic data have recently and significantly expanded. Clinical genomic sequencing services are now readily available at a large number of hospitals and healthcare providers.[22] There are also direct-to-consumer genomic sequencing services, such as Nebula Genomics and Dante Labs, and DNA services laboratories—high-throughput sequencing facilities that will usually perform sequencing on any piece of DNA sent to them.[23] This does not include third-party interpretation services, companies that analyse genomic sequence data, from charting complicated bioinformatics comparisons to drawing simple family trees.[24]

Despite the variety of genomic analysis services, all of them, at a basic level, operate in the same fashion. An individual provides a genomic sample, typically from their saliva or blood. The sample is then minimally prepared and sent to a sequencing facility. The facility then processes the sample on a genetic sequencing machine that produces a data file containing the sequence information and, often, some additional information about accuracy (often synonymously referred to as 'quality').[25] These data are usually stored in a series of files in the BAM file format.[26] Because the human genome contains a string of six billion chemical base pairs—and because modern genetic sequencing machines often read each area of DNA in the genome dozens of times—these files can be quite large, in some cases over 100 GB per subject.[27] The sequencing service then delivers these files to the individual via physical storage, for example flash drive or hard disk, or, more frequently, an encrypted cloud storage platform, such as Genomics on Amazon Web Services, for download.[28]

Raw sequence data alone are practically useless. The goal of sequencing, instead, is interpretation: what the sequence data *means* for some particular use.[29]

18 International Human Genome Sequencing Consortium, 'Initial Sequencing and Analysis of the Human Genome' (2001) 409 Nature 860, 860.

19 ibid.

20 John E Gorzynski and others, 'Ultrarapid Nanopore Genome Sequencing in a Critical Care Setting' (2022) 386 New England J Med 700, 700.

21 Sherkow (n 6) 1051–52.

22 Julianne M. O'Daniel and others, 'A Survey of Current Practices for Genomic Sequencing Test Interpretation and Reporting Processes in US Laboratories' (2017) 19 Genetics in Med 575, 575

23 Sherkow (n 6) 1058–60.

24 Christi J. Guerrini and others, 'Who's on Third? Regulation of Third-Party Genetic Interpretation Services' (2020) 22 Genetics in Med 4, 4.

25 Somak Roy and others, 'Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists' (2018) 20 J Molecular Diagnostics 4, 5.

26 ibid 6.

27 Ben Langmead and Abhinav Nellore, 'Cloud Computing for Genomic Data Analysis and Collaboration' (2018) 19 Nature Rev Genetics 208.

28 ibid 210.

29 Clayton and others (n 1) 3.

Individuals who wish to use the interpretive genomic services described above must consequently share their underlying genomic data, that is, the DNA sequence, with the interpretive service.[30] This can be trivial, for example, where the sequencing service and interpretive service are one and the same, such as some of the interpretive services provided by Illumina, an industrial DNA sequencing giant.[31] At other times, however, sharing genomic data can be a complex undertaking, requiring individuals to transfer their genomic data using insecure means or for individuals to consent to the interpretive service having far more access to the genomic data than desired or even necessary.[32] Users wishing to conduct a robust ancestry analysis, for example, must often disclose a significant amount of sensitive genomic information to the interpretive service, even though the end information sought is far less sensitive.[33] This is exacerbated where an interpretive service needs additional sensitive information to validate its results or where the individual, the location of the genomic data, and the interpretive service span several jurisdictions.[34]

Privacy concerns about the usual approach

The usual approach to genomic sequencing places a step of removal between the sequencing subject, that is the patient or consumer, and the holder of the resulting genomic data. In the health-care context, for example, genomic data are often—and often, exclusively—turned over to the health-care provider.[35] In other cases, the data—while being made available to their respective data subjects—are held in cloud storage controlled by another entity.[36] Historically, this separation between a subject and their genomic data had some purchase; individuals do not typically have enough electronic storage to keep such vast amounts of data or the technical expertise to securely analyse it.[37] But this separation does suggest several privacy concerns.

The most immediate concern arises when the interpretation service is also the genomic sequencer. To the extent that data subjects wish to make use of only a subset of interpretation services, for exampl to determine their propensity for a narrowly drawn genetic illness, the user cannot practically do so without disclosing *all* of their genomic information to the same entity.[38] Similarly, a data subject seeking to engage in interpretation services for a non-sensitive application, such as determining familial geographic origin, must often also disclose sensitive information—such as health information—with their request.[39]

In many instances, DNA sequencers are aware of these difficulties and have established privacy policies that either limit data subjects' access to their own data or promptly delete it upon request.[40] But these policies suffer from the risk of overpromise. Sequencers may not follow through on their own recitals due to accident, negligence, deceit, or outright fraud. In 2023, for example, the US Federal Trade Commission (FTC) settled a complaint against Vitagene, Inc., charging that the company deceptively retained subjects' genomic data even after requests to delete them.[41] Relatedly, because genomic sequences are shared among consanguineous family members, there is a risk of inferring sensitive genomic information among users with differing privacy requests.[42]

But even where interpretive services do follow through on their promises of limiting access to subjects' genomic data, computing security risks abound. Using a single or small number of centralized electronic storage systems for genomic data makes such systems a prime target for hackers. In 2023, for example, hackers gained access to seven million users' genetic data from 23andMe, a combination sequencer and interpretive service.[43] The fallout of this incident ultimately led to the company's bankruptcy—and the sale of subjects' underlying data to another entity.[44] The cybersecurity of genomic data remains an ongoing and complex problem.[45]

On the data subjects' side, there is also the risk of inadvertent sharing, even where data subjects are in possession or full control of their genomic data. Users

30 Guerrini and others (n 24) 4.

31 ibid.

32 Clayton and others (n 1) 6–7.

33 ibid 16–18.

34 Fruzsina Molnár-Gábor and others, 'Bridging the European Data Sharing Divide in Genomic Science' (2022) 24 J Med Internet Res e37236.

35 Roy and others (n 25) 6.

36 Langmead and Nellore (n 27) 208.

37 ibid.

38 ibid 217.

39 US National Institute of Standards and Technology, *Cybersecurity of Genomic Data* (NIST Internal Report No. 8432, Dec 2023) 6–7 <https://s3.documentcloud.org/documents/24234732/nistir8432.pdf> [https://perma.cc/L9RA-EYPX].

40 James W Hazel and Christopher Slobogin, 'Who Knows What, and When?: A Survey of the Privacy Policies Proffered by U.S. Direct-to-Consumer Genetic Testing Companies' (2018) 28 Cornell JL and Public Policy 35, 49–52.

41 Decision and Order, *In re 1Health.io Inc d/b/a Vitagene, Inc* No. C-4798 (6 Sept 2023).

42 Natalie Ram, 'DNA by the Entirety' (2015) 115 Columbia L Rev 873, 899–903.

43 Mack DeGeurin, 'Hackers Got Nearly 7 million People's Data from 23andMe. The Firm Blamed Users in "Very Dumb" Move' *The Guardian* (London, 15 February 2024) <https://www.theguardian.com/technology/2024/feb/15/23andme-hack-data-genetic-data-selling-response> [https://perma.cc/RRY9-W5EN].

44 Gerke and others (n 9) 937.

45 *Cybersecurity of Genomic Data* (n 39) 9–10.

delivering their genomic data to third-party interpretive services may divulge more information than they intended, or fail to appreciate the scope of the third-party's use of their information.[46] Third parties, in turn, may share that information with other services in order to improve their analyses—something users may be unaware of, even when detailed in an interpretive service's privacy agreement.[47] This, too, was a cause of complaint by the FTC against Vitagene,[48] and speaks to a broader concern about separating data subjects from control of their data.

Secure computing platforms: a new approach

Secure computing environments

Our project has developed a new approach to sharing genomic data—what we deem a 'sociotechnical' approach—by relying on the ubiquity, trust, and recent development of secure computing environments. In the usual genomic data-sharing approaches outlined above, interpretive services necessarily access users' underlying genomic data to provide any requested analyses.[49] In our system, by contrast, access to subjects' genomic data are controlled by a host application running within a secure computing environment. Any interpretive analyses, therefore, must be run entirely within that environment. In other words, whereas previous systems required bringing genomic data to the computation, our approach brings the computation to the data.

Generally, secure computing platforms house executable software within a confined computing area and do not—without express permission from the system's user—allow data to escape.[50] Any data or computation results within the environment can further be encrypted to ensure that a 'data leak' is not interpretable outside the secure environment.[51] In this way, data analyses can be performed without the risk that the analyser programme will share the underlying data outside of the secure environment.

Until recently, using secure computing environments for genomic data was impractical. But advances in both genomics and secure computing availability have now made such operations feasible. As detailed above, sequencing has become easier, from direct-to-consumer whole-genomic sequencing services to run-of-the-mill DNA service laboratories.[52] True: the original sequencers may have a copy of subjects' underlying genomic data. But many DNA services laboratories do, in fact, delete users' data shortly after producing it.[53] And the large size of whole-genomic sequence datasets, specifically, make indefinitely housing such data quite costly.[54] It is now possible—and relatively inexpensive—to obtain and control whole-genomic sequence data specific to an individual data subject.[55]

Secondly, secure computing environments—once the province of only cutting-edge computer science research—are now widely commercially available. Many smartphones, for example, house secure computing environments. Apple's Secure Enclave, for example, is 'a dedicated secure subsystem integrated into Apple systems on chip (SoCs) … [and] isolated from the main processor to provide an extra layer of security'.[56] Secure Enclave 'is designed to keep sensitive user data secure even when the Application Processor kernel becomes compromised'.[57] The technology is available on virtually all iPhones today. Similarly, Android's Trusted Execution Environment (TEE) is a secure computing environment 'isolated from the rest of the system by both hardware and software', present on all Android smartphones released or updated since 2017.[58]

The host application use case

Our approach begins with a user uploading their genomic data to a smartphone with a secure computing environment, such as recent models of Apple or Android smartphones. A 'host application' running within the secure computing environment then controls all access to the data through a combination of encryption technologies (eg salted Merkle trees) and access permissions. This has a dual purpose: it prevents other applications from gaining access to the genomic data outside the secure computing environment, and prevents any data analysed within the secure computing environment from leaving.

A user seeking to analyse their genomic data could then download a third-party application for the purpose. For example, if a user wanted to determine their propensity for a particular genetic disorder, they would

46 Hazel and Slobogin (n 40) 54–6.
47 ibid 55.
48 *1Health* (n 41).
49 Langmead and Nellore (n 27) 208.
50 Brandão, Resende and Martins (n 11) 2.
51 ibid.
52 Sherkow (n 6) 1060.
53 Hazel and Slobogin (n 40) 50.
54 Langmead and Nellore (n 27) 216–217.
55 ibid.
56 'Secure Enclave' (Apple Platform Security, 7 May 2024) <https://support.apple.com/guide/security/secure-enclave-sec59b0b31ff/web> [https://perma.cc/N4UU-VAMA].
57 ibid.
58 'Trustee TEE' (*Android Open Source Project*, 29 April 2024) <https://source.android.com/docs/security/features/trusty> [https://perma.cc/D5AX-QYH5].

download a third-party application to do so. The host application would then allow the third-party application to run within the secure computing environment, where the third-party application would then have access to the genomic data. Importantly, any results would then be returned to the host application, still running within the secure computing environment, where the user could view them. Once the third-party application's operations are complete, it brings neither the results of its analysis nor any underlying genomic data outside the secure computing environment, preventing the leakage of any genomic data. As an added layer of security, users could also make use of application verification stores—such as Apple's App Store—to ensure that any third-party applications have been vetted to perform within the secure environment.[59]

Additionally, the results of any analysis need not be the genomic data itself, ie a user's underlying genomic sequence. For example, a third-party application designed to ascertain a user's susceptibility to a certain genetic disease could simply return results such as 'High Risk' or 'Low Risk' without disclosing any underlying genomic information. The technology also allows users to limit access to portions of the data by the third-party application, such as certain genes or areas of the genome, or to refuse information about the user's personal characteristics or the provenance of the sample. This further limits the risk of data leaks and is akin to the concept, familiar in computer science, of 'zero-knowledge proofs': ways of disclosing the receipt or use of information without disclosing the underlying information itself.[60]

Secure computing environments as pro-privacy

This approach solves several privacy concerns surrounding genomic data-sharing. Significantly, it prevents the problem of interpretive services permanently possessing users' genomic data or analysis results, whether by design or accident. By sequestering access to genomic data from within a secure computing environment, interpretive services, even if temporarily inside, cannot remove them. Relatedly, users need not worry about disclosing the entirety of their genomic information simply to perform interpretive analyses on a small part. This limits interpretive services from performing unwanted inference analyses, such as linking users to genetically related families. The disclosure of genomic data by an analysis service to a third party, as in the *Vitagene* case, is technically impossible.

Making use of secure computing environments also prevents accidental disclosures. By separating interpretive services from the underlying data, interpretive services cannot accidentally disclose their contents by leak or neglect; they never come into possession of the data in the first instance. For the same reason, it also makes the interpretive services themselves poor targets for hackers wishing to steal genomic information, because there is no single computing system housing such data. Genomic data theft, as in the recent example of 23andMe, is wholly impractical.

And the system also seems to act as a bulwark against outright fraud. An interpretive service, even purposefully attempting to remove genomic data from the secure environment to their own systems, would need to defeat both the host application's permissions and the environment's secured layer—an extremely difficult task we think implausible. While attempts to defeat security layers in secure computing environments have been reported as possible, they are costly, laborious, and difficult—and, with some recent advances in hardware, increasingly difficult to do.[61]

Such risks are further mitigated by smartphone platforms' verification of apps on app stores. The verification process typically ensures that any apps made available through branded app stores hew to certain software and developer policies, many of which ensure that apps isolated to a secure computing environment stay there.

Sociotechnical trust

We refer to this approach as a 'sociotechnical' one because it relies on both technical advances in secure computing and, significantly, social trust. Social trust—in the context of privacy—generally refers to 'a resource of social capital between or among two or more persons concerning the expectations that others will behave according to accepted norms'.[62] For data privacy, those norms include adhering to promises of security or encryption—that data promised to be secured will, in fact, be secured, and that data promised to be encrypted will, in fact, remain so.

Social trust is an important—arguably the *most* important—facet of any data platform.[63] Despite the technical safeguards of secure computing environments,

59 Apple, 'Apple Platform Security' (May 2024) 118, <https://help.apple.com/pdf/security/en_US/apple-platform-security-guide.pdf> [https://perma.cc/4PJ2-JDZA].

60 Mahimna Kelkar and others, 'Complete Knowledge: Preventing Encumbrance of Cryptographic Secrets' (2023) Cryptology ePrint Archive (Paper 2023/044).

61 Brandão, Resende and Martins (n 11) 3–4.

62 Ari Ezra Waldman, *Privacy as Trust* (Cambridge UP 2018) 51.

63 ibid 61

users who do not trust the safety of the system will not use it. And without a significant user base, DNA interpretation services are unlikely to develop applications for the system outlined above. The alternative—where DNA interpretation services build apps to be used outside of a secure computing environment—is effectively the system currently in place, with data subjects dumping genomic data online en masse, and interpretive services taking advantage of them.[64]

The privacy flaws in the current system are problematic for data subjects. But they are also problematic at a broader, societal-level scale as the amount of freely available, unrestricted genomic data increases. This includes the risks attendant to identifying individuals by familial inference—that is, effectively deducing an individual from pseudonymous genomic data based on identifiable close genetic relatives.[65] Famously, the presence of genomic data in ancestry databases has been used by police in the USA to identify criminal suspects from genomic family profiles.[66] Beyond law enforcement, the availability of DNA has also been used to tie specific individuals to anonymous genomic DNA samples where that individual's relative has deposited genomic information online. In a highly cited paper, researchers were able to uniquely identify several individuals from anonymous genomic samples if one of their close relatives deposited genomic data online.[67] By 2018, it was estimated that over 90 per cent of Americans of European descent were identifiable.[68] That number, as of 2025, has certainly increased. Our project seeks to ameliorate these societal privacy flaws by making it easier—and no less convenient—to sequester genomic data from prying eyes.

To that end, part of this grant-funded project is to investigate users' willingness to adopt our system. Aside from developing genomic data-sharing software, we are also measuring the potential for user adoption through a combination of app testing and follow-up surveys. These include exploring which types of interpretive services users would be willing to take advantage of; whether those differ from currently available analyses; whether users are interested in sharing the results of the genomic testing without sharing the underlying data; and whether users will make use of disclosing to an interpretive service some portions of their genome but not others. Such uses are features of our technology, not available under current genomic data-sharing regimes.

Sharing computationally secure genomic data in the EU and USA

Technical safeguards and social trust for secure genomic data-sharing are still only half of the equation. The development and adoption of secure genomic data-sharing platforms also rely on law. Laws that, for example, place significant legal burdens on third-party DNA interpretive services—even if they cannot, by design, report any underlying sequence data—are likely to hamper their development. Similarly, laws that 'propertise' genomic data more generally may yield significant legal risks to genomic data-sharing software, even if such software operates in secure computing environments.[69] A functioning, secure, genomic data-sharing ecosystem requires appropriate levels of legal protection and liability.

This is yet even more complicated because genomic data-sharing is, of course, an international phenomenon.[70] At its core, genomic data are simply digital information that—like other digital information—can hardly be restricted at national borders.[71] Human genomic sequencing occurs in virtually every country, and there are few barriers for individuals of one nationality engaging in sequencing in another jurisdiction.[72] Genomic data research, especially, tends to be a global affair.[73] And in the EU in particular, human genomic samples are very likely to cross Member States' borders.[74]

The internationalization of genomic sequencing touches on significant legal pressure points for data collectors and analysers.[75] Understanding how these operate is necessary to understand whether particular policies or best practices for secure genomic data-sharing should be adopted, and whether there are opportunities (or pitfalls) for cross-border arbitrage between or within the EU and the USA.

Here, we look at EU and US law because the two jurisdictions are responsible for a substantial portion of

64 Guerrini and others (n 24) 4.

65 Ram, 'DNA by the Entirety' (n 42) 899–903.

66 Natalie Ram, 'Genetic Privacy After *Carpenter*' (2019) 105 Virginia L Rev 1357, 1359–63.

67 Yaniv Erlich, Tal Shor, Itsik Pe'er and Shai Carmi, 'Identity Inference of Genomic Data Using Long-Range Familial Searches' (2018) 362 Science 690, 690.

68 ibid.

69 Jessica L Roberts, 'Progressive Genetic Ownership' (2018) 93 Notre Dame L Rev 1105, 1130–1131.

70 Kärt Pormeister, 'Genetic Research and Applicable Law: The Intra-EU Conflict of Laws as a Regulatory Challenge to Cross-Border Genetic Research' (2018) 5 J L & Biosci 706, 720.

71 Irma Klünker and Heiko Richter, 'Digital Sequence Information between Benefit-Sharing and Open Data' (2022) 9 J L & Biosci lsac035, 5–8.

72 OECD, 'Building and Sustaining Collaborative Platforms in Genomics and Biobanks for Health Innovation', OECD Science, Technology and Industry Policy Papers No. 102 (March 2021) 8–10.

73 ibid 11–12.

74 Pormeister (n 70) 720.

75 Clayton and others (n 1) 2.

human genomic sequencing today and have a deep saturation of smartphone usage.[76] Of additional interest, the EU and the USA have stark differences in their legal approaches to genomic data privacy. In brief, the EU operates under the ever-present GDPR, which sharply regulates the personal data of EU citizens.[77] In contrast, the USA has virtually no national law that compels any particular approach to secure genomic data-sharing.[78] But a great number of US states, at the subnational level, have promulgated broad privacy laws over just the past 2 years. This part analyses these regimes and concludes with some comparative thoughts about the EU and US systems.

EU law

The GDPR and EU Regulation 2017/746 (*in vitro* devices)

The EU has a uniform, comprehensive, and widely applicable privacy regime, the GDPR.[79] At its most general level, the GDPR places restrictions on the collection, control, and processing of all 'personal data', defined as 'any information relating to an identified or identifiable natural person'.[80] This specifically includes 'genetic data', for which some special protections are given.[81] These provisions, and indeed the whole of the GDPR, have a broad territorial scope: they apply to the data of any data subjects who are under the jurisdiction of an EU Member State, whether the processing of their data occurs there or not.[82]

In crafting its protections, the GDPR broadly categorizes the entities that encounter personal data into three groups: the 'data subject', the 'data controller', and the 'data processor'. The data subject is the natural person to whom the personal data relate—its owner, in simpler terms—and is afforded a variety of rights regarding the use of their data.[83] The 'data controller' is the entity that 'determines the purposes and means of the processing of personal data'.[84] And the data processor is an entity that 'processes personal data on behalf of the controller'.[85] This division is meant to ensure ownership of personal data to the data subject, as well as focus liability—if something goes wrong—on the data controller.[86]

Here, the user of our system should be construed as the 'data subject', since they would be the natural person to whom the genomic data relates. To be sure, we could envision circumstances in which that is not the case—for example, where a data subject transfers their genomic data to another trusted individual for analysis. But given the personal nature of smartphones, we think these circumstances would be rare; likely analysed similarly given the data subject's consent under Article 6(1)(b); and, in any case, beyond the scope of this article.

The host application, meanwhile, should be construed as the controller; it is responsible for 'implement[ing] appropriate technical and organisational measures to ensure and to be able to demonstrate that processing is performed in accordance' with the GDPR.[87] By default, these technical measures, captured in a controller's 'data protection policy,' should strive for 'data minimization,' that is, allowing the processing of 'only [the] personal data which are necessary for each specific purpose of the processing are processed'.[88] The technical measures should also prevent the transmission of personal data 'without the individual's intervention[,] to an indefinite number of natural persons'.[89] And for whatever processing is performed, the controller is obligated to keep sufficient records of their occurrences.[90]

Accordingly, construing the host application as a data controller, the system outlined here likely meets, and probably exceeds, these strictures. The technical measures available to secure computing platforms can readily ensure data minimization; as a matter of computing hardware, processors will have access to only the data available to them within the secure computing environment. The host application can also limit processors' access to specific portions of the genome, further minimizing the amount of data processed. In addition, because the results of processing cannot readily escape the secure computing environment, there is little risk that processors will transmit any personal data to 'an indefinite number of natural persons'.[91] This sequestering

76 IQVIA, 'Understanding the Global Landscape of Genomic Initiatives' (12 May 2020) <https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/understanding-the-global-landscape-of-genomics-initiatives.pdf> [https://perma.cc/WN2G-N2RH]; GSMA, 'The State of Mobile Internet Connectivity 2023' (Oct 2023) <https://www.gsma.com/r/wp-content/uploads/2023/10/The-State-of-Mobile-Internet-Connectivity-Report-2023.pdf?utm_source=website&utm_medium=button&utm_campaign=somic23> [https://perma.cc/S96D-S5DM].

77 GDPR (n 2), art 5.

78 Clayton and others (n 1) 6–7.

79 Regulation (EU) 2016/679 of the European Parliament and of the Council (n 2).

80 ibid art 4(1).

81 ibid 4(13), art 9.

82 ibid art 3(2).

83 ibid arts 12–22.

84 ibid art 4(7).

85 ibid art 4(8).

86 ibid art 5(2).

87 ibid art 24(1).

88 ibid art 25(2).

89 ibid.

90 ibid art 30(1).

91 ibid art 25(2).

of data is, perhaps, the principal design advantage of secure computing environments under the GDPR.

The GDPR's requirements for processors—here, the third-party interpretive services—are much more stringent. Processors must process only the personal data designated to them by the controller, ensure data confidentiality, implement technical security measures, and 'delete or return' processed data to the controller after processing, among other requirements.[92] Fortunately, the nature of the host application makes this relatively simple. Because the host application controls which parts of the genome the processor has access to, third-party interpretive services can only process designated genomic data. Data confidentiality, by extension, is baked into the technical component of the larger architecture because data cannot leave the secure environment. Nor does the processor 'see' or 'possess' the data outside of the secured environment, making deletion and return automatic. This means that strictures concerning pseudonymization, such as those in Article 6(4)(e), are unnecessary. As a result, the technical security measures from secure computing environments are, effectively, provided by the host application—the controller—fulfilling Article 28(3)(e) and (f) with respect to the processor assisting the controller in implementing security measures.

Beyond the GDPR, EU Regulation 2017/746 may also be applicable to some uses of the host application.[93] The Regulation generally pertains to *in vitro* diagnostic medical devices, defined as 'any medical device … intended by the manufacturer to be used *in vitro* for the examination of specimens, including blood and tissue donations, derived from the human body' for a broad variety of purposes.[94] Given the Regulation's recitals and other provisions, this seems to include genetic diagnostics by implication.[95] The Regulation requires all *in vitro* device marketers to conduct post-market performance follow-up reports (PMPF) to ensure their devices operate properly in real-world settings.[96] For genetic diagnostics, this appears to include documentation concerning scientific and clinical validity, that is, whether the test result accurately analyses the underlying genetic information and provides a clinically valid interpretation.[97] At the same time, other provisions of the Regulation—such as the need to pair some genetic tests with genetic counselling—only apply where the *in vitro* device is used within a healthcare setting.[98] This makes these restrictions likely inapplicable to situations where the data subject chooses which third-party applications will interact with the host application outside of the healthcare context. We think, therefore, that the host application system described above is robustly compliant under the GDPR and EU Regulation 2017/746.

Commentary and the potential for conflicts

This analysis is certainly not the first to assess novel genetic technologies under the GDPR; much has been written, for example, about the GDPR's role in policing genetic research. Edward S. Dove, for example, has written extensively on the role of individual rights in the processing of genomic data, and in particular, the deficiencies of broad consent for secondary uses.[99] Dara Hallinan, by contrast, has taken the opposite view on broad consent.[100] By extension, Mahsa Shabani and Pascal Borry published an influential analysis of how 'pseudonymized' genomic data processing for research purposes would be construed under the GDPR, noting the practice's deficiencies in that context.[101] These opposing concerns regarding broad consent and individual rights under the GDPR are, indeed, problematic, especially where any underlying genomic information is shared with the processor. But one advantage of the approach described here is that it has the capacity, if implemented correctly, to prevent unauthorized sharing and surreptitious use arising from broad consent previously identified in the literature.

In addition to the implementation concerns previously recognized, others have noted the potential for intra-EU legal conflict. This arises from gaps—intentional or not—in the GDPR concerning genetic research from previously collected samples. As explained by Kärt Pormeister:

92 ibid art 28.

93 Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU [2017] OJ L117.

94 ibid art 2(2).

95 ibid recitals 9, 10, art 4.

96 ibid art 10(3).

97 ibid art 56(3).

98 ibid art 4(2).

99 Regina Becker, Davit Chokoshvili, and Edward S Dove, 'Legal bases for effective secondary use of health and genetic data in the EU: time for new legislative solutions to better harmonize data for cross-border sharing?' (2024) 14 Intl Data Privacy L 223; Edward S Dove, 'Collection and protection of genomic data' in Sarah Gibbon and others (ed), *Routledge Handbook of Genomics, Health and Society* (2018); Edward S Dove, 'The EU general data protection regulation: implications for international scientific research in the digital era' (2018) 46 J L, Med & Ethics 1013–30.

100 Dara Hallinan, 'Broad Consent Under the GDPR: An Optimistic Perspective on a Bright Future' (2020) 16 Life Sci Society & Policy 1.

101 Mahsa Shabani and Pascal Borry, 'Rules for Processing Genetic Data for Research Purposes in View of the New EU General Data Protection Regulation' (2017) 26 Eur J Human Genetics 149.

> [G]enetic research might be subject to the Oviedo convention and its additional protocol concerning biomedical research, depending on whether these instruments have been ratified and implemented into national law by a given country. The Oviedo convention does establish a general rule for the secondary use of biosamples under Article 22. However, Article 22 only sets a minimum threshold of due notification, and leaves it for national laws to regulate the matter.[102]

This establishes a potential conflict of laws concerning the control and processing of genetic data in clinical trials, especially where genetic data had been previously collected from subjects and are sought to be reused without first notifying them. But again, the nature of genomic data processing in secure computing environments makes this all but impossible. Because users do not share the underlying genomic data with third-party processors—and must affirmatively consent each time such an analysis is performed—research derived from such platforms cannot 'reuse' data in this sense. Member States' implementations of the Oviedo Convention would nonetheless continue to be controlled by the GDPR's regulation of data controllers and processors.

US law

Unlike the EU, the USA does not have a comprehensive and nationally territorial privacy law regime; there is no omnibus 'privacy' or 'data protection' law akin to the GDPR. US law also spans a large and diverse federal system, comprising federal law that applies nationally in addition to the separate laws of 50 states, 14 territories, and one federal district. Some of these subnational jurisdictions do, in fact, have broad-based privacy laws, as discussed below. But the point is that data privacy in the USA is not a national monolith.

Federal law

US federal data privacy laws—to the extent they exist at all—are entirely sectoral. That is, they depend entirely on the *use* of the data and, to a lesser extent, where the data originate from. For example, 'health data', at the federal level, is less a discrete category of data than it is a claim about its purpose—the provisioning of health care. Without a single, overarching federal privacy statute, any federal act that controls the flow of information could, in some sense, be construed as a data privacy law. Given that such a categorization is too broad to be useful, only two federal statutes govern genomic data-sharing more or less explicitly: the Health Insurance Portability and Accountability Act (HIPAA) and the Genetic Information Nondiscrimination Act (GINA).

HIPAA generally prohibits the disclosure of 'protected health information' (PHI)—including genomic data—by a 'covered healthcare entity' without the consent of a patient, also termed an 'authorization'.[103] At the same time, myriad exceptions exist, contrary to HIPAA's focus on consent, including four permissible 'Privacy Practices', mainly concerned with payment and billing for healthcare services; and twelve 'Public Purpose' exceptions, as diverse as furthering cadaveric organ donation and improving the reach of law enforcement surveillance.[104]

First and foremost, with regard to computationally secure genomic data, it appears that data subjects would have previously authorized the sharing of their data with the host application and, by extension, the processing application, since the data subject would have chosen which processing application to use. A valid authorization, under HIPAA, must contain certain elements concerning the nature of the information sought to be disclosed and several recitations regarding data subjects' rights.[105] In the USA, these can easily be obtained via a 'click through' authorization screen, already in use for numerous mobile health apps. Even if data subjects do not separately authorize the sharing of their data with the processing app, this still likely meets HIPAA's authorization standard, given that the subject voluntarily made use of and directed the processing application to interact with the host application. More expansively—and this is a common misconception—nothing in HIPAA forbids, or grants a remedy to, a user who voluntarily shares their own PHI with another.

More formalistically, HIPAA is unlikely to apply in any case because none of the likely actors—the user, the secure data host, or the processing app—would constitute a 'covered healthcare entity'. Under HIPAA, those entities constitute healthcare providers, health plans, health clearinghouses, and many business associates of the above entities. For example, in the case where a user wishes to assess their genomic data for variants of the *ApoE* gene, a gene strongly predictive of Alzheimer's, and uses a processing app within the secured data environment to do so, neither the host application nor the processing app would constitute a covered healthcare entity.

102 Kärt Pormeister, 'Genetic Research and Applicable Law: The Intra-EU Conflict of Laws as a Regulatory Challenge to Cross-Border Genetic Research' (2018) 5 J L & Biosci 706, 709.

103 45 CFR s 164.508 (25 January 2013).

104 ibid s 164.512 (26 April 2024).

105 ibid s 164.512(b)(1)(i).

GINA, analogously, prohibits the use of 'genetic information' by employers or health insurers, including going so far as to prohibit their mere acquisition of genetic information in a variety of contexts.[106] And 'genetic information,' under GINA, has been defined broadly to include not just DNA sequence information but also 'genetic-test results, genetic-test results of family members, and family medical history'.[107] As a result, employers may not, for example, require, request, or purchase an employee's genetic information.[108] But, like HIPAA, an individual can, in some circumstances, waive these prohibitions by consent—namely, 'when an employee consents to the disclosure and the subsequent use of the information offers the employee some kind of benefit', such as a workplace wellness programme.[109] The upshot of this is that, at the federal level in the USA, if the secure computing platform is not used by a covered entity under HIPAA, or an employer or health insurer under GINA, there are no explicit restrictions on the platform's use.

State law

In contrast to US federal law, there is a robust variety of US state-level statutes regulating genomic data privacy. While there is not a modal approach, staste laws on genomic data privacy can generally be grouped into three categories: privacy laws specifically concerning genomic information; general privacy laws that pertain to or include genomic data under their protection; and an absence of any comprehensive consumer data privacy laws.

As of this writing—although the field is fast-moving—twenty-two states have enacted specific, standalone statutes pertaining to genetic data privacy.[110] Some, like those in Illinois and Massachusetts, parallel—and often, predate—the federal GINA statute; they principally focus on non-discrimination regulations pertaining to genetic information rather than data privacy or portability.[111] Others, such as those in Alaska or Florida, affirmatively make genetic information the legal property of the person from whom the data originated.[112] Yet others, like Washington's My Health My Data Act, impose regulatory restrictions and penalties on collectors of genetic data, absent consent.[113]

Other states, meanwhile, have adopted broad data privacy statutes, many of which are modelled after the GDPR and that include—expressly or implicitly—genomic data in their definitions of 'sensitive data'.[114] The State of Colorado, for example, has enacted one of the most restrictive laws concerning genetic privacy under its general privacy statute, the Colorado Privacy Act.[115] The Colorado Privacy Act effectively demands that any data controllers—using a definition adopted from the GDPR—grant data subjects several rights concerning their data, including a right to access, correction, deletion, and portability.[116] The Act applies to all controllers that market to Colorado residents and either control more than 25,000 persons' data or process data from more than 100,000 people per year.[117] This expressly includes 'genetic data'—*any* data generated by an analysis of a user's genome.[118] Connecticut,

106 42 USC s 2000ff–5 (21 May 2008); 78 Fed Reg 5659 (25 January 2013).

107 Bradley A Arehart and Jessica L Roberts, 'GINA, Big Data, and the Future of Employee Privacy' (2019) 128 Yale LJ 710, 732; 42 USC s 2000ff(4)(A) (2022).

108 42 U.S.C. s 2000ff-1(b) (2022).

109 Arehart and Roberts (n 107) 776.

110 Alabama Genetic Data Privacy Act, HB 21 (15 May 2024); Alaska Genetic Privacy Act, Alaska Stat s 18.13.020 (2024); Arizona Genetic Information Privacy Act, Ariz Rev Stat Ann s 44-8001 (2023); California Genetic Information Privacy Act, SB 41 (9 September 2021); Florida Protecting DNA Privacy Act, HB 833 (1 October 2021); Idaho Genetic Testing Privacy Act, Idaho Code s 39-8303 (2006); Illinois Genetic Information Privacy Act, 410 Ill Comp Stat 513/1 (1998); Kentucky Genetic Information Privacy Act, HB 502 (8 April 2022); Maryland Genetic Information Privacy Act, Md Code Ann, Comm s 14-4401 (2022); Massachusetts Genetic Privacy Act, Mass Gen Law ch 111, s 70G (2000); Montana Genetic Information Privacy Act, SB 351 (7 June 2023); Nebraska Genetic Information Privacy Act, LB 308 (17 July 2024); Nevada Consumer Health Data Privacy Law, SB 370 (15 June 2023); New Mexico Genetic Information Privacy Act, NM Stat Ann s 24-21-3 (1998); Oregon Genetic Privacy Law, Or Rev Stat s 192.535 (2023); South Carolina Privacy of Genetic Information Act, SC Code Ann s 38-93-10 (1998); South Dakota Genetic Data Privacy Law, SB 178 (21 Mar 2021); Tennessee Genetic Information Privacy Act, HB 1310 (1 May 2023); Utah Genetic Information Privacy Act, SB 227 (17 Mar 2021); Virginia Genetic Data Privacy Act, SB 1087 (1 July 2023); Washington My Health My Data Act, Wash Rev Code Ann s 19.373.005 (2023); Wyoming Genetic Data Privacy Law, HB 86 (8 Mar 2022).

111 Illinois Genetic Information Privacy Act, 410 Ill Comp Stat 513/1 (1998); Massachusetts Genetic Privacy Act, Mass Gen Law ch 111, s 70G (2000).

112 Alaska Genetic Privacy Act, Alaska Stat s 18.13.020 (2024); Florida Protecting DNA Privacy Act, HB 833 (1 October 2021); see also Jessica L. Roberts, 'Progressive Genetic Ownership' (2018) 93 Notre Dame L Rev 1105, 1128–1129.

113 Washington My Health My Data Act, Wash Rev Code Ann s 19.373.005 (2023).

114 Arkansas Personal Information Protection Act, Ark Code Ann s 4-110-101 (2005); Colorado Privacy Act, SB 21-190 (7 July 2021); Connecticut Data Privacy Act, SB 6 (10 May 2022); Delaware Personal Data Privacy Act, HB 154 (11 September 2023); Indiana Consumer Data Protection Act, SB 5 (1 May 2023); Iowa Data Privacy Law, SF 262 (28 March 2023); Kentucky Consumer Data Protection Act, HB 15 (4 April 2024); Minnesota Consumer Data Privacy Act, HF 4757 (24 May 2024); New Hampshire Privacy Act, SB 225 (6 Mar 2024); New Jersey Data Privacy Law, SB 332 (16 Jan 2024); Rhode Island Data Privacy Act, SB 2500 (28 June 2024); Texas Data Privacy and Security Act, Tex Bus & Comm Code Ann. s 541.001 (2023); Vermont Data Privacy Act, HB 121 (11 May 2024).

115 Colorado Privacy Act, SB 21-190 (7 July 2021).

116 ibid s 6-1-1306.

117 ibid s 6-1-1304.

118 ibid s 6-1-1303(24).

Delaware, Iowa, and Texas, among others, have taken similar approaches.[119] Meanwhile, some larger states, like New York, do not currently have either a general data privacy law or a genomic-specific privacy law on the books.

Despite this variance in subnational approaches, none, it seems, would prohibit the sharing of computationally secure genomic data in the manner described above. For those states taking a GDPR-style approach to genetic data, such as Colorado and Texas, the host application approach would likely satisfy those laws' requirements for the same reasons they would do so under the GDPR.[120] Non-GDPR-style states that nonetheless impose notice and consent restrictions on mere possessors of genomic data, like Washington, would likely similarly permit the technology because the data subject directs the controller, that is the host application, on when and how to process the subject's data.[121] Generally, putting users in control of their own genomic data—along with who gets to process it—satisfies many of the 'notice and consent' impositions of a wide variety of US state laws.

This applies with even greater force to those states that have effectively properties' genomic data, such as Alaska.[122] In those states, genomic data and the results of any genomic test belong to the person from whom the sample originated. For computationally secure genomic tests, because the underlying data are not made available to the third-party processor—nor has the ability to leave the secure computing environment—this would make it all but impossible for other entities to 'misappropriate' a user's genomic data. Again, securing genomic data through a host application and 'bringing the computation to the data' accords with such laws' conceptions of both property and licensing.

At the same time, because each state's genomic privacy law—for those that have them—is differently constrained as to whom or which data they cover, there are likely to be future complications regarding choice of law. This will be especially present where a user crosses state boundaries as a non-resident. Nonetheless, the technology described here seems to satisfy the minimum requirements for all states, at least today. For now, processing genomic data through secure computing environments seems to serve as a technological solution to the United States' wide-ranging federalist approach.

Jurisdictional arbitrage: a comparison

The technology described here may seem to present an opportunity for jurisdictional arbitrage: using the technology in one jurisdiction to evade the requirements of another.[123] But a closer comparison of EU and US regimes shows that such schemes are, perhaps counterintuitively, unnecessary because the technology is likely to be treated similarly across both jurisdictions.

First, the secure housing of genomic data within data subjects' control seems to be incredibly significant across both EU and US jurisdictions. In both instances, placing the user at the centre of the data controlling function seems to limit the liabilities typically imposed on data controllers.[124] This seems to fulfil the legal requirements imposed on controllers in both the EU and all 50 states.[125] Similarly, granting subjects effective control over their data in secure environment limits the possibility that processors will have access or use genomic data beyond the privileges granted to them by the controller—again, seemingly in conformity with both the GDPR and subnational data privacy laws across the USA.[126]

Secondly, processing data in a user-controlled, secure computing environment also appears to fulfil the differing purposes of statutes across both jurisdictions. For example, the sociotechnical approach to genomic data privacy described here seems to accord with the GDPR's recital of data protection as a 'fundamental right' (Recital 1) and control by natural persons of 'their own personal data' (Recital 7).[127] As explained above, no US data protection statutes frame their purposes similarly; they instead often focus on improving data 'portability', making data a form of legal property, or identifying users as 'consumers'.[128] Nonetheless, the approach outlined here fulfils these objectives just as well.

Indeed, the overlap between the EU and subnational US opportunities is so great that the opportunities for jurisdictional arbitrage appear to be quite narrow, if non-existent. The GDPR would practically apply to any

119 Connecticut Data Privacy Act, SB 6 (10 May 2022); Delaware Personal Data Privacy Act, HB 154 (11 Sept 2023); Iowa Data Privacy Law, SF 262 (28 Mar 2023); Texas Data Privacy and Security Act, Tex Bus & Comm Code Ann. s 541.001 (2023).

120 See nn 88–97.

121 Washington My Health My Data Act, Wash Rev Code Ann s 19.373.005 (2023).

122 Alaska Genetic Privacy Act, Alaska Stat s 18.13.020 (2024).

123 This is separate from the question about data *transfers* between jurisdictions, which is not addressed in this paper. Briefly though: because, as described, no underlying genomic data is transferred from the secured computing environment to outside of it, an analysis of cross-border data transfers is likely unnecessary.

124 See nn 92–95 and accompanying text.

125 ibid, n 132–134.

126 ibid.

127 GDPR (n 2) recital 1, 7.

128 Washington My Health My Data Act, Wash Rev Code Ann s 19.373.05 (2) ('portability'); Alaska Genetic Privacy Act, Alaska Stat s 18.13.010 ('property'); Texas Data Privacy and Security Act, Tex Bus & Comm Code Ann s 541.001(7) ('consumer').

host application offered anywhere in the USA, so long as any EU citizen would have the ability to download it within or outside the USA.[129] As is true with other smartphone applications, it is wholly impracticable to attempt to restrict use by EU citizens within the United States. As a corollary, the host application could not circumvent the GDPR by limiting genomic data services to US territory. Similarly, jurisdictional arbitrage would be no more effective were US citizens to have their genomic data processed within the EU to take advantage of the GDPR.

Lessons for sharing sensitive data through secure computing environments

Beyond the immediate application of genomic data platforms, there are broader lessons to be learned about using secure computing environments to process sensitive data and how they relate to international data privacy law. These concern the harmonization of data privacy laws after years of misalignment; the importance of social trust in a computing system; and the significance of the data subject's direction over the controller.

Legal harmonization through technology

Although we live in a digitally interconnected world, there has been a lot of lamentation that privacy laws are not harmonized across borders.[130] This has the effect of altering the user experience of certain technologies in different countries and making privacy compliance an expensive game, as international privacy compliance becomes costly and difficult.

But user-controlled secure computation appears to be one example where the underlying technology achieves the same results across two of the largest coordinated markets in the world today, achieving the same results as would formal compliance review, even for sensitive data, from Colorado to Croatia. As a result—and unlike Facebook or other forms of social media—the user experience should be the same across a wide variety of jurisdictions.

The approach here suggests one path forward—a technology-based solution—even where legal harmonization is not available. That is, one can find or develop *technological* solutions that broadly satisfy the privacy regimes of varying large technology markets. This is akin to international harmonization through private technological standardization, more commonly seen in the networking technology or hardware space, for example various Wi-Fi standards or USB-C.[131]

Additionally, the technology here suggests that such standardization can be achieved internationally, even where compliance is at its most fraught for one country's *subnational* laws. Here, a privacy technology that complies with both Alaska's and Texas's privacy laws smooths differences between the USA and the EU even in the absence of larger federal legislation. Going forward, and especially where national-level intervention is difficult, one seeking to adopt the principles of the GDPR in the USA should focus not on federal laws but on state ones. In the USA, the path towards compliance with Brussels runs not through Washington but Wisconsin.

The need of social trust for data privacy

The degree to which secure computing works depends not on law but on social trust. Social trust, in Ari Ezra Waldman's account, is inextricably tied to privacy and sharing with internet intermediaries; we only disclose sensitive information to computational intermediaries if we trust the purveyors of the technology and believe that the technology will guard against unauthorized disclosures.[132] Even where privacy law is robust and actionable—as with the GDPR—a technology will fail in the market if users are afraid to use it.

Despite secure computing environments' technical bulwarks against unauthorized disclosure, the trust quotient for the technology remains rather high. To get there, users must upload sensitive information to their phone and trust that the host application will prevent access to it. Users must also trust that the underlying hardware will work as described—namely, that it will safeguard the underlying information from third-party processors. Even in these cases— where a number of intermediaries are regulated by the GDPR and other privacy protections—users will simply not use the technology if any of these planks of trust are unsteady.

This facet of secure computing environments is further evidence of the importance of trust in technical solutions to privacy problems. Woodrow Hertzog has noted that '[t]he design of technologies sends us signals that shape our expectations of trust and risk calculations. Companies design technologies that will convince people to trust them."[133] Previous scholarship has also noted this about design, albeit more about user

129 GDPR (n 2) art 3.

130 W Gregory Voss, 'Obstacles to Transatlantic Harmonization of Data Privacy Law in Context' (2019) Univ Illinois JL Tech & Policy 405, 431.

131 Zachariah Davies and Arnaud Van Waeyenberge, 'Better Regulation by Standards? Harmonised Technical Standards, Transparency, and the Rule of Law' (2025) 62 Common Market L Rev 147, 148.

132 Waldman (n 62) 47–61.

133 Woodrow Hartzog, Privacy's Blueprint 99 (2018).

interfaces and less about hardware.[134] Yet secure computing environments take this a step further, oddly suggesting that trust is actually interdependent on technical solutions. That is, even with maximum amounts of user trust, and even in adopting best practices for design, users will nonetheless refuse to adopt the technology if they do not trust that the technology *works*—as exemplified by the public's reluctance to adopt self-driving cars.[135]

Securely sharing genomic data, therefore, likely requires public education on how the technology ultimately works—a cost that may hinder the development of the technology. This could be offset by increased publicity from smartphone manufacturers regarding their adoption of secure computing environments. Yet while the technical details of smartphones' secure computing environments are widely available to developers, they are not a point of significant advertising for manufacturers. Apple's Secure Computing Enclave, for example, has not been a part of the company's most recent iPhone advertising campaigns.[136] This contrasts with other security technical features, such as end-to-end encryption.[137] Social trust must be acquired for secure computing environments to flourish as a privacy-mediating device.

The importance of a secure computing zone of control

Processing genomic data in a secure computing environment also demonstrates the importance of situating data's 'zone of control' on a user's device. There is no concern, for example, that the data will be 'made accessible without the individual's intervention to an indefinite number of natural persons' because, in a secure, locally housed environment, the data are not shared with anyone other than the user.[138] This model further accords with those laws from US states allowing subjects to 'own' their genetic data as property because it makes it all but impossible for other data stewards to delete, transfer, or license user data without the user doing so themselves.[139]

By contrast, secure computing environments with data hosted 'in the cloud' run the risk of non-compliance. User authentication breaches may allow unauthorized individuals access to the underlying data, as the 23andMe example demonstrates. And even if the data are truly secured from intrusion, there is nonetheless the possibility that data could be accidentally destroyed—thus, violating some US state laws on ownership—or transferred to another entity pursuant to a corporate restructuring, such as a bankruptcy.[140] Furthermore, to the degree that data hosting encompasses users from around the globe, such involuntary data transfers would raise significant legal harmonization problems. What should a secure data host do if compelled by a US bankruptcy court to transfer data in a manner that nonetheless violates the GDPR?[141]

Nonetheless, there are limits to the benefits of secure computing models for data-sharing. The technology outlined here works because processors can deliver salient results without also returning the underlying data, for example, that as a user is at risk of developing a certain genetic illness without returning the underlying sequence itself. This makes sequestering data feasible without diminishing the utility of processing apps. But the system comes with risks where returned results must also include the underlying data to be evaluated. For example, a processing query on which specific gene variant a user has effectively asks the processor to return the data itself. While the safeguards in the system are likely to prevent the 'leakage' of data from the secure computing environment, this could not be guaranteed in the same way if the sequence data itself is not returned to the processor. Caution with processors—even in a secured environment—is still required.

Conclusion

The bankruptcy of 23andMe should remind us that the sharing of genomic data is a complex problem exacerbated by the fragmentation of international privacy laws. Further, the perceived lack of regulatory harmonization and data security may restrict the development of follow-on, consumer-focused genomic technologies. The recent development and widespread adoption of secure computing environments, however, may pave a smoother path towards the development of genomic data processing technologies. At their core, secure computing environments sequester—through both hardware and software—sensitive data from a larger computing environment, ensuring that certain data cannot 'leak out' without a user's consent.

134 ibid.

135 Maria Alonso, 'Driving Trust: Paving the Road for Autonomous Vehicles', Forum Institutional (10 Sept 2024) https://www.weforum.org/stories/2024/01/driving-trust-paving-the-road-for-autonomous-vehicles/.

136 Maria Alonso, 'Driving Trust: Paving the Road for Autonomous Vehicles', Forum Institutional (10 Sept 2024) <https://www.weforum.org/stories/2024/01/driving-trust-paving-the-road-for-autonomous-vehicles/> [https://perma.cc/4RCV-VVWY]

137 ibid.

138 ibid art 25(2).

139 See (n 125).

140 Gerke and others (n 9) 937.

141 ibid.

This technological advance has the added virtue of appearing to comply with a broad variety of international privacy protections—namely, EU privacy laws and privacy laws in the USA and individual US states. This suggests that the best solutions to some problems of international privacy law have not a legal but a technological solution, so long as users trust the underlying technology and are in complete control of the data source—a sociotechnical solution to data privacy. While harmonizing international privacy laws remains a worthy goal, the simplest path forward, in an age of political fragmentation, may be a technological one.

Funding

This work was funded, in part, by a National Human Genome Research Institute [grant no. R01HG012249-01].

Conflict of interest statement

None declared.

https://doi.org/10.1093/idpl/ipaf018

International Data Privacy Law, 2025, 00, 1–15
https://doi.org/10.1093/idpl/ipaf018
Article