

# Principles and Policy Recommendations for Comprehensive Genetic Data Governance

Vivek Ramanan\*, Ria Vinod\*, Cole Williams\*, Sohini Ramachandran†, Suresh Venkatasubramanian†

Brown University

vivek\_ramanan@brown.edu, ria\_vinod@brown.edu, cole\_williams@brown.edu, sramachandran@brown.edu, suresh\_venkatasubramanian@brown.edu

## Abstract

Genetic data collection has become ubiquitous, producing genetic information about health, ancestry, and social traits. However, unregulated use—especially amid evolving scientific understanding—poses serious privacy and discrimination risks. These risks are intensified by advancing AI, particularly multi-modal systems integrating genetic, clinical, behavioral, and environmental data. In this work, we organize the uses of genetic data along four distinct ‘pillars’, and develop a risk assessment framework that identifies key values any governance system must preserve. In doing so, we draw on current privacy scholarship concerning contextual integrity, data relationality, and the Belmont principle. We apply the framework to four real-world case studies and identify critical gaps in existing regulatory frameworks and specific threats to privacy and personal liberties, particularly through genetic discrimination. Finally, we offer three policy recommendations for genetic data governance that safeguard individual rights in today’s under-regulated ecosystem of large-scale genetic data collection and usage.

## Introduction

The scale of genetic data generation has increased at a staggering rate since 2001 (Loos 2020; Bycroft et al. 2018). A primary driver of this is decreased DNA sequencing costs (Wetterstrand 2023). For example, there are whole genome sequences combined with medical records from approximately 400,000 US residents in the National Institutes of Health (NIH) All of Us Research Study, with plans to increase sampling to include at least 1 million US residents (National Institutes of Health 2024). Direct-to-consumer (DTC) genetic testing companies have much more data: 23andMe has 14 million customer sequences (23andMe 2023a) and Ancestry has 25 million (Ancestry 2025). The increase in genetic data and improvements in technologies has resulted in the following aspirational “bold prediction” for 2030 by the NIH: “a person’s complete genome sequence along with informative annotations can be securely and readily accessible on their smartphone” (Gunter and Green 2023). From a scientific perspective, the collection

of genetic data and subsequent genetic information at this scale has been critically valuable to understanding health, ancestry, and driving public health initiatives.

The unregulated collection of genetic data has created new avenues for surveillance, exploitation, and systemic harm. Since 2020, the U.S. Department of Homeland Security has accelerated its collection of genetic data, growing the federal DNA database (CODIS) by more than 1.5 million people, the majority of whom are people of color (Glaberson, Tse, and Tucker 2024). 23andMe suffered a data breach after a hacker accessed customer profiles, primarily targeting individuals of Chinese and Ashkenazi Jewish descent (Carballo, Schmall, and Tumin 2024). Such pressing, real-world discrimination and privacy concerns are exacerbated by the sordid history of eugenics and forced sterilization in U.S. states in the 20th century (a model later adopted by Nazi Germany (Spiegel 2019)), and more recent practices of genetic determinism (Epstein 2003), racial pseudoscience (Duello et al. 2021; Lala and Feldman 2024; Carlson et al. 2022; Panofsky, Dasgupta, and Iturriaga 2021), race-motivated violence (Carlson et al. 2022), privacy and consent violations (Jillson 2024; Sterling 2011; Strand 2016), and broad scientific distrust (Saulsberry and Terry 2013; Parkman et al. 2015; Kaye 2012). These serious concerns must be addressed today given the number of recent proposals to use genetic data and information in public infrastructure—e.g., to assess intelligence proxies, like IQ (Plomin and von Stumm 2018); insurance premiums (Karls-son Linnér and Koellinger 2022); or educational outcomes (Harden et al. 2020). The only federal-level legislation in effect is the Genetic Information Non-Discrimination Act (GINA), which prohibits using genetic information in employment and health insurance decisions but not in long-term care, disability, or life insurance.

**Our Work.** In this paper, we make three contributions to construct a robust genetic data governance system. First, we expand upon Wan et al.’s (2022) Four Pillars of genetic data collectors and argue that inconsistent and *leaky* regulation across these pillars introduces opportunity for genetic discrimination and privacy violations. Second, we present a risk assessment framework for modern genetic data governance, derived from theoretical analysis and existing U.S. policy. Third, we make policy recommendations to (1) define genetic data and genetic information as a distinct protected cat-

\*These authors contributed equally

†These authors jointly supervised this work

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

egory of sensitive data, (2) extend the Genetic Information Nondiscrimination Act (GINA) to include this category, and (3) establish a unified regulatory framework to govern all collection of genetic data and applications of genetic information. We expect that our proposals for reformed genetic data governance can better protect the privacy and agency of individuals while still promoting scientific research in human genetics.

### **Background: How Health and Social Information Is Derived From Genetic Data**

An individual’s genome is over 3 billion DNA bases long and derived equally from both parents. Genetic variants—units of DNA that differ between individuals—are scattered throughout the genome and contain information about genetically influenced traits (from physical traits to behavioral traits). To predict traits or disease risk, geneticists analyze an individual’s unique set of genetic variants. However, this is a challenging task: while diseases like cystic fibrosis or Huntington’s disease are determined by a single genetic variant, most traits are influenced by many variants across the genome as well as non-genetic factors (complex traits). In this section, we make the connection between genetic data—the raw genomic material—and genetic information, which refers to health, ancestry, and social traits inferred through statistical analysis of that genetic data.

### **Genetic Data Is Inherently Relational**

The inheritance of DNA—where each parent passes down a random half of their genome to their offspring—results in relatives sharing portions of their DNA, with closer relatives sharing a larger proportion. Shared genetic material means that decisions made about any individual’s genome extend to their genetic relatives. At the population level, all individuals can be found in a latent web of interconnectedness. This means that a privacy breach in a small part of this web (that represents those in a genetic dataset, for example) may propagate to larger parts of the web and implicate individuals who were not in the original dataset. Thus, *relationality* applies to genetic data, where value is derived from shared connections among individuals (Viljoen 2024; Hargreaves 2017). This relationality complicates discussions of privacy in the context of genetic data (Costello 2022).

A powerful example of the relationality of genetic data was highlighted in 2018, when law enforcement leveraged it to identify the Golden State Killer. Authorities used DNA left at a crime scene to identify distant relatives through DNA matching on a public website, GEDMatch. By combining these matches with publicly available genealogical records, they identified the suspect despite the closest identifiable relatives sharing only a great-great-great-grandfather (Kaiser 2018; Zabel 2019). An individual’s genome is shared with tens of thousands of other individuals, but at the same time is uniquely theirs: even the genomes of identical twins contain differences (Jonsson et al. 2021). While a single genome in isolation may not always be immediately identifiable, the relational nature of genetic data can enable re-identification. Researchers

have argued that genomic data should be treated as “always identifiable in principle” due to this inherent interconnectedness (Bonomi, Huang, and Ohno-Machado 2020; Shabani and Marelli 2019).

### **Insight From Genetic Data Is a Function of the Cohort**

Trait prediction models require a reference genetic dataset, which is analogous to the training dataset in machine learning. Reference datasets contain genetic data and their “ground truth” annotations of trait values, whether it be ancestry, disease status, height, etc. Confounders like environmental exposures and social determinants of health can introduce bias in a genetic study. For example, the average participant in the UK Biobank tends to be more educated and healthier than the average Briton (Fry et al. 2017). The choice of the *cohort*—that is *who* is in the dataset and how their data is labeled—is thus a critical consideration of trait prediction models and their generalizability, i.e., their out-of-distribution (OOD) performance in individuals outside of the cohort. Prioritizing “cohort matching” in a genetic study ensures that the test sample closely aligns with the reference cohort in key factors such as ancestry, age, geography, and social or environmental influences. This alignment minimizes confounding and enables a more accurate interpretation of strictly genetic effects on traits.

### **Trait Prediction Models From Genetic Data**

Inference from genetic data requires knowledge of a genetic variant’s effect on a trait or disease. This can sometimes be tested in animals by controlled experiments that manipulate the animal’s genome. However, this approach is infeasible and unethical to pursue in humans, and so researchers instead rely on large genetic datasets to ask “Do individuals who have the variant *tend* to have the disease (or trait of interest) compared to those who do not have the variant?” To answer this question, researchers typically conduct genome-wide association studies (GWAS), which identify genetic variants that are associated with a trait and quantify the *effect* of each variant on the trait. The learned effect sizes of variants (also called *weights*) are directly interpretable and can be used for trait prediction. The polygenic score (PGS), considered the gold standard for trait prediction, is calculated as a weighted sum of genetic variants, with each variant scaled by its corresponding GWAS effect size. GWAS/PGS approaches have been widely adopted for trait prediction.

The deployment of AI technologies for genetic inference is nascent but growing: predictive models are increasingly deployed to enable precision medicine therapies (Feiszt and Kenny 2025). Recently, there have been calls to integrate genetic data with other information—rich clinical, behavioral, and environmental data using foundation models. Critically, biological data—particularly genomic sequences—is large-scale and shaped by noise, sparsity, and biased sampling across the diversity of life. In the domains of vision and language, integrating multi-modal data has shown success in improving predictive performance on sparse and noisy data domains. In the medical domain, Google’s Med-Gemini pro-

cesses electronic health records, imaging data (e.g., pathological and X-ray images), and genomic data (in the form of PGS) for trait prediction (Yang et al. 2024). They claim Med-Gemini-polygenic—their model for predicting traits with PGS—to be highly performant on OOD data, a highly valued result in AI models for biology. However, it is important to consider that genetic data is often correlated with and confounded by environmental or non-biological variables. Thus, while OOD traits may appear to be genetically correlated with in-distribution (ID) traits (implying shared biological pathways), these correlations—particularly for behavioral phenotypes—are often confounded by non-genetic factors that mimic biological signal, undermining their interpretability as true genetic correlations (Border et al. 2022). Examples of this phenomenon have been demonstrated in places like genetic risk testing for opioid use disorder, where environmental covariates confounded genetic data (Pozzi 2023; Hatoum et al. 2021; Davis et al. 2025).

### Blindspots for AI Applications Using Genetic Information

We believe that the increasing inclusion of genetic data into foundational models—while the interpretability of even simple linear models remains opaque—is of grave concern. GWAS are fraught with confounding: the effect sizes may be statistical artifacts due to correlations between genetic variants and environmental exposures; this is particularly true with behavioral traits. If variant-environmental correlations are not present in another population, the model will generalize poorly. Further, the non-zero effect size of the correlated variant will give the illusion of a genetic or biological basis for the trait, supporting ideas of genetic determinism. Statistical geneticists have developed study designs that control for almost *all* environmental confounding (Veller and Coop 2024; Howe et al. 2022; Nivard et al. 2022; Okbay et al. 2022), but these studies have not been widely adopted.

At the same time, researchers are using these *existing* studies—which suffer from confounding—to train AI models. We believe that the lack of consideration for this nuance in both training data curation and biological foundation model evaluations can exacerbate pre-existing issues in existing approaches. While AI can improve trait prediction or interpretation, it may do so by modeling nonlinear interactions among correlated variables and leveraging gene-environment correlations. Yet, there remains an outsized excitement to develop increasingly complex AI models with genetic data (Yang et al. 2024). We are particularly concerned about the use of multi-modal AI models incorporating genetic data to predict behavior, educational attainment, and other social outcome traits. PGS have already been developed for a variety of these traits, including educational attainment (Okbay et al. 2022), intelligence (Plomin and von Stumm 2018; Selzam S et al. 2019), standardized testing and mathematics scores (Harden et al. 2020; Smith-Woolley et al. 2018), substance use disorder (Barr et al. 2020), and aggression (Kretschmer et al. 2023). Med-Gemini-Polygenic was applied to health traits (i.e., not the aforementioned traits), and leveraged genetically correlated traits to perform well OOD (in this case, OOD refers to the

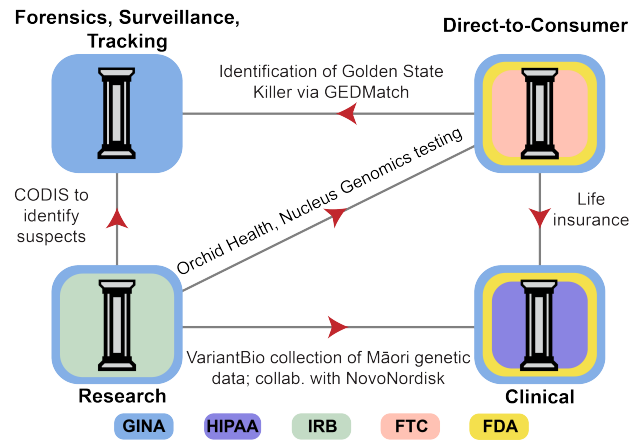


Figure 1: The genetic ecosystem consists of the Four Pillars of genetic data collection and regulatory considerations. Data collectors within each pillar operate with different collection strategies and intent to store, use, and transfer the data. However, there is only a patchwork of federal and state-level legislation that governs each Pillar. The resulting regulatory gaps lead to *leaky protections* through which genetic data or genetic information can be used in other pillars with little to no oversight. Real-world examples of leakage are indicated by red arrows. Leakage refers to both (1) genetic data itself and (2) genetic information, i.e., inferences derived from genetic data.

trait, not the individuals). Genetically correlated health traits may reflect shared biology, but for behavioral and social traits, correlations can be induced by non-biological factors (Border et al. 2022) and thus an AI model may not be “learning” any biological insight. If even simple genetic models struggle with confounding from factors like environmental exposures and cohort biases, complex multi-modal AI models should not be assumed to inherently distinguish true genetic effects from these confounded associations, and may even leverage them, potentially exacerbating existing issues and compounding uncertainties.

### Background: Current Protections

In this section, we outline the foundation of today’s genetic data collectors and the regulatory and legislative environments that they are subject to. To better understand the vast landscape over which genetic data is being collected, we build upon the Four Pillars (Wan et al. 2022) of genetic data collectors. We categorize these data collectors based on their motivations for collecting genetic data and the level of consent, control, and ownership afforded to the individuals from whom the data is collected. We define the updated Four Pillars as the following:

- **Forensics, Tracking, Surveillance:** Entities focused on monitoring, identifying, and analyzing genetic data to support public health outcomes, security, and law enforcement objectives.
- **Direct to Consumer (DTC):** Entities developing products to provide consumers with direct access to genetic

insights.

- **Clinical:** Entities requiring medical professionals' oversight to conduct tests, often enabling insurance coverage and generating clinically actionable insights.
- **Research:** Entities engaging in research and development (R&D) and basic research across private, public, and academic sectors. This includes public consortia, biobanks, and collaborations with DTC entities.

The genetic data ecosystem encompasses not only the entities within the Four Pillars but also the broader network of legislatures, regulators, and individuals involved in or affected by genetic data collection and use. Here, we highlight the regulatory gaps that allow genetic data and genetic information to leak across pillars. We provide real-world examples of this leakage in Figure 1, illustrating the *sociality problem* formalized by Viljoen to contextualize how relational data's social effects are categorized (Viljoen 2024).

## Federal Protections

**Statute: Genetic Information Non-Discrimination Act (GINA).** GINA was enacted in 2008 and prohibits genetic discrimination by employers or health insurers. Specifically, GINA bars employers from making hiring, firing, or promotion decisions based on genetic information (and information disclosed about relatives) and bans health insurance providers from denying coverage or charging higher premiums based on genetic information, including genetic predisposition for disease (Green, Lautenbach, and McGuire 2015). GINA does not apply to life, disability, or long-term care insurance (Gunter and Green 2023) or any other venues outside of employment and health insurance (e.g., education). The US Equal Employment Opportunity Commission (EEOC) regulates workplace discrimination for employees and thus, is responsible for enforcing GINA in the workplace. Before the passing of GINA, multiple cases of civil rights being violated in employment practices using genetic data have been recorded (Ajunwa 2016). Enforcement for health insurers falls under the purview of several federal agencies, including the Department of Health and Human Services (HHS); Department of Labor; Centers for Medicare and Medicaid Services; and the Department of the Treasury (Vermont Legislature 2019).

**Statute: Health Information Portability and Accountability Act (HIPAA).** HIPAA, signed into law in 1996, grants security and privacy protections for patient health information (PHI). HIPAA applies to covered entities that maintain this information, such as healthcare and health insurance companies, but does not apply to many other entities that might retain health information, including search engines, medical information sites, or dating sites (Citron 2020). Importantly, HIPAA does not apply to two of the largest DTC genetic testing companies 23andMe and AncestryDNA (Sklar 2020). HIPAA does not cover deidentified data: PHI data with unique identifying information (names, address, phone numbers, dates, etc.) removed do not fall under HIPAA. Research consortia, biobanks, and research collaborations with DTC and clinical entities primarily use

genetic data that is considered deidentified, and so HIPAA generally does not apply. However, the relational aspect of genetic data has motivated calls to “consider genomic data as, in principle, always identifiable” (Bonomi, Huang, and Ohno-Machado 2020). The inability to completely strip data of *all* identifiable features introduces critical vulnerabilities for genome owners and exacerbates the risk of data leakage across the Four Pillars.

**Agency: Food and Drug Administration (FDA).** The FDA “is responsible for protecting the public health by ensuring the safety, efficacy, and security of human and veterinary drugs, biological products, and medical devices” (Commissioner 2023). In the context of DTC genetic testing companies, the FDA only regulates well-defined high risk medical tests with clinical actions, such as 23andMe's health tests for diseases like Parkinson's, breast cancer, and late-onset Alzheimer's. If the FDA approves a specific test, Centers for Medicare and Medicaid Services then choose to approve the test for insurance coverage (Daval and Kesselheim 2023). Most genetic tests are not covered by Medicare, particularly DTC tests.

**Agency: Federal Trade Commission (FTC).** The FTC enforces federal laws that protect consumers from “fraud, deception and unfair business practices” (Commission 2024). The FTC has warned DTC genetic testing companies that the results they return to customers must be backed by “reliable science”, and also warns companies against making exaggerated claims about the use of AI in their products (Jillson 2024). Additionally, the FTC remains vigilant of deceptive practices regarding privacy policy changes and “dark patterns” designed to coerce consumers into consenting to data sharing. The power and broad scope of the FTC—and their stated intentions of cracking down on DTC genetic testing companies (Jillson 2024)—make it a major player in the regulatory space, and it will likely grow in importance if the federal government prioritizes genetic data privacy.

**Policy: The Common Rule.** This is a federal policy that protects human subjects in research settings (Sciences et al. 2014). It was first codified in 1981 by the Department of Health, Education, and Welfare (now HHS) but adopted more widely by other federal agencies in 1991. The Common Rule institutionalizes protections for vulnerable participants, protections overseen by Institutional Review Boards (IRBs); IRBs are groups “formally designated to review and monitor biomedical research involving human subjects” (Research 2019). Revisions enacted in 2017 focus on informed consent for participants, but do not cover deidentified data (Menikoff, Kaneshiro, and Pritchard 2017; Protections OHRP). IRBs ensure that consistent protection of human subjects are maintained and necessary for research to begin and proceed if involving human participants. The maintenance of the Common Rule and HIPAA, in the case of medical data, is performed by IRBs.

## State Protections

State laws offer far more protections than their federal counterparts, but a patchwork of these laws means a patchwork

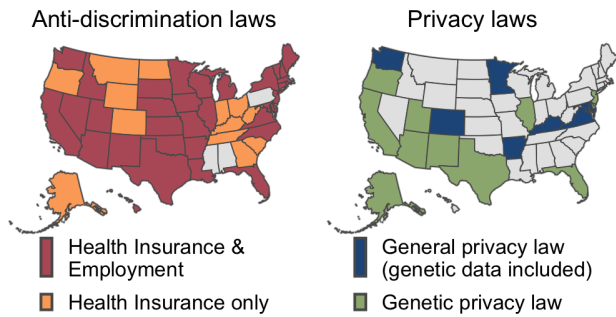


Figure 2: State-level genetic discrimination and privacy legislation varies widely beyond federal HIPAA and GINA protections. **Left:** States with genetic anti-discrimination laws covering both health insurance and employment (red) versus health insurance only (orange). **Right:** States with general privacy laws including genetic data (blue) and genetic-specific privacy laws (green). States with both are green.

in protections, as shown in Figure 2. Here, we focus on two categories of laws: genetic anti-discrimination laws and data privacy laws.

**Genetic Anti-Discrimination Laws.** Most states have passed some version of a genetic anti-discrimination law (initially prompted by the passage of GINA) although the sector, scope, and strength of the laws differs between states. Some states offer fewer protections than GINA: Georgia, for example, has a statute outlawing genetic discrimination by health insurers, but not employers. Other states offer more protections than GINA. California’s CalGINA (Legislature 2011) outlaws genetic discrimination in several sectors: employment, health insurance, housing, and by any state agency or entity receiving state funding (such as emergency medical services). Florida includes life, long-term care, and disability insurers (of Representatives 2020). Many states (including Massachusetts and New Mexico) that include life insurance in their non-discrimination laws exempt companies if genetic data reliably, and based on “sound actuarial principles” (Legislature 2025), gives information relating to mortality or morbidity (see Section 5.4 for a detailed case study).

**Genetic Data Privacy Laws.** Many states have genetic data privacy laws, either as a standalone law or as part of a broader data privacy law. These laws were generally passed after GINA with the rise of DTC genetic testing companies. Standalone genetic data privacy laws tend to be written specifically to cover DTC genetic testing companies and are often extensions of genetic non-discrimination laws passed around the time of GINA.

Both data privacy and genetic privacy laws have similar provisions (Prince et al. 2021), including a customer right to request deletion of genetic data and samples and the ne-

cessity of consent for data transfers. Some states allow consumers to bring lawsuits for violations of the privacy law (a private right of action). In general though, privacy laws do not apply to deidentified genetic data (as per HIPAA). An important contrast to the US federal and state bills on privacy is the EU General Data Protection Regulation (GDPR). GDPR not only regulates all personal data within the EU but also the transfer of data outside of the EU, data minimization, and compliance. Genetic data is labeled as “special category data” in GDPR as it has many identifiers associated with it.

## Blindspots for Regulatory Bodies Governing Genetic Data Collection and Usage

A critical tension within the current protective framework, predating even advanced AI applications, lies between traditional single-gene tests and newer probabilistic measures of genetic risk, notably polygenic scores (PGS). Legislation like GINA was largely envisioned for single-gene tests that detect specific variants directly linked to conditions (e.g., BRCA1/2 for breast cancer, HTT for Huntington’s). These tests offer relatively deterministic insights, which existing laws sought to shield from discriminatory use.

However, PGS operate differently. They don’t “detect” individual genotypes or mutations in the direct sense that GINA outlines in its definition of “genetic information” (an analysis that “detects genotypes, mutations, or chromosomal changes” (Congress 2008)). Instead, PGS represent a secondary computational analysis, transforming raw genetic data into statistical models of relative risk across populations. This transformation from direct detection to probabilistic, interpretive analysis arguably places PGS outside GINA’s explicit protections for employment and health insurance decisions (Callier and Prince 2024; Yanes et al. 2024). This potential loophole is significant because the law focuses on the detection of genetic information rather than the interpretive analyses and risk scores derived from it.

This distinction also manifests in regulatory oversight. The FDA, for instance, regulates well-defined single-gene tests with clear clinical actions, but PGS largely fall outside its purview. While the FTC might address PGS claims under fraud or deceptive practice statutes, this offers a different, and potentially less direct, form of consumer protection than the specific anti-discrimination mandates of GINA or the clinical validation oversight of the FDA.

## Risk Assessment Framework

Our goal in this section is to develop a risk assessment framework for genetic data governance. We adopt the structure of the aims of the Blueprint for the AI Bill of Rights (Park 2023) and motivate our framework with three central questions: (1) What values (i.e., moral principles and civil liberties) should be preserved? (2) What are the vulnerabilities in the current system that can compromise these values? (3) What are the harms that result from the vulnerabilities?

## What Values Should Be Preserved?

We seek to identify values that are future-proof and can be encoded in a system of genetic data collection, considering practical ideas of control, consent, and non-discrimination surrounding genetic data. Using Viljoen's framework of the relational theory of data governance (Viljoen 2024), we identify vertical relationships (between data collectors, i.e. the Four Pillars, and individuals) and horizontal relationships (between genetically connected individuals). This framework must protect personal and civil liberties for both data submitters and their implicated genetic relatives, encompassing both vertical and horizontal relationality. Through this analysis, six key values emerge. The first four values are drawn particularly from Nissenbaum's framework of contextual integrity, highlighting genetic data transfer while preserving personal freedoms (Nissenbaum 2019, 2004; Barth et al. 2006).

1. **Right to action:** The individual, and the individual only, has the choice to submit (and the freedom to not submit) their genome.
2. **Ownership of the genome:** The individual owns their genome and therefore controls the usage of their genome, including *who* has access to it and for what reason.
3. **Right to privacy:** The individual has a right to privacy with regards to their genome and inferences made from their genome.
4. **Right to knowledge:** The individual has a right to know or *not* know about inferences made from their genome.
5. **Protecting opportunities for advancement:** Genetic data should not be used to deprive the individual of opportunities in any domain, including (but not limited to) education, access to financial tools, health insurance, housing, social services, and reproductive choices, thus preventing a "genetic underclass" (Ajunwa 2016).
6. **Benefits of inclusion:** The Belmont principle states "those who bear the burdens of research (i.e., those who are exposed to the discomforts, inconveniences, and risks) should receive the benefits in equal measure to the burdens" (U.S. Department of Health & Human Services (HHS) 1979). We believe that this applies to individuals and their genetic data.

## What Are the Vulnerabilities in the Current System That Can Compromise These Values?

The current genetic data ecosystem has many vulnerabilities that compromise the values listed above. These include:

1. **Unsettled science:** The role of genetics in shaping complex traits is poorly understood, and integrating genetic data into multi-modal AI models will likely not improve this understanding. As a consequence, the deterministic nature of genetics may be overstated, particularly for behavioral and cognitive traits.
2. **Rapid evolution of genetic data/methods:** As genetic data collection, sequencing, and methodologies quickly evolve, legal protections fall behind. For example, education discrimination is not included in GINA, which was

passed five years before the first large GWAS on educational attainment (Rietveld et al. 2013).

3. **Guilt by association:** DNA databases used for criminal investigations can impact not only the individuals whose DNA is stored, but also their biological relatives, as genetic information is shared among family members.
4. **Geographical legislative patchwork:** Genetic privacy and anti-discrimination laws vary widely across states, with unclear rules on data removal, third-party sharing, and other protections. This patchwork leaves individuals vulnerable when transferring data across state lines, and many states lack specific safeguards for genetic data.

## What Are the Harms That Arise as a Result?

We outline the consequences that could arise from violations of our key values through the above vulnerabilities. These downstream effects, which we refer to as harms, can not only affect an individual and their immediate genetic relatives and but can also have a broader effect on communities and the population at large.

1. **Leakage to the family:** Any conclusion about an individual's genetic trait may be linked back to their relatives, even if no action is taken by their relatives in either (1) submitting data or (2) opting in to receive information. Any derived secondhand knowledge can immediately impact insurance (health, life, etc.) and medical treatment plans, as well as compromise identity.
2. **Loss of anonymity:** It is straightforward to identify some traits about an individual from their genetic data, and subsequently from their genetic relatives. Common examples include: race, gender, ethnicity, and markers for certain diseases. As the refinement of reference datasets continues and data collection grows, leaky genetic data interfaces can thus compromise information about an individual, their genetic relatives, and unborn children that the individual (and implicated relatives) may wish to maintain as private.
3. **Loss of Data Control:** Private institutions offering genetic tests control the entire lifecycle of individuals' data from collection and storage to monetization. This control creates significant transparency issues, as it is unclear how data is monetized, which parties benefit, or what agreements include an individual's data. Users may opt out of participation, but unregulated data flows ensure that their genetic footprint persists indefinitely. This lack of transparency is particularly concerning in cases of bankruptcy or acquisition. For example, in 2024, Tempus AI acquired Ambry Genetics, a company that sequences genetic data for 400,000 patients annually (Tempus 2025). Tempus stated its intent to "leverage this data and augment Tempus' current data offering," raising questions about how patient data is repurposed and monetized in such deals. Ultimately, an individual's decision to submit genetic data for testing often entails relinquishing full ownership of their genetic information. Without robust safeguards, privacy risks emerge from opaque data practices and leaky interfaces, leaving individuals vulnerable to misuse or unauthorized access.

4. **Misinformed Actions:** The interpretation of genetic test results and the subsequent actions taken by individuals are profoundly personal, often influencing lifestyle changes, significant financial decisions, and critical medical choices. For instance, individuals with BRCA1 gene mutations indicating an elevated breast cancer risk may opt for preventive interventions. These decisions are particularly sensitive as the interpretation of genetic data evolves alongside advancements in scientific methods. However, the relational nature of genetic data complicates individual autonomy and access to information. It is possible for someone to receive conclusions about themselves indirectly through the test results of genetic relatives. This dynamic, combined with the widespread availability of unregulated genetic tests and the evolving nature of genetic science, increases the potential for harm if individuals act on information that is incomplete, inaccurate, or probabilistic rather than deterministic.
5. **Financial impact:** Individuals may face financial repercussions if insurance or legal policies fail to provide adequate protection or if coverage is denied. Beyond medical procedures, financial impacts can also arise on a case-by-case basis, such as in situations where genetic data is held for ransom, legal defense is required, or individuals are forced to purchase direct-to-consumer tests (e.g., in order to purchase a life insurance policy).

## Case Studies

We apply our risk assessment framework to four case studies which highlight regulatory gaps across the Four Pillars. These case studies, ranging from past events to speculative scenarios, all involve privacy loss or potential discrimination. We explore case studies that illustrate genetic data leakage in both, a pillar-to-pillar and one-to-many pillar settings.

### Genetics and Education

Polygenic scores (PGS) are popular in the social/behavioral sciences for predicting social outcome traits, such as educational attainment (EA; number of schooling years completed by an adult). We will use EA as the example throughout this case study, but note that there is interest in predicting standardized testing scores, mathematics performance, and other traits with substantial environmental influences.

A common metric for assessing PGS accuracy is the percentage of trait variance it explains: higher percentages indicate better predictive performance. A recent 23andMe EA PGS, based on data from over 3 million customers of European descent, explains 12-16% of the variance in educational attainment (EA) (Okbay et al. 2022). This effectively means that 50-70% of individuals with PGS scores in the top 10% for EA are predicted to graduate college. However, the PGS accuracy significantly decreases when applied to African American customers. This is an example of the commonly observed “portability problem” (Martin et al. 2017), where a PGS derived from GWAS in one population predicts poorly in another due to confounding.

There have been several calls to use EA PGS to inform education policy. Harden et al. propose the use of math-

performance PGS to identify “leaks” in the education system: for example, by identifying high math PGS students who perform poorly, they claim educators could pinpoint why and how students are failing to reach their potential (Harden et al. 2020). Plomin & von Stumm take it further: they use the term “precision education” (akin to precision medicine) to propose a tailor-made, individualized education that is genetics-informed (Plomin and von Stumm 2018). Statements like this, combined with statements such as “students with higher polygenic scores for years of education have, on average, higher cognitive ability, better grades and come from families with higher SES [socioeconomic status]” (Smith-Woolley et al. 2018) are cause for concern because they invoke a sense of genetic determinism. However, other predictors (parents’ educational status, socioeconomic status) explain similar amounts of variance in EA (Morris, Davies, and Davey Smith 2020; Mostafavi et al. 2020) and—unlike DNA—are mutable through social policy changes.

Several of our values would be violated if children were required to submit their DNA (Right to Action) or educational opportunities were denied to children based on their genetic potential (Opportunities for Advancement). Through the vulnerabilities of unsettled science and the rapid evolution of genetic methods, leakage to the family can occur and affect not only children, but their families and future. Crucially, these profound challenges exist even before considering the added complexities and potential for misuse introduced by integrating such scores with advanced AI systems.

### DTC Genetic Data Brokerage and Leakage

The fragility of genetic data governance is starkly illustrated by direct-to-consumer (DTC) company 23andMe. Beyond FDA-approved genetic tests (e.g., BRCA1/2 variants), carrier screenings, health risk assessments, and genealogy services, the company also collects extensive non-genetic survey data and conducts its own GWAS.

In 2023, a credential stuffing attack compromised 14,000+ user accounts. Exploiting 23andMe’s relative linkage feature hackers accessed data for an additional 5.5M users (23andMe 2023b) in a powerful example of genetic data’s relationality. The hackers also reportedly targeted Chinese and Ashkenazi Jewish profiles (Carballo, Schmall, and Tumin 2024). A critical factor enabling this breach was the lack of consensus around security responsibility, which resulted in the disproportionate and harmful targeting of underrepresented groups. The breach highlighted how an individual’s genetic privacy is hostage to others’ security practices, aptly summarized by a Reddit user: “Your genetic data is only as secure as your relatives’ passwords.”

The security and integrity of consumer genetic data was once again compromised when 23andMe filed for bankruptcy and announced its database and assets were for sale (23andMe 2025). This prompted widespread calls for data deletion (Eddy 2025) and a federal probe into the sale’s privacy and national security risks (U.S. House Committee on Oversight and Government Reform 2025). In May 2025, it was announced that pharmaceutical giant Regeneron won the bid to buy 23andMe’s assets for \$256M (Regeneron

Pharmaceuticals, Inc. 2025). Regeneron would gain access to existing genetic data, extensive non-genetic information, model weights, and stored biological samples, which allow for future whole-genome sequencing. Regeneron, in compliance with the terms of the sale (23andMe, Inc. 2025), pledged to uphold 23andMe's terms and conditions (T&Cs), which include the right to delete data (Regeneron Pharmaceuticals, Inc. 2025). More than two dozen states sued to block the sale and bidding was reopened (Ruwitch 2025). In June 2025, it was announced that 23andMe founder and ex-CEO Anne Wojcicki's newly-formed non-profit TTAM Research Institute would buy the assets for \$305M.

The 23andMe saga reveals critical vulnerabilities and rights violations within our risk assessment framework. The data breach infringing on Right to Privacy, while the (failed) asset sale to Regeneron highlighted contested Ownership of the Genome, as users' deletion rights via T&Cs contrast with the de facto transfer of dataset ownership. This underscores the dangers of a fragmented regulatory landscape: Californians have protections during third-party sales, but those in states like Pennsylvania lack equivalent federal or state safeguards for DTC genetic data sales. Finally, Regeneron's proposed acquisition of biological samples and non-genetic data would have exposed vulnerabilities from rapidly evolving technology and unsettled science, enabling applications far beyond users' original consent or understanding.

### **Genetic Data Collection of Detained Noncitizens**

A recent Georgetown University report from the Center of Privacy and Technology (CPT) examines the U.S. federal government's practice of collecting DNA from detained migrants (Glaberson, Tse, and Tucker 2024). This practice, which began with the 2005 DNA Fingerprint Act and expanded significantly in 2020—a mandate that the Department of Homeland Security must collect DNA from all detainees, even briefly detained—has led to a dramatic increase in detainee representation in CODIS (the federal DNA database), rising from 0.21% in 2019 to 9.21% in 2023. By 2020, approximately 25,000 noncitizens were added to the database under the “detainee” classification.

The CPT report highlights several concerning aspects of this practice, particularly its disproportionate impact on migrants of color and issues of consent. In our Risk Assessment Framework, this immediately violates the Right to Action and Right to Knowledge. Many migrants undergo DNA collection without understanding its implications, sometimes believing it to be a COVID-19 test or submitting under threat of criminal prosecution. The report argues that this practice violates the Fourth Amendment in collecting DNA from detainees without probable cause. Additionally, even though CODIS was designed to be privacy-protective by collecting only 20 markers—a tiny slice of the genome believed to be medically neutral—this limited data can still identify relatives through partial matching and, as an analysis 25 years later would show, are slightly informative of disease risk (Bañuelos et al. 2022). The expansive collection of genetic data has far-reaching implications through the guilt-by-association vulnerability. Since CODIS records are difficult to expunge, they can restrict Opportunities for Advance-

ment not only for detainees but also for their relatives/descendants.

### **Underwriting Life Insurance with AI and Genetic Data**

Life insurance underwriting is the process whereby an insurance company uses personal and health information to assess the risk of insuring an applicant. The relationship between applicants and insurance companies is already fraught: (Devnos, Wang, and Iyer 2016) found that patients are more likely to share their genetic information with Google than insurance companies. The future of life insurance underwriting is expected to become increasingly computational and automated through the usage of AI and the collection of more personalized, individual genetic data (Balasubramanian, Libarikian, and McElhaney 2021; Filabi and Duffy 2021; Maier et al. 2020; Perumalsamy, Konidena, and Krothapalli 2023). While medical records and demographic factors can be used for mortality analysis, the inclusion of genetic factors can mean that risk prediction can be performed much earlier in an applicant's life without necessarily the same amount of records as an older individual, emphasizing the risk of genetic discrimination (Karls-son Linnér and Koellinger 2022) based on the potential to have risk factors (e.g., a genetic risk for high blood pressure versus clinically-measured high blood pressure). Life insurance is not covered by GINA and life insurance companies can access medical records (which may include genetic test results) as part of an application. There is evidence of industry pressure and lobbying to shift the focus of bills to not protect genetic discrimination in life insurance (Rothstein 2018). Bills have been introduced in several states that would restrict the use of genetic information in underwriting, but these efforts largely failed: as of 2022, out of thirty-seven proposed bills across all states, three were introduced, eight were signed by the governor, demonstrating the geographical dependency of policy-based protection (Vermont Legislature 2019).

Life insurance companies may be interested in using PGS to assess an individual's risk for various diseases. In early 2024, U.S.-based life insurance company MassMutual and U.K.-based Genomics PLC announced a partnership, offering free genetic testing to MassMutual's life insurance customers (MassMutual 2024). However, the press release stressed that MassMutual would not receive individual results and that current premiums/policies would be unaffected. This example particularly highlights the leaky interface between pillars, where the Right of Privacy and Right of Action are called into question. AI insurance underwriting algorithms already suffer from racial biases (Lee 2022). That PGS suffer from the portability problem (itself caused by systematic biases in training datasets) means that these biases could be perpetuated in genetics-informed life insurance underwriting. Additionally, given the relational aspect of genetic data, genetic-based underwriting could affect other biological relatives' applications, harming not only relatives but any others with the same genetic markers that are associated with mortality risks. The vulnerabilities of (1) unsettled science and (2) rapid evolution and usages of genetic

data can exacerbate the potential harms of using genetic data as part of underwriting in life insurance.

## Recommendations

### Recommendation 1: Redefining Genetic Data

**Issue:** Existing policy language does not adequately define genetic data or genetic information.

**Recommendation:** Given that genetic data possess unique characteristics distinct from other identifying information, we propose a two-tier definitional framework.

1. **Genetic data** refers to any information relating to an individual’s genetic characteristics, including but not limited to DNA or RNA sequences, genomic variants, gene expression profiles, epigenetic markers, or genetic sequences from biological relatives. This data shall always be considered personally identifiable information (PII) as it pertains to unique biological attributes that can potentially be linked to a specific individual, their biological relatives, or identifiable group. For the purposes of this definition, any de-identified, pseudonymized, or anonymized genetic information shall be treated as genetic data, regardless of measures taken to mask individual identities, recognizing that the inherent characteristics of genetic data may enable re-identification through advanced technological or data cross-referencing methods.
2. **Genetic information** encompasses any downstream inferences, predictions, prior inferences, or derivations extracted from genetic data, including but not limited to health predispositions, ancestry determinations, trait predictions, and algorithmic outputs based on genetic data analysis.

While genetic information may not be immediately identifiable as traditional PII, we recommend that both genetic data and genetic information be treated as a *distinct protected category for sensitive data* requiring enhanced safeguards and the same level of protections as PII and protected health information (PHI). This approach draws from established legal precedent in biometric privacy law, specifically the distinction between “biometric identifiers” and “biometric information” established in Illinois’s Biometric Information Privacy Act (Illinois General Assembly 2008).

### Recommendation 2: Extending Protections for Genetic Discrimination

**Issue:** While GINA covers health insurance and employment, it does not protect life, long-term care, and disability insurance, as well as other domains (e.g., housing, education).

**Recommendation:** We recommend that additional federal laws should be enacted to extend GINA’s coverage beyond employment and health insurance and for GINA to adopt and encompass the definitions of genetic data and genetic information provided in Recommendation 1. The most comprehensive state genetic anti-discrimination law is California’s CalGINA, which includes housing, mortgage brokerage, education, and more. However, despite these extensions, CalGINA does not cover life, disability, or long-term

care insurance. We recommend that a comprehensive update to GINA should explicitly cover life, disability, long-term care, as well as education and any other opportunities for advancement to be protected against genetic discrimination. Legislation should be written in such a way that prohibits any barrier to opportunity, even those not anticipated at the time of writing. Additionally, it should explicitly bar genetic risk for complex traits known to have significant environmental influences (such as cardiovascular disease) from being considered a preexisting condition.

### Recommendation 3: A Genetic Data Regulation Framework

**Issue:** Current protections do not comprehensively apply to the full genetic data ecosystem which leads to *leaky* protections (e.g., clinical tests being used in life insurance).

**Recommendation:** We suggest a uniform regulatory framework that applies the same standards of collection, consent, and control to all data collectors in the genetic data ecosystem. In the spirit of the *legitimacy* challenge posed by Viljoen, we build upon specific values inherent in privacy rights to ensure that responsibility is placed on the organizations that hold and analyze data, rather than the individuals from whom it was collected, and the approach to consider data as a democratic medium (Solove 2022; Viljoen 2024).

1. **Entity approval:** Entities—whether they be universities, health care providers, hospital systems, or corporations—that collect or house any type of genetic data (DNA, RNA, or even genetic test results) must have prior regulatory approval. Approval requires a clear commitment to the basic rights of the individuals whose genetic data the entity will collect or own, such as data privacy, security, the right to deletion, etc.
2. **Test approval:** After entity approval, we propose regulations for the inferential tests, specifically those that return results to either an individual or a medical practitioner for any actionable result. At a minimum, entities would be required to publicly release detailed white papers for each test that detail laboratory procedures, quality control steps, inferential models used to validate the testing process for AI models, presentation of results, and any other necessary details that would allow one to recreate the analysis if given the same data.
3. **Powers given to the individual**
  - Individuals should be able to request data removal. This would include specifically (1) destruction of the physical sample, (2) deletion of data, (3) removal of any identifiers, and (4) any removed data no longer influence the results of any downstream models.
  - Third-party data transfers should only occur between companies authorized to collect or own genetic data, with recipient companies subject to the same regulations. We recommend providing individuals with the option of blanket opt-out from data transfers and specific opt-in choices for each third party. Individuals must be notified of the transfer and given a reasonable time to opt-out.

- Research usage is common among DTC companies, academia, and hospitals, where data provided from individual tests can be pooled together as a dataset for research (both internally and via external collaborations). We suggest an opt-out strategy here for individuals, with a specification on whether they consent to their data becoming publicly available in any form (e.g. public datasets, open sourced GWAS weights, trained models, etc).
  - Incidental genetic discoveries can be possible in research use cases where data for a particular test was used for other tests. In these cases, where knowledge of a particular trait can also be harmful and uninformative to the individual, we suggest an opt-in strategy for an individual to blanket choose if they wish to hear any secondary discoveries, or forward any secondary discoveries to their medical practitioner.
4. **Bankruptcy and Acquisition:** As detailed by third-party data transfers, the owning of genetic data would also apply to companies who acquire any genetic data as part of assets through acquisitions or bankruptcy. Thus, all entities involved must already have approval to handle genetic data. In the rare event that there are no entities to handle the data, we suggest a similar protocol structure as nuclear waste, where all physical sample, data, and models are subsequently destroyed and cannot be recovered.

## Conclusions and Open Challenges

In this paper we outline the current goals and approaches of scientific research in genetics, the genetic data ecosystem, and state of public protections. We identify critical gaps in current regulatory federal and state-level frameworks that prevent the productive, ethical, and safe adoption of genetic data for broad societal applications. To address this, we propose a novel risk assessment framework that offers key public protections and preserve's individuals personal liberties, and finally make three concrete policy recommendations. Taken together, we aim to create a more robust data governance system that institutes comprehensive individual protections while still enabling responsible innovation in human genetics, recognizing the need to balance privacy with the advancement of beneficial research.

While our policy recommendations focus on enhancing individual protections, we also identify three key challenges in regulating genetic data usage that fall outside the scope of our work and that would benefit from further research.

## The Relational Theory of Privacy

**Should Genetic Data Submissions Require Consent From Biological Relatives?** When an individual does a genetic test they are, indirectly, testing (part of) their relative's genome. What (if any) consent or privacy measures should be taken to protect their biological relatives who may not consent to such a test? Additionally, what should happen upon death? Do children "own" their parents' genome (and what ownership would, say, a full sibling have?) and can they make decisions (e.g., to delete the data)?

**Who Has Ownership of Children's DNA?** Do parents have a "fiduciary duty" to protect their child's genetic data and protect them from the system vulnerabilities that we have outlined? Part of this fiduciary duty would involve whether a genetic test should be taken in the first place, e.g., the difference between a medically-necessary test and a DTC genetic test for insights into the child's ancestry and genealogical relatives. The relational aspect of genetic data also comes into play. Should consent be required from both parents to sequence their child's genome? See (Bala 2023) for a discussion of child's genome ownership.

## International Genetic Data Transfers

Our paper focuses on U.S. regulatory and legislative bodies and U.S.-based companies. Should there be restrictions on genetic data transfer between countries? For example, one hypothetical scenario might involve a political refugee who has medically-necessary genetic testing done within the U.S, but because of leaky data transfer rules, genetic information is transferred to their home country where it might be used to identify family members now at risk for persecution.

## Collective Rights for Genetic Data

Privacy scholars acknowledge the necessity of collective rights when building on the relational nature of genetic data. Government-collected and/or -controlled genetic databases pose a fundamental challenge to democratic legitimacy, creating power imbalances that may alter citizen-state relationships and undermine collective political engagement with legal systems. Safeguarding individual data-subject rights, while important, does not address the deeper issue: the lack of collective, democratic decision-making regarding whether, how, and for what purposes genetic data is created and used. Key scholars addressing collective rights include King, who demonstrates this phenomenon in DTC genetic testing (King 2019); Solow-Neiderman, who details how individual genetic submissions compromise collective privacy (Solow-Niederman 2021); Costello, who raises issues of group consent and responsibility (Costello 2022); and Viljoen, who argues for democratizing data for the public collective (Viljoen 2024).

While robust theoretical arguments support collective rights frameworks, their path to implementation is less clear. Our analysis bridges existing policy approaches with emerging theoretical frameworks, providing a scaffold for comprehensive genetic data governance. In an increasingly data-driven society, there is an urgent need to address major gaps in genetic data governance, which informs our process to develop a risk assessment framework and outline three concrete policy recommendations. This work represents not an endpoint but a call to action for policymakers, researchers, and ethicists to collaborate on transforming theoretical insights into practicable governance solutions that safeguard genetic privacy for present and future generations.

## Acknowledgments

We would like to acknowledge Rui-Jie Yew for helpful discussions in formulating this project and Emily Tucker for

comments on an earlier draft of this paper. This research was supported by grants from the MacArthur Foundation and Ford Foundation—along with private donations—to Brown University’s Center for Technological Responsibility, Reimagination, and Redesign (CNTR); by National Institutes of Health grants R35 GM139628 (S.R.) and T32 GM128596 (S.R., C.W., R.V., V.R.); and by the National Science Foundation (C.W.).

## References

- 23andMe. 2023a. 23andMe Reports FY2023 Fourth Quarter and Full Year Financial Results. *23andMe Investor Relations*. Press Release. Accessed January 16, 2025.
- 23andMe. 2023b. Addressing Data Security Concerns – Action Plan. *23andMe Blog*. Originally published October 6, 2023. Updated December 5, 2023. Accessed January 16, 2025.
- 23andMe. 2025. 23andMe Provides Update Regarding Court-Supervised Sale Process. *23andMe Investor Relations*. Press Release.
- 23andMe, Inc. 2025. 23andMe Confirms All Potential Buyers Must Agree to Comply With Privacy Policy and Applicable Law. *Global Newswire*. March 26, 2025.
- Ajunwa, I. 2016. Genetic Data and Civil Rights. *Harvard Civil Rights-Civil Liberties Law Review*, 51(1): 75–114.
- Ancestry. 2025. Company Facts: The Many Ways Our Family Has Grown. *Ancestry Corporate*. Accessed January 22, 2025.
- Bala, N. 2023. Who Owns Children’s DNA? *Michigan Law Review*, 122: 457.
- Balasubramanian, R.; Libarikian, A.; and McElhaney, D. 2021. Insurance 2030— The impact of AI on the future of insurance. Technical report, McKinsey and Company.
- Barr, P. B.; Ksinan, A.; Su, J.; Johnson, E. C.; Meyers, J. L.; Wetherill, L.; Latvala, A.; Aliev, F.; Chan, G.; Kuperman, S.; Nurnberger, J.; Kamarajan, C.; Anokhin, A.; Agrawal, A.; Rose, R. J.; Edenberg, H. J.; Schuckit, M.; Kaprio, J.; and Dick, D. M. 2020. Using polygenic scores for identifying individuals at increased risk of substance use disorders in clinical and population samples. *Translational Psychiatry*, 10(1): 196.
- Barth, A.; Datta, A.; Mitchell, J.; and Nissenbaum, H. 2006. Privacy and contextual integrity: framework and applications. In *2006 IEEE Symposium on Security and Privacy (S&P’06)*, 15 pp.–198. Berkeley/Oakland, CA: IEEE. ISBN 978-0-7695-2574-7.
- Bañuelos, M. M.; Zavaleta, Y. J. A.; Roldan, A.; Reyes, R.-J.; Guardado, M.; Chavez Rojas, B.; Nyein, T.; Rodriguez Vega, A.; Santos, M.; Huerta-Sanchez, E.; and Rohlf, R. V. 2022. Associations between forensic loci and expression levels of neighboring genes may compromise medical privacy. *Proceedings of the National Academy of Sciences*, 119(40): e2121024119.
- Bonomi, L.; Huang, Y.; and Ohno-Machado, L. 2020. Privacy Challenges and Research Opportunities for Genomic Data Sharing. *Nature Genetics*, 52(7): 646–654.
- Border, R.; Athanasiadis, G.; Buil, A.; Schork, A. J.; Cai, N.; Young, A. I.; Werge, T.; Flint, J.; Kendler, K. S.; Sankaranarayanan, S.; Dahl, A. W.; and Zaitlen, N. A. 2022. Cross-trait assortative mating is widespread and inflates genetic correlation estimates. *Science*, 378(6621): 754–761. Publisher: American Association for the Advancement of Science.
- Bycroft, C.; Freeman, C.; Petkova, D.; Band, G.; Elliott, L. T.; Sharp, K.; Motyer, A.; Vukcevic, D.; Delaneau, O.; O’Connell, J.; Cortes, A.; Welsh, S.; Young, A.; Effingham, M.; McVean, G.; Leslie, S.; Allen, N.; Donnelly, P.; and Marchini, J. 2018. The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 562(7726): 203–209. Publisher: Nature Publishing Group.
- Callier, S.; and Prince, A. E. 2024. The Legal Uncertainties of Sociogenomic Polygenic Scores. *Harvard Journal of Law & Technology*, 38(2): 615–676. Symposium 2024.
- Carballo, R.; Schmall, E.; and Tumin, R. 2024. 23andMe Breach Targeted Jewish and Chinese Customers, Lawsuit Says. *The New York Times*. Accessed January 16, 2025.
- Carlson, J.; Henn, B. M.; Al-Hindi, D. R.; and Ramachandran, S. 2022. Counter the weaponization of genetics research by extremists. *Nature*, 610(7932): 444–447.
- Citron, D. K. 2020. A New Compact for Sexual Privacy. Commission, F. T. 2024. Enforcement.
- Commissioner, O. o. t. 2023. What We Do. Publisher: FDA.
- Congress, U. S. 2008. Genetic Information Nondiscrimination Act of 2008. *United States Statutes at Large*, 122: 881–922. Public Law 110-233, 42 U.S.C. § 2000ff, May 21, 2008.
- Costello, R. Á. 2022. Genetic Data and the Right to Privacy: Towards a Relational Theory of Privacy? *Human Rights Law Review*, 22(1): ngab031.
- Daval, C. J. R.; and Kesselheim, A. S. 2023. Authority of Medicare to Limit Coverage of FDA-Approved Products: Legal and Policy Considerations. *JAMA Internal Medicine*, 183(9): 999–1004.
- Davis, C. N.; Jinwala, Z.; Hatoum, A. S.; Toikumo, S.; Agrawal, A.; Rentsch, C. T.; Edenberg, H. J.; Baurley, J. W.; Hartwell, E. E.; Crist, R. C.; Gray, J. C.; Justice, A. C.; Gelernter, J.; Kember, R. L.; Kranzler, H. R.; and VA Million Veteran Program. 2025. Utility of Candidate Genes From an Algorithm Designed to Predict Genetic Risk for Opioid Use Disorder. *JAMA Network Open*, 8(1): e2453913.
- Devnos, L.; Wang, T.; and Iyer, S. 2016. The genomics inflection point: Implications for healthcare | Rock Health.
- Duello, T. M.; Rivedal, S.; Wickland, C.; and Weller, A. 2021. Race and genetics versus ‘race’ in genetics. *Evolution, Medicine, and Public Health*, 9(1): 232–245.
- Eddy, M. 2025. 23andMe Just Filed for Bankruptcy. You Should Delete Your Data Now. *The New York Times*. Updated March 25, 2025.
- Epstein, C. J. 2003. Is modern genetics the new eugenics? *Genetics in Medicine*, 5(6): 469–475.
- Feiszt, P.; and Kenny, P. A. 2025. Abstract 2773: The Gundersen Precision Oncology Cohort: a biospecimen resource

- for mutation characterization and target discovery. *Cancer Research*, 85(8\_Supplement\_1): 2773–2773.
- Filabi, A.; and Duffy, S. 2021. AI-Enabled Underwriting Brings New Challenges for Life Insurance: Policy and Regulatory Considerations. *Journal of Insurance Regulation*.
- Fry, A.; Littlejohns, T. J.; Sudlow, C.; Doherty, N.; Adamska, L.; Sprosen, T.; Collins, R.; and Allen, N. E. 2017. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *American Journal of Epidemiology*, 186(9): 1026–1034.
- Glaberson, S.; Tse, E.; and Tucker, E. 2024. Raiding the Genome: How the United States Government is Abusing Its Immigration Powers to Amass DNA for Future Policing. *Center On Privacy & Technology at Georgetown Law*.
- Green, R. C.; Lautenbach, D.; and McGuire, A. L. 2015. GINA, Genetic Discrimination, and Genomic Medicine. *New England Journal of Medicine*, 372(5): 397–399. Publisher: Massachusetts Medical Society. eprint: <https://www.nejm.org/doi/pdf/10.1056/NEJMp1404776>.
- Gunter, C.; and Green, E. D. 2023. To boldly go: Unpacking the NHGRI's bold predictions for human genomics by 2030. *American Journal of Human Genetics*, 110(11): 1829–1831.
- Harden, K. P.; Domingue, B. W.; Belsky, D. W.; Boardman, J. D.; Crosnoe, R.; Malanchini, M.; Nivard, M.; Tucker-Drob, E. M.; and Harris, K. M. 2020. Genetic associations with mathematics tracking and persistence in secondary school. *npj Science of Learning*, 5(1): 1.
- Hargreaves, S. 2017. 'Relational Privacy' & Tort. *William & Mary Journal of Women and the Law*, 23(3): 433.
- Hatoum, A. S.; Wendt, F. R.; Galimberti, M.; Polimanti, R.; Neale, B.; Kranzler, H. R.; Gelernter, J.; Edenberg, H. J.; and Agrawal, A. 2021. Ancestry may confound genetic machine learning: Candidate-gene prediction of opioid use disorder as an example. *Drug and Alcohol Dependence*, 229(Pt B): 109115.
- Howe, L. J.; Nivard, M. G.; Morris, T. T.; Hansen, A. F.; Rasheed, H.; Cho, Y.; Chittoor, G.; Ahlskog, R.; Lind, P. A.; Palviainen, T.; Van Der Zee, M. D.; Cheesman, R.; Mangino, M.; Wang, Y.; Li, S.; Klaric, L.; Ratliff, S. M.; Bielak, L. F.; Nygaard, M.; Giannelis, A.; Willoughby, E. A.; Reynolds, C. A.; Balbona, J. V.; Andreassen, O. A.; Ask, H.; Boomsma, D. I.; Campbell, A.; Campbell, H.; Chen, Z.; Christofidou, P.; Dahm, C. C.; Evans, L. M.; De Geus, E. J. C.; Giddaluru, S.; Gordon, S. D.; Harden, K. P.; Hill, W. D.; Hughes, A.; Kerr, S. M.; Kim, Y.; Kweon, H.; Latalva, A.; Lawlor, D. A.; Li, L.; Lin, K.; Magnus, P.; Magnusson, P. K. E.; Mallard, T. T.; Martikainen, P.; Mills, M. C.; Nølstad, P. R.; Pedersen, N. L.; Porteous, D. J.; Reid, J.; Silventoinen, K.; Southey, M. C.; Stoltenberg, C.; Tucker-Drob, E. M.; Wright, M. J.; Kaprio, J.; Medland, S. E.; Walters, R. G.; Evans, D. M.; Smith, G. D.; Hayward, C.; Brumpton, B.; Hemani, G.; Davies, N. M.; Åsvold, B. O.; Naess, Ø.; Millwood, I. Y.; Willer, C. J.; Benjamin, D. J.; Turley, P.; Koellinger, P. D.; Social Science Genetic Association Consortium; and Within Family Consortium. 2022. Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. *Nature Genetics*, 54(5): 581–592.
- Illinois General Assembly. 2008. Biometric Information Privacy Act. 740 ILCS 14/1 et seq. Illinois Compiled Statutes.
- Jillson, E. 2024. The DNA of privacy and the privacy of DNA.
- Jonsson, H.; Magnusdottir, E.; Eggertsson, H. P.; Stefansson, O. A.; Arnadottir, G. A.; Eiriksson, O.; Zink, F.; Helgason, E. A.; Jonsdottir, I.; Gylfason, A.; Jonasdottir, A.; Jonasdottir, A.; Beyter, D.; Steingrimsdottir, T.; Norddahl, G. L.; Magnusson, O. T.; Masson, G.; Halldorsson, B. V.; Thorsteinsdottir, U.; Helgason, A.; Sulem, P.; Gudbjartsson, D. F.; and Stefansson, K. 2021. Differences between germline genomes of monozygotic twins. *Nature Genetics*, 53(1): 27–34.
- Kaiser, J. 2018. We will find you: DNA search used to nab Golden State Killer can home in on about 60% of white Americans. *Science*.
- Karlsson Linnér, R.; and Koellinger, P. D. 2022. Genetic risk scores in life insurance underwriting. *Journal of Health Economics*, 81: 102556.
- Kaye, J. 2012. The Tension Between Data Sharing and the Protection of Privacy in Genomics Research. *Annual Review of Genomics and Human Genetics*, 13(1): 415–431.
- King, J. 2019. "Becoming Part of Something Bigger": Direct to Consumer Genetic Testing, Privacy, and Personal Disclosure. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW): 158:1–158:33.
- Kretschmer, T.; Ouellet-Morin, I.; Vrijen, C.; Nolte, I. M.; and Hartman, C. 2023. Polygenic risk for aggressive behavior from late childhood through early adulthood. *European Child & Adolescent Psychiatry*, 32(4): 651–660.
- Lala, K. N.; and Feldman, M. W. 2024. Genes, culture, and scientific racism. *Proceedings of the National Academy of Sciences*, 121(48): e2322874121.
- Lee, R. 2022. AI can perpetuate racial bias in insurance underwriting. *Yahoo Finance*. Accessed January 2, 2025.
- Legislature, C. S. 2011. SB-559 Discrimination: genetic information. *California Legislative Information*, Chapter 261. Approved September 6, 2011.
- Legislature, M. 2025. Section 120E: Unfair discrimination based on results of genetic test. *Massachusetts General Laws*, Part I, Title XXII, Chapter 175.
- Loos, R. J. F. 2020. 15 years of genome-wide association studies and no signs of slowing down. *Nature Communications*, 11(1): 5900. Publisher: Nature Publishing Group.
- Maier, M.; Carlotto, H.; Saperstein, S.; Sanchez, F.; Balogun, S.; and Merritt, S. 2020. Improving the Accuracy and Transparency of Underwriting with AI to Transform the Life Insurance Industry. *AI Magazine*, 41(3): 78–93. Number: 3.
- Martin, A. R.; Gignoux, C. R.; Walters, R. K.; Wojcik, G. L.; Neale, B. M.; Gravel, S.; Daly, M. J.; Bustamante, C. D.; and Kenny, E. E. 2017. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *The American Journal of Human Genetics*, 100(4): 635–649.

- MassMutual. 2024. Genomics plc and MassMutual's program enables more policyowners to understand health risks through innovative genetic testing. *MassMutual Newsroom*. Accessed January 2, 2025.
- Menikoff, J.; Kaneshiro, J.; and Pritchard, I. 2017. The Common Rule, Updated. *New England Journal of Medicine*, 376(7): 613–615. Publisher: Massachusetts Medical Society. eprint: <https://www.nejm.org/doi/pdf/10.1056/NEJMp1700736>.
- Morris, T. T.; Davies, N. M.; and Davey Smith, G. 2020. Can education be personalised using pupils' genetic data? *eLife*, 9: e49962. Publisher: eLife Sciences Publications, Ltd.
- Mostafavi, H.; Harpak, A.; Agarwal, I.; Conley, D.; Pritchard, J. K.; and Przeworski, M. 2020. Variable prediction accuracy of polygenic scores within an ancestry group. *eLife*, 9: e48376.
- National Institutes of Health. 2024. All of Us Research Program Strategic Goals. *All of Us Research Program*. Accessed January 2, 2025.
- Nissenbaum, H. 2004. Privacy as Contextual Integrity. *Washington Law Review*, 79(1): 119.
- Nissenbaum, H. 2019. Contextual Integrity Up and Down the Data Food Chain. *Theoretical Inquiries in Law*, 20(1): 221–256.
- Nivard, M. G.; Belsky, D.; Harden, K. P.; Baier, T.; Andreassen, O. A.; Ystrom, E.; Van Bergen, E.; and Lyngstad, T. H. 2022. Neither nature nor nurture: Using extended pedigree data to understand indirect genetic effects on offspring educational outcomes.
- of Representatives, F. H. 2020. HB 1189: Genetic Information for Insurance Purposes. *Florida Senate*.
- Okbay, A.; Wu, Y.; Wang, N.; Jayashankar, H.; Bennett, M.; Nehzati, S. M.; Sidorenko, J.; Kweon, H.; Goldman, G.; Gjorgjieva, T.; Jiang, Y.; Hicks, B.; Tian, C.; Hinds, D. A.; Ahlskog, R.; Magnusson, P. K. E.; Oskarsson, S.; Hayward, C.; Campbell, A.; Porteous, D. J.; Freese, J.; Herd, P.; Watson, C.; Jala, J.; Conley, D.; Koellinger, P. D.; Johansson, M.; Laibson, D.; Meyer, M. N.; Lee, J. J.; Kong, A.; Yengo, L.; Cesarini, D.; Turley, P.; Visscher, P. M.; Beauchamp, J. P.; Benjamin, D. J.; and Young, A. I. 2022. Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nature Genetics*, 54(4): 437–449. Publisher: Nature Publishing Group.
- Panofsky, A.; Dasgupta, K.; and Iturriaga, N. 2021. How <span style="font-variant:small-caps;">White</span> nationalists mobilize genetics: From genetic ancestry and human biodiversity to counterscience and metapolitics. *American Journal of Physical Anthropology*, 175(2): 387–398.
- Park, E. 2023. The AI Bill of Rights: a step in the right direction. *Orange County Lawyer Magazine*, 65(2).
- Parkman, A. A.; Foland, J.; Anderson, B.; Duquette, D.; Sobotka, H.; Lynn, M.; Nottingham, S.; Dotson, W. D.; Kolor, K.; and Cox, S. L. 2015. Public awareness of genetic nondiscrimination laws in four states and perceived importance of life insurance protections. *Journal of Genetic Counseling*, 24(3): 512–521.
- Perumalsamy, J.; Konidena, B. K.; and Krothapalli, B. 2023. AI-Driven Risk Modeling in Life Insurance: Advanced Techniques for Mortality and Longevity Prediction. *Journal of Artificial Intelligence Research and Applications*, 3(2): 392–422. Number: 2.
- Plomin, R.; and von Stumm, S. 2018. The new genetics of intelligence. *Nature reviews. Genetics*, 19(3). Publisher: Nat Rev Genet.
- Pozzi, G. 2023. Automated opioid risk scores: a case for machine learning-induced epistemic injustice in healthcare. *Ethics and Information Technology*, 25(1): 3.
- Prince, A. E.; Uhlmann, W. R.; Suter, S. M.; and Scherer, A. M. 2021. Genetic testing and insurance implications: Surveying the US general population about discrimination concerns and knowledge of the Genetic Information Nondiscrimination Act (GINA). *Risk management and insurance review*, 24(4): 341–365.
- Protections (OHRP), O. f. H. R. 2009. Federal Policy for the Protection of Human Subjects ('Common Rule'). Last Modified: 2024-03-27T10:49:52-0400.
- Regeneron Pharmaceuticals, Inc. 2025. Regeneron Enters into Asset Purchase Agreement to Acquire 23andMe® for \$256 Million; Plans to Maintain Consumer Genetics Business and Advance Shared Goals of Improving Human Health and Wellness. *Regeneron*. Press Release, May 19, 2025.
- Research, C. f. D. E. a. 2019. Institutional Review Boards (IRBs) and Protection of Human Subjects in Clinical Trials. *FDA*. Publisher: FDA.
- Rietveld, C. A.; Medland, S. E.; Derringer, J.; Yang, J.; Esko, T.; Martin, N. W.; Westra, H.-J.; Shakhbazov, K.; Abdellaoui, A.; Agrawal, A.; Albrecht, E.; Alizadeh, B. Z.; Amin, N.; Barnard, J.; Baumeister, S. E.; Benke, K. S.; Bielak, L. F.; Boatman, J. A.; Boyle, P. A.; Davies, G.; De Leeuw, C.; Eklund, N.; Evans, D. S.; Ferhmann, R.; Fischer, K.; Gieger, C.; Gjessing, H. K.; Hägg, S.; Harris, J. R.; Hayward, C.; Holzapfel, C.; Ibrahim-Verbaas, C. A.; Ingelsson, E.; Jacobsson, B.; Joshi, P. K.; Jugessur, A.; Kaakinen, M.; Kanoni, S.; Karjalainen, J.; Kolcic, I.; Kristiansson, K.; Kutalik, Z.; Lahti, J.; Lee, S. H.; Lin, P.; Lind, P. A.; Liu, Y.; Lohman, K.; Loitfelder, M.; McMahon, G.; Vidal, P. M.; Meirelles, O.; Milani, L.; Myhre, R.; Nuotio, M.-L.; Oldmeadow, C. J.; Petrovic, K. E.; Peyrot, W. J.; Polašek, O.; Quaye, L.; Reinmaa, E.; Rice, J. P.; Rizzi, T. S.; Schmidt, H.; Schmidt, R.; Smith, A. V.; Smith, J. A.; Tanaka, T.; Terracciano, A.; Van Der Loos, M. J. H. M.; Vitart, V.; Völzke, H.; Wellmann, J.; Yu, L.; Zhao, W.; Allik, J.; and et al. 2013. GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment. *Science*, 340(6139): 1467–1471.
- Rothstein, M. A. 2018. Time to End the Use of Genetic Test Results in Life Insurance Underwriting. *The Journal of law, medicine & ethics : a journal of the American Society of Law, Medicine & Ethics*, 46(3): 794–801.

- Ruwitch, J. 2025. Judge OKs sale of 23andMe — and its trove of DNA data — to a nonprofit led by its founder. *NPR*.
- Saulsberry, K.; and Terry, S. F. 2013. The Need to Build Trust: A Perspective on Disparities in Genetic Testing. *Genetic Testing and Molecular Biomarkers*, 17(9): 647–648.
- Sciences, C. o. R. t. t. C. R. f. t. P. o. H. S. i. R. i. t. B. a. S.; Board on Behavioral, C.; Statistics, C. o. N.; Population, C. o.; Education, D. o. B. a. S. S.; ; and Council, N. R. 2014. Federal Policy for the Protection of Human Subjects (“Common Rule”). In *Proposed Revisions to the Common Rule for the Protection of Human Subjects in the Behavioral and Social Sciences*. National Academies Press (US).
- Selzam S; Sj, R.; Jb, P.; Ca, R.; Pf, O.; and R, P. 2019. Comparing Within- and Between-Family Polygenic Score Prediction. *American journal of human genetics*, 105(2). Publisher: Am J Hum Genet.
- Shabani, M.; and Marelli, L. 2019. Re-identifiability of genomic data and the GDPR: Assessing the re-identifiability of genomic data in light of the EU General Data Protection Regulation. *EMBO reports*, 20(6): e48316.
- Sklar, E. B. 2020. Be Careful Where you Spit: Do HIPAA-Covered Genetic Tests Actually Provide Greater Privacy Protection to Consumers? *Seton Hall Journal of Legislation and Public Policy*, 44(1).
- Smith-Woolley, E.; Pingault, J.-B.; Selzam, S.; Rimpfeld, K.; Krapohl, E.; von Stumm, S.; Asbury, K.; Dale, P. S.; Young, T.; Allen, R.; Kovas, Y.; and Plomin, R. 2018. Differences in exam performance between pupils attending selective and non-selective schools mirror the genetic differences between them. *npj Science of Learning*, 3(1): 1–7. Publisher: Nature Publishing Group.
- Solove, D. J. 2022. The Limitations of Privacy Rights. *Notre Dame Law Review*, 975.
- Solow-Niederman, A. 2021. Information Privacy and the Inference Economy. *SSRN Electronic Journal*.
- Spiegel, A. M. 2019. THE JEREMIAH METZGER LECTURE: A BRIEF HISTORY OF EUGENICS IN AMERICA: IMPLICATIONS FOR MEDICINE IN THE 21ST CENTURY. *Transactions of the American Clinical and Climatological Association*, 130: 216–234.
- Sterling, R. L. 2011. Genetic Research among the Havasupai: A Cautionary Tale. *AMA Journal of Ethics*, 13(2): 113–117.
- Strand, N. 2016. Shedding Privacy Along With our Genetic Material: What Constitutes Adequate Legal Protection against Surreptitious Genetic Testing? *AMA Journal of Ethics*, 18(3): 264–271.
- Tempus. 2025. Tempus Investor Presentation. Accessed: 2025-01-18.
- U.S. Department of Health & Human Services (HHS). 1979. The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. <https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c-FINAL.pdf>. Accessed: 2025-01-22.
- U.S. House Committee on Oversight and Government Reform. 2025. Comer Launches Investigation into 23andMe Bankruptcy, Seeks Testimony to Protect Americans’ Personal DNA Data. *U.S. House Committee on Oversight and Government Reform*. Press Release, April 15, 2025.
- Veller, C.; and Coop, G. M. 2024. Interpreting population- and family-based genome-wide association studies in the presence of confounding. *PLOS Biology*, 22(4): e3002511.
- Vermont Legislature. 2019. Genetic Testing Underwriting Restrictions Legislation in the States Summary 2019-2022.
- Viljoen, S. 2024. A Relational Theory of Data Governance. *Yale Law Journal*.
- Wan, Z.; Hazel, J. W.; Clayton, E. W.; Vorobeychik, Y.; Kantarcioglu, M.; and Malin, B. A. 2022. Sociotechnical safeguards for genomic data privacy. *Nature Reviews Genetics*, 23(7): 429–445. Publisher: Nature Publishing Group.
- Wetterstrand, K. A. 2023. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). *National Human Genome Research Institute*. Accessed January 2, 2025.
- Yanes, T.; Tiller, J.; Haining, C. M.; Wallingford, C.; Otlowski, M.; Keogh, L.; McInerney-Leo, A.; and Lacaze, P. 2024. Future implications of polygenic risk scores for life insurance underwriting. *npj Genomic Medicine*, 9(1): 25.
- Yang, L.; Xu, S.; Sellergren, A.; Kohlberger, T.; Zhou, Y.; Ktena, I.; Kiraly, A.; Ahmed, F.; Hormozdiari, F.; Jaroensri, T.; Wang, E.; Wulczyn, E.; Jamil, F.; Guidroz, T.; Lau, C.; Qiao, S.; Liu, Y.; Goel, A.; Park, K.; Agharwal, A.; George, N.; Wang, Y.; Tanno, R.; Barrett, D. G. T.; Weng, W.-H.; Mahdavi, S. S.; Saab, K.; Tu, T.; Kalidindi, S. R.; Etemadi, M.; Cuadros, J.; Sorensen, G.; Matias, Y.; Chou, K.; Corrado, G.; Barral, J.; Shetty, S.; Fleet, D.; Eslami, S. M. A.; Tse, D.; Prabhakara, S.; McLean, C.; Steiner, D.; Pilgrim, R.; Kelly, C.; Azizi, S.; and Golden, D. 2024. Advancing Multimodal Medical Capabilities of Gemini. Version Number: 1.
- Zabel, J. 2019. The Killer Inside Us: Law, Ethics, and the Forensic Use of Family Genetics. *Berkeley Journal of Criminal Law*, 24(2): 47.