San Diego Law Review

Volume 52 | Issue 3 Article 7

9-1-2015

Whose Genome Is It Anyway?: Re-identification and Privacy Protection in Public and Participatory Genomics

Sejin Ahn

Whose Genome Is It Anyway?: Re-identification and Privacy Protection in Public and Participatory Genomics

SEJIN AHN*

TABLE OF CONTENTS

---

* © 2015 Sejin Ahn. J.D. Candidate, University of San Diego School of Law, 2017; Ph.D. Genetics, Harvard University, 2010; B.S. Biological Sciences, Seoul National University, 2002. I would like to thank my family and friends for their continued moral support.

I. INTRODUCTION

Xenon Valiant was always curious about his ancestry. He did not know his parents very well because they passed away when he was very young. Xenon discovered on the Internet that he could get his whole genome sequenced[1] and possibly trace his ancestry and disease risks for $10,395,[2] which was out of his reach. While he was debating whether to open a savings account, he saw an Internet ad calling for participants for a whole genome sequencing study.[3] Xenon seized the opportunity to learn about

1. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, PRIVACY AND PROGRESS IN WHOLE GENOME SEQUENCING 22 (Oct. 2012), http://bioethics.gov/sites/default/files/PrivacyProgress508_1.pdf [http://perma.cc/5XXV-DXGJ]. Whole genome sequencing refers to the process of reading the complete deoxyribonucleic acid (DNA) sequence of the organism. *Whole Genome Sequencing (WGS) Program*, U.S. FOOD & DRUG ADMIN., (Mar. 5, 2015), http://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS [http://perma.cc/TFP8-7YQB].

2. *Research Genetics*, GENE BY GENE, https://www.genebygene.com/pages/research?goto=cma [https://perma.cc/8HPJ-E54K] (last visited July 3, 2015) (providing direct-to-consumer whole genome sequencing for $10,395).

3. *See*, *e.g.*, *Clinical Research at Yale*, YALE SCHOOL OF MED., http://www.yalestudies.org/ [http://perma.cc/C6VB-CTEW] (last visited Dec. 18, 2014); *Patient Recruitment at the NIH Clinical Center*, NAT'L INSTS. OF HEALTH CLINICAL CTR., http://www.cc.nih.gov/recruit [http://perma.cc/R9ES-L3H4] (last updated July 24, 2015); *Recruiting Study Subjects*, JOHNS HOPKINS MED. (Feb. 2015), http://www.Hopkinsmedicine.org/institutional_review_board/guidelines_policies/guidelines/study_subject_recruitment.html [http://perma.cc/PJ82-XGLR]. Research institutes and hospitals often recruit human subjects for genomic and other medical studies through traditional print advertisements, websites or through word-of-mouth. *See* Robert Lins, *Clinical Trials: Recruitment Challenges for Proof-of-Concept Viral Challenge Trials*, DRUG DEV. & DELIVERY (Sept. 3, 2014), http://drug-dev.com/Main/Back-Issues/CLINICAL-TRIALS-Recruitment-Challenges-for-Proofof-748.aspx [http://perma.cc/XPP9-JVRC]; Deborah F. Tate et al., *Recruitment of Young Adults into a Randomized Controlled Trial of Weight Gain Prevention: Message Development,*

his ancestry and genetic conditions through the whole genome sequencing study. He consented to publishing his genome online on the OpenGenome.org website,[4] which assured him that it would not post any information identifying him with his genome information.[5] Xenon was delighted with the opportunity to contribute his genome information to the scientific community and find out more about his genetic makeup and ancestry. He was also excited at the prospect of finding out more potential health related information, because he did not have access to health insurance. He shared his experience in a Facebook post. Xenon's genome sequencing revealed he has a rare mutation in the ABCD1 gene, which causes X-linked Adrenoleukodystrophy (X-ALD), a neurological disorder that can manifest itself late into adulthood.[6]

Xenon's brother, Krypton, did not understand why genomics intrigued Xenon. Krypton was a single dad who just wanted to carry on with his job and raise his children. He applied to purchase disability insurance as a safety net for his children. The insurance company hired a genome

Methods, and Cost, 15 TRIALS 326 (2014), http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4150977/pdf/13063_2014_Article_2206.pdf [http://perma.cc/C2UA-5SBQ].

4. OpenGenome.org is a hypothetical open-source public genomics project. Actual open-source public genomics projects include the International HapMap Project, 1000 Genomes, and the Personal Genome Project. *See* INT'L HAPMAP PROJECT, http://hapmap.ncbi.nlm.nih.gov/ [http://perma.cc/GFE2-BMRN] (last visited July 25, 2015); 1000 GENOMES, http://www.1000genomes.org/ [http://perma.cc/8NCN-9V7Q] (last visited July 25, 2015); PERS. GENOME PROJECT: HARV., http://www.personalgenomes.org [http://perma.c/9FYT-TAGB] (last visited July 25, 2015).

5. See *infra* note 129 for examples of personally identifiable information.

6. Marc Engelen et al., *X-Linked Adrenoleukodystrophy (X-ALD): Clinical Presentation and Guidelines for Diagnosis*, *Follow-Up and Management*, 7 ORPHANET J. RARE DISEASES 51, 1–3 (2012). X-ALD is a rare genetic disorder that caused by a mutation in the ABCD1 gene on the X chromosome. *Id.* at 1. The mutation affects the metabolism of lipids. *Id.* In males, who have one X chromosome, the onset, severity and types of symptoms can vary widely. *Id.* at 2–3. Symptoms can range from childhood onset cognitive deficits, auditory impairments, and visual impairments that can lead to death in two to four years, to early onset spastic paraplegia, to slowly progressive myelopathy (disease of the spinal cord), and to devastating cerebral demyelination (damage to the myelin sheath of neurons). *Id.* In females, who have two X chromosomes, heterozygous individuals (individuals with a mutation in one of the two copies of the gene) can be asymptomatic, or develop neurological symptoms around the age of sixty years. *Id.* at 5–6.

hacker[7] to scrutinize Krypton's application.[8] The hacker discovered that Krypton's brother Xenon had a publicly available genome sequence. The

7. There is not a consensus definition of "genome hacking," as people have used the terms in various different contexts. *See*, *e.g.*, Peter Aldhous & Michael Reilly, *Special Investigation: How My Genome Was Hacked*, NEW SCIENTIST, Mar. 2009, at 6 (using the term "genome hacking" to refer to surreptitious genomic testing without the consent of the subject); Damian Counsell, *Hacking the Genome*, LINUX USER, June 2001, at 26, 28 (using the term "genome hacking" to refer to the computational analysis of genomic sequences by providing algorithms and computational tools); Erika Check Hayden, *Privacy Protections: The Genome Hacker*, 497 NATURE 172, 173 (2013) (using the term "genome hacker" to refer to Massachusetts Institute of Technology genomic data scientist Yaniv Erlich who revealed some of the risks of genomic data re-identification by uncovering the identity of some public genomics participants); Karen Hopkin, *Hacking the Genome*, THE SCIENTIST (June 1, 2012), http://www.the-scientist.com/?articles.view/articleNo/32153/title/Hacking-the-Genome/ (using the term "genome hacking" to refer to the study of evolutionary genetics and genome structure). Hackers have attacked and accessed private electronic medical records in the past. Jennifer Dobner, *Fallout Grows from Hacking of Utah Health Database*, REUTERS (Apr. 9, 2012, 11:55 PM), http://www.reuters.com/article/2012/04/10/us-usa-hackers-utah-idUSBRE83904G20120410 [http://perma.cc/9YAF-8S8C]. For example, in 2012, the Utah Department of Health electronic medical records were hacked, leading to approximately 500,000 victims having their sensitive personal information stolen. *Id.* Although the term "genome hacking" has not been used to refer to breach of genomic data or privacy, "genome hacking" or "DNA hacking" would be appropriate terms to refer to such genomic data security breaches or malicious re-identification of genomic data. *See* David Ewing Duncan, *Hacking Your DNA*, NEWSWEEK (Mar. 12, 2014, 12:58 PM), http://www.newsweek.com/2014/03/21/hacking-your-dna-247975.html [http://perma.cc/CX6Y-J9X7].

8. Insurance companies, during the insurance underwriting process, consider many basic factors about the applicant to calculate risks, insurance amount, and benefits. *See, e.g.,* John H. Dodge & David J. Christianson, *Genetic Testing and Disability Insurance: An Alternative Opinion*, 35 J.L. MED. & ETHICS (SPECIAL SUPP.) 33, 33 (2007). Opinions differ as to whether genetic information can be used in underwriting disability and life insurance. *See*, *e.g.*, *id.* at 33–35; Susan M. Wolf & Jeffrey P. Kahn, *Genetic Testing and the Future of Insurance: Ethics*, *Law & Policy*, 35 J.L. MED. & ETHICS (SPECIAL SUPP.) 6, 6 (2007); AM. ACAD. OF ACTUARIES, ISSUE BRIEF: THE USE OF GENETIC INFORMATION IN DISABILITY INCOME AND LONG-TERM CARE INSURANCE (2002), http://www.actuary.org/files/publications/genetic_25apr02.pdf [http://perma.cc/9N44-53FT]; MD. INS. ADMIN., REPORT ON GENETIC INFORMATION AND GENETIC TESTING: INSURANCE AND PERSONALIZED MEDICINE (2009), http://www.mdinsurance.state.md.us/sa/docs/documents/home/reports/report-geneticstesting1-10.pdf [http://perma.cc/K7NN9RL2]. Although there are guidelines about limitations of information gathering during the underwriting process, public records and medical records are often used in the underwriting process. *See*, *e.g.*, *Education Center: Underwriting*, AM. INT'L GROUP, http://www.aig.com/underwriting_3789_536492.html [http://perma.cc/XMV2-T37Y] (last updated Mar. 27, 2014); *NAIC Insurance Information and Privacy Protection Model Act*, NAT'L ASS'N OF INS. COMM'RS (Oct. 1992), http://www.naic.org/store/free/ MDL-670.pdf [http://perma.cc/6GC5-WD9L]. Recently, a website, called Hacker's List, opened, that matches professional hackers with anonymous customers for private hacking service. Matthew Goldstein, *Need Some Espionage Done? Hackers Are for Hire Online*, N.Y. TIMES DEALBOOK (Jan. 15, 2015 9:09 PM), http://dealbook.nytimes.com/2015/01/15/need-some-espionage-done-hackers-are-for-hire-online/?_r=0 [http://perma.cc/4GD8-CFVU]; HACKER'S LIST, https://hackerslist.com/

hacker re-identified[9] Xenon's genome from OpenGenome.org, noted the high likelihood Krypton had for developing X-ALD, and reported back to the insurance company, who denied Krypton's application. Years later, Krypton's children received a large bill for Krypton's care when his neurological symptoms from X-ALD started manifesting.

This is a hypothetical, but plausible scenario.[10] Recent developments in next-generation DNA sequencing[11] and cloud-based data sharing services[12] have led to the rapid advancement of genomics,[13] including many successful

---

[https://perma.cc/J7AT-6ZN3] (last visited Jan. 17, 2015). Although it is unclear whether "genome hackers" are currently available for hire, it is conceivable that they may become available in the future for insurance companies to hire to scrutinize publicly available information to re-identify applicants' information.

9. "Re-identification" refers to the linking the de-identified data (data without any information that can identify the individual source of the data) to the source of the data. Bradley Malin & Latanya Sweeney, *How (Not) To Protect Genomic Data Privacy in a Distributed Network: Using Trail Re-Identification To Evaluate and Design Anonymity Protection Systems*, 37 J. BIOMEDICAL INFORMATICS 179, 180 (2004); Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1703–04 (2009).

10. In 2013, data scientists successfully re-identified individuals' data from public genomics projects. *See* Melissa Gymrek et al., *Identifying Personal Genomes by Surname Inference*, 339 SCIENCE 321, 321–24 (2013). In 2005, a fifteen-year-old boy identified his anonymous sperm donor father through Y chromosome genotyping and internet searches. Alison Motluk, *Anonymous Sperm Donor Traced on Internet*, NEW SCIENTIST (Nov. 3, 2005), http://www.newscientist.com/article/mg18825244.200 [http://perma.cc/S64W-VMK8]. See *infra* Part III.A for discussion of successful re-identification of genetic data.

11. See *infra* Part II.B for a further discussion of next-generation sequencing.

12. Cloud-based data sharing services are web-based storage services where users store and access the data on a host server instead of an individual computer. Jonathan Strickland, *How Cloud Computing Works*, HOW STUFF WORKS, http://computer.howstuffworks.com/cloud-computing/cloud-computing.htm [http://perma.cc/GH8P-XDZF] (last visited Jan. 17, 2015). For examples of cloud-based data sharing services for large-scale genomic data, see *BaseSpace Genomics Cloud Computing*, ILLUMINA, https://basespace.illumina.com/home/index [https://perma.cc/E995-8YTX] (last visited July 25, 2015) (providing storage and applications for analysis of genomic sequencing data performed on Illumina sequencing platforms); *Services for Genomics*, AMAZON WEB SERVICES, http://aws.amazon.com/health/genomics/services/ [http://perma.cc/J82R-H8VH] (last visited July 25, 2015) (providing storage and tools for analysis of high-throughput genomic data).

13. Genomics refers to the study of the genome, including whole genome sequencing and genome-scale analysis of diseases, such as genome-wide association studies (GWAS). *A Brief Guide to Genomics*, NAT'L HUM. GENOME RES. INST., http://www.genome.gov/18016863 [http://perma.cc/YAL3-9UPJ] (last updated Apr. 4, 2015); *Genome-Wide Association Studies*, NAT'L HUM. GENOME RES. INST., http://www.genome.gov/20019523 [http://perma.cc/R4UX-TLHS] (last updated Mar. 31, 2015).

large-scale genotyping and genome sequencing projects.[14] Genomics is a data-driven discipline of science, and the goal of public and participatory genomics projects[15] is to make a large amount of genomic sequence data publicly available and accessible by researchers to foster discoveries.[16] In addition to the reference genome sequenced as part of the initial Human Genome Project,[17] the availability of many more genome-scale data significantly contributed to the society's understanding of human genetics, human evolution and migration, and hereditary diseases.[18]

Publicly available genomic data often excludes information that can identify the subject,[19] preserving anonymity and protecting the participant's privacy rights.[20] However, genomic data inherently contains permanent information that is unique to an individual, and scientists have successfully determined the identity of some individuals whose genomic sequences are publicly available.[21] This poses unique privacy concerns about publicly available genomic information, particularly due to the ease of cross-referencing other public information, such as information on social media

---

14. George M. Church, *Genomes for All*, 294 SCI. AM., Jan. 2006, at 47, 47–48 (2006). Genotyping generally refers to experiments that determine the genotype (genetic information) of an individual. *Genotyping*, GENETICS HOME REFERENCE, http://ghr.nlm.nih.gov/glossary=genotyping [http://perma.cc/SV9B-WQLC] (last visited July 25, 2015). Common large-scale genotyping techniques include single nucleotide polymorphism (SNP) arrays and next-generation sequencing. *See* Ayman Grada & Kate Weinbrecht, *Next-Generation Sequencing: Methodology and Application*, 133 J. INVESTIGATIVE DERMATOLOGY 1, 1–4 (2013); Kevin L. Gunderson et al., *A Genome-Wide Scalable SNP Genotyping Assay Using Microarray Technology*, 37 NATURE GENETICS 549 (2005). "Genome sequencing" refers to the determination of how deoxyribonucleic acid (DNA) base pairs are organized in a string in the organism's genome. *See The Human Genome Project Completion: Frequently Asked Questions*, NAT'L HUM. GENOME RES. INST., http://www.genome.gov/11006943 [http://perma.cc/3DBH-BF9E] (last updated Oct. 30, 2010). For further discussion of genome sequencing projects, see *infra* Parts II.A and II.B.

15. "Public genomics projects" refers to genomics projects that recruit a large number of participants from the public, and often have the data publicly available. John M. Conley et al., *Enabling Responsible Public Genomics*, 20 HEALTH MATRIX 325, 330–32 (2010). For further discussion of public genomics projects, see *infra* Part II.B.

16. Conley et al., *supra* note 15, at 330–32.

17. *See infra* Part II.A for a more detailed discussion on the Human Genome Project.

18. *See*, *e.g.*, Church, *supra* note 14, at 48; Conley et al., *supra* note 15, at 333–34; *About the 1000 Genomes Project*, 1000 GENOMES, http://www.1000genomes.org/about [http://perma.cc/5WDA-VPDA] (last visited July 3, 2015).

19. This process is often called "de-identification" of data. *How Can Covered Entities Use and Disclose Protected Health Information for Research and Comply with the Privacy Rule?*, NAT'L INSTS. OF HEALTH, http://privacyruleandresearch.nih.gov/pr_08.asp [http://perma.cc/U8QJ-KEXE] (last updated Feb. 2, 2007).

20. *Id.* At least in theory, the data should be free of personally identifying information. *Id.*; Ohm, *supra* note 9, at 1716; PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 63–65.

21. This process is often called "re-identification." *See supra* note 9. *See infra* Part III.A for further discussion and examples of re-identification.

and ancestry websites.[22] Future privacy implications are even more complex because the full significance and possible implications of the genomic information are currently unknown.[23] The scientific community and administrative agencies[24] have actively discussed the problems, needs, and possible solutions to balance participants' privacy rights with encouraging participation and information availability.[25] In 2008, Congress passed the Genetic Information Nondiscrimination Act (GINA), which prohibits discrimination based on genetic information in certain situations, to address some of the emerging issues from widely available genomic information.[26] Unfortunately, GINA does not provide privacy protection,[27] and there is currently not a clear solution for the balance.[28]

---

22. Yaniv Erlich & Arvind Narayanan, *Routes for Breaching and Protecting Genetic Privacy*, 15 NATURE REVIEWS GENETICS, 409, 410–11 (2014).

23. *See*, *e.g.*, Conley et al., *supra* note 15, at 344; Erlich & Narayanan, *supra* note 22, at 417.

24. Administrative agencies that participate in and oversee genomic research include the National Institutes of Health (NIH), Centers for Disease Control and Prevention, Department of Energy, Food and Drug Administration (FDA), National Science Foundation, and United States Department of Agriculture. *Other Federal Agencies Involved in Genomics*, NAT'L HUM. GENOME RES. INST., http://www.genome.gov/10003899 [http://perma.cc/J3Z8-Z9XY] (last visited July 25, 2015).

25. *See*, *e.g.*, PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1; Bartha Maria Knoppers, *Framework for Responsible Sharing of Genomic and Health-Related Data*, 8 HUGO J. 1, 3 (2014), http://www.thehugojournal.com/content/8/1/3 [http://perma.cc/TL76-RAAF]; Dina Paltoo et al., Commentary, *Data Use Under the NIH GWAS Data Sharing Policy and Future Directions*, 46 NATURE GENETICS 934 (2014); P3G Consortium et al., *Public Access to Genome-Wide Data: Five Views on Balancing Research with Privacy and Protection*, 5 PLOS GENETICS e1000665, at 1–4 (2009), http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1000665 [http://perma.cc/U5UA-2RSP]; Abdul-Kareem Ahmed, *Unhidden Traits: Genomic Data Privacy Debates Heat Up*, SCI. AM. (Aug. 14, 2013), http://www.scientificamerican.com/article/unhidden-traits-genomic-data-privacy-debates-heat-up/ [http://perma.cc/X5NK-D2C3]; Michelle Meyer, *Online Symposium on the Law, Ethics & Science of Re-identification Demonstrations*, BILL OF HEALTH, HARV. L. PETRIE-FLOM CTR. (May 13, 2013), http://blogs.law.harvard.edu/billofhealth/2013/05/13/online-symposium-on-the-law-ethics-science-of-re-identification-demonstrations/ [http://perma.cc/5MJS-QNBD]; *NIH Issues Finalized Policy on Genomic Data Sharing*, NAT'L INSTS.OF HEALTH (Aug. 27, 2014, 1:00 PM), http://www.nih.gov/news/health/aug2014/od-27.htm [http://perma.cc/ZML4-TLUR].

26. Genetic Information Nondiscrimination Act (GINA) of 2008, Pub. L. No. 110-233, 122 Stat. 881 (codified as amended in scattered sections of 29 U.S.C. & 42 U.S.C.).

27. *Id.*

28. See *supra* note 25 for examples of discussion about balancing participant privacy and scientific information availability.

This Comment advocates for a comprehensive solution to achieve the balance between privacy rights and availability of information. In particular, a strong ban on malicious re-identification and broader anti-discrimination and privacy legislation are necessary to ensure the participants' privacy protection and encourage participation in genomics projects. In addition, the scientific community should establish data standards that can aid in implementation of protective measures to minimize privacy violations.[29] Part II provides an overview of recent developments in genomic technologies and public and participatory genomics. Part III summarizes the privacy issues present in public genomics. Part IV reviews current legislation on genetic information and research participation, including their limitations. Part V proposes a multi-faceted solution, including legislative and research governance solutions to adequately balance participants' privacy with information availability.

II. OVERVIEW OF GENOMICS AND PUBLIC AND PARTICIPATORY GENOMICS

A. Primer on Genetics and Genomics

The genome is the complete set of genetic information of an organism, including the genes and the regions between the genes, stored as deoxyribonucleic acid (DNA).[30] Any two humans' genomes are more

29. This Comment does not address other issues that may be implicated in public genomics, such as intellectual property rights, clinical use of research data and secondary findings, and implication in criminal law, which are reviewed elsewhere. *See*, *e.g.*, PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, ANTICIPATE AND COMMUNICATE: ETHICAL MANAGEMENT OF INCIDENTAL AND SECONDARY FINDINGS IN THE CLINICAL, RESEARCH, AND DIRECT-TO-CONSUMER CONTEXTS 2–19 (2013), http://bioethics.gov/sites/default/files/FINALAnticipateCommunicate_PCSBI_0.pdf [http://perma.cc/Z86A-K84R] (discussing the issue of secondary findings of genomic information); Teneille Brown & Kelly Lowenberg, *Biobanks*, *Privacy*, *and the Subpoena Power*, 1 STAN. J.L. SCI. & POL'Y 88, 89, 101 (2009) (discussing using genetic data in the context of law enforcement); Jose L. Contreras, *Bermuda's Legacy: Policy, Patents, and the Design of the Genome Commons*, 12 MINN. J.L. SCI. & TECH. 61, 111–18 (discussing intellectual property issues).

30. NAT'L HUMAN GENOME RESEARCH INST., NIH PUBLICATION NO. 07-6284, A GUIDE TO YOUR GENOME 1, 3 (2007), http://www.genome.gov/Pages/Education/AllAbouttheHumanGenomeProject/GuidetoYourGenome07_vs2.pdf [http://perma.cc/768U-K5J3]. A human genome contains approximately 3 billion pairs of nucleotides, organized into twenty-three pairs of chromosomes and approximately 20,000 protein coding genes. *Id.* at 3–4; Iakes Ezkurdia et al., *Multiple Evidence Strands Suggest that There May Be as Few as 19,000 Human Protein-Coding Genes*, 23 HUM. MOLECULAR GENETICS 5866, 5872–73 (2014); *Human Assembly and Gene Annotation*, ENSEMBL, http://uswest.ensembl.org/Homo_sapiens/Info/Annotation#assembly [http://perma.cc/X5EU-MRJF] (last visited July 25, 2015). The nucleus of the cell stores fundamental genetic instructions of an organism in DNA form. NAT'L HUM. GENOME RESEARCH INST., *supra* at 4–5. Strings of adenine (A), thymine (T), guanine (G), cytosine (C) DNA nucleotides and their complements encode

than 99% identical, with the less than 1% difference accounting for the expressed physical variations, such as height, hair color, and disease state.[31] DNA sequencing, the determination of how DNA base pairs are organized in a string, is one of the first steps in understanding the relationship between the genotype–genetic information—and the phenotype–physical expression of the genotype.[32]

Early human genetics led to the discovery of classic Mendelian hereditary diseases, and DNA sequencing allowed the identification of the actual causal gene for some of these diseases.[33] Many of the most prevalent diseases, such as cancer and heart disease, are not classic Mendelian diseases and involve multiple genetic and environmental factors, making the study of genetic components of these diseases difficult.[34]

In the past two decades, the discipline of genetics evolved from locating individual genes and analyzing their sequences, to sequencing the entire genome of an organism, and to understanding the organization and

---

the instructions to make the proteins which perform the majority of functions in our bodies. *Id.* at 4. A gene is an individual organized unit of information that can encode for a protein. *Id.* at 4. A gene can also encode other functional molecules such as a non-coding ribonucleic acid (RNA), but the majority of genes encode proteins. Mark B. Gerstein et al., *What is a Gene*, *Post-ENCODE? History and Updated Definition*, 17 GENOME RES. 669 (2007).

31. Lance W. Hahn et al., *Multifactor Dimensionality Reduction Software for Detecting Gene-Gene and Gene-Environment Interactions*, 19 BIOINFORMATICS 376, 376 (2003); *International HapMap Project*, NAT'L HUM. GENOME RES. INST., http://www.genome.gov/10001688 [http://perma.cc/4SZB-TBJ3] (last updated May 1, 2012).

32. NAT'L HUMAN GENOME RESEARCH INST., *supra* note 30, at 3; NCI-NHGRI Working Grp. on Replication in Ass'n Studies, *Replicating Genotype–Phenotype Associations*, 447 NATURE 655, 655 (2007).

33. Mendelian hereditary diseases, named after the genetic inheritance patterns discovered by Gregor Mendel, are caused by a single gene mutation. Gerstein et al., *supra* note 30, at 669–70; Heidi Chial, *Mendelian Genetics: Patterns of Inheritance and Single-Gene Disorders*, SCITABLE BY NATURE EDUC., http://www.nature.com/scitable/topicpage/Mendelian-Genetics-Patterns-of-Inheritance-and-Single-966 [http://perma.cc/26L8-QVYP] (last visited July 25, 2015). One example of a disease with Mendelian inheritance disease is sickle cell anemia, which occurs due to a single substitution of an A nucleotide with a T nucleotide in the β-globin gene. Allison Ashley-Koch et al., *Sickle Hemoglobin (Hb S) Allele and Sickle Cell Disease: A HuGE Review*, 151 AM. J. OF EPIDEMIOLOGY 839, 839–45 (2000).

34. Hahn et al., *supra* note 31, at 376. Many common diseases involve a complex gene-gene interaction (interaction of multiple genetic factors) and gene-environment interaction (interaction of genetic factors and environmental factors). *Id.* Epigenetic factors, which are inherited changes that are not encoded in the DNA, and the environment, also contribute to expressed physical variations. Gerstein et al., *supra* note 30, at 672.

interaction between genes within the context of the genome.[35] Human genetics is a discipline based on statistics, so the availability of a large amount of genomic data is crucial to understanding the human genetic history and human diseases.[36]

B. Public and Participatory Genomics

The first public genomics project, the Human Genome Project (HGP), started in 1990.[37] The HGP began as an international scientific research consortium, whose participants included the National Institutes of Health (NIH), the U.S. Department of Energy, and universities and research institutes around the world.[38] At the same time, a parallel sequencing project was underway at a private corporation, Celera Genomics.[39] The HGP and Celera Genomics published two draft reference genomes in 2001,[40] and the HGP completed the initial reference human genome in

35. *Frequently Asked Questions About Genetic and Genomic Science*, NAT'L HUM. GENOME RES. INST., http://www.genome.gov/19016904 [http://perma.cc/3B3W-XPFX] (last updated Feb. 14, 2014).

36. *Genome-Wide Association Studies*, *supra* note 13. Determination and aggregation of phenotypes (physically expressed traits) is also necessary to fully understand the implication of the genetic data. *Id.* Linking the genotypic data to the phenotypic data allows scientists to determine the association between a specific mutation and a phenotype. *Id.* An example of these types of studies is genome-wide association studies (GWAS). *Id.*

37. *The Human Genome Project Completion: Frequently Asked Questions, supra* note 14.

38. *Id.*

39. COMM. ON INTELLECTUAL PROP. RIGHTS IN GENOMIC & PROTEIN RESEARCH AND INNOVATION, NAT'L RESEARCH COUNCIL OF THE NAT'L ACADS., REAPING THE BENEFITS OF GENOMIC AND PROTEOMIC RESEARCH: INTELLECTUAL PROPERTY RIGHTS, INNOVATION, AND PUBLIC HEALTH 36 (Stephen A. Merrill & Anne-Marie Mazza eds., 2006). Celera Genomics was founded as a commercial alternative to generating and commercializing genomic information. *About Us*, CELERA, https://www.celera.com/celera/history [https://perma.cc/E64H-F9DT] (last visited July 25, 2015). During the initial genomics era, the HGP and Celera were competing to complete the human genome, and Celera's shotgun approach, sequencing shorter fragments and assembling them later, tended to be quicker, and the HGP later changed their approach to shotgun sequencing as well. NAT'L RESEARCH COUNCIL OF THE NAT'L ACADS., *supra*; J. Craig Venter et al., *The Sequence of the Human Genome*, 291 SCIENCE 1304, 1305–06 (2001), http://www.sciencemag.org/content/291/5507/1304.full.pdf [http://perma.cc/X8AJ-X6RK].

40. Editorial, *E Pluribus Unum*, 7 NATURE METHODS 331, 331 (2010). A representative sample of five humans were the subjects in the initial draft reference genome. Int'l Human Genome Sequencing Consortium, *Initial Sequencing and Analysis of the Human Genome*, 409 NATURE 860, 860 (2001), http://www.nature.com/nature/journal/v409/n6822/pdf/409860a0.pdf [http://perma.cc/YR99-FXRW]; Venter et al., *supra* note 39, at 1305. The draft genome had around 150,000 gaps, including gaps in specific small gap regions that are difficult to sequence, such as repeats, in particular structural regions such as centromeres and heterochromatic regions, or in regions with high diversity. Int'l Human Genome

2003.[41] Since the HGP's completion, DNA genotyping[42] and sequencing technologies have advanced tremendously, reducing the time and cost of obtaining genomic data and increasing the amount of available sequences.[43]

One of the earliest genome-scale genotyping technologies that has become widely available is single-nucleotide polymorphism (SNP) genotyping.[44] SNPs are single base pair differences between individuals' DNA and constitute the small differences between individuals.[45] There are approximately ten million SNPs in the human genome, occurring roughly once every 300 base pairs.[46] Although the variations constitute less than 1% of the genome, their role as landmarks can help capture most of the genetic variation between individuals.[47] These landmarks are used

---

Sequencing Consortium, *supra* at 874; Venter et al., *supra* note 39, at 1311; Editorial, *supra*.

41. *International Consortium Completes Human Genome Project: All Goals Achieved; New Vision for Genome Research Unveiled*, NAT'L HUM. GENOME RES. INST. (April 14, 2003), https://www.genome.gov/11006929 [https://perma.cc/B9V3-53ZQ]. The initial "completion" of the human reference genome covered approximately 99% of the human genome (excluding heterochromatic regions), and reduced the number of gaps to 341. Int'l Human Genome Sequencing Consortium, *Finishing the Euchromatic Sequence of the Human Genome,* 431 NATURE 931, 931–45 (2004), http://www.nature.com/nature/journal/v431/n7011/pdf/nature03001.pdf [http://perma.cc/3W9U-3ZY8]. Newer builds of the human reference genome, in which scientists are trying to determine the sequence of the gaps, continues to be assembled and released. Vivien Marx, *A Star Is Born: the Updated Human Reference Genome*, NATURE METHODS: METHAGORA (Dec. 24, 2013, 12:19 PM), http://blogs.nature.com/methagora/2013/12/the_updated_human_reference_genome.html [http://perma.cc/BG2Q-53NA]. A recently released build, Genome Reference Consortium build 38, included sequences of some of the gaps in the centromere. *Id.*

42. For an explanation of genotyping, see *supra* note 14.

43. *DNA Sequencing Costs*, NAT'L HUM. GENOME RES. INST., http://www.genome.gov/sequencingcosts/ [http://perma.cc/2MA5-7VMA] (last updated June 15, 2015).

44. Ann-Christine Syvänen, *Toward Genome-Wide SNP Genotyping*, 37 NATURE GENETICS S5, S5 (Supp. 2005), http://www.nature.com/ng/journal/v37/n6s/pdf/ng1558.pdf [http://perma.cc/42B8-UHKB]. Microarray chip-, bead- or sequencing-based genomic scale SNP genotyping can be used to distinguish SNPs on the genomic scale, up to millions of SNPs. *Id.*; Gunderson et al., *supra* note 14, at 549–54. For the purposes of this paper, "public genomics data" include both whole genome sequencing and genomic scale SNP.

45. *What Are Single Nucleotide Polymorphisms (SNPs)?*, GENETICS HOME REFERENCE, http://ghr.nlm.nih.gov/handbook/genomicresearch/snp [http://perma.cc/9L3H-YVBG] (last visited July 25, 2015).

46. *Id.* SNPs are typically interspersed throughout the genome in protein coding genes and in regions between them. *Id.*

47. *SNP*, BROAD INST., https://www.broadinstitute.org/education/glossary/snp [https://perma.cc/4JWE-C3C2] (last visited Mar. 28, 2015). SNPs can be used as landmarks of inheritance because genetic variations that are physically close to each other on the chromosome are more likely to be inherited together than variations that are located far

in genome-wide association studies (GWAS) to identify SNPs that occur more frequently in people with a particular physical trait, such as a disease, than in people without.[48] Inexpensive genome-wide SNP genotyping became the basis of direct-to-consumer genomics, with companies such as 23andMe, deCODE Genetics, Navigenics and Pathway Genomics providing SNP genotyping information to private consumers, who pay as little as $99 for their ancestry and disease-related information.[49] One popular direct-to-consumer genomics company, 23andMe, temporarily suspended providing health-related customer reports, due to Food and Drug Administration approval problems[50] but continues to provide and

---

apart. *Id.* Some of the landmark SNPs can be used to deduce the nearby variations. *See* Eric S. Lander, *Initial Impact of the Sequencing of the Human Genome*, 470 NATURE 187, 190–91 (2011); *What Is the HapMap?* INT'L HAPMAP PROJECT, http://hapmap.ncbi.nlm.nih.gov/whatishapmap.html.en [http://perma.cc/L7YL-8SAW] (last visited Mar. 28, 2015).

48. *Genome-Wide Association Studies*, *supra* note 13. For example, in GWAS, SNP information (genotype) and phenotype information, such as having a particular disease of interest, from many individuals are aggregated, to identify regions of the genome that may be linked to the phenotype. *Id.* A set of common genetic variants are compared in individuals with a condition, such as a disease being studied, and control individuals who do not have the condition. *Id.* With a large number of SNPs and sample size, scientists can narrow particular genomic regions by the landmark SNPs that may be associated with the specific trait. *See* Lander, *supra* note 47; Syvänen, *supra* note 44, at S5.

49. 23ANDME, https://www.23andme.com/ [https://perma.cc/5GAE-9BKH] (last visited Mar. 28, 2015); PATHWAY GENOMICS, https://www.pathway.com/ [https://perma.cc/Q34W-68RP] (last visited Mar. 28, 2015); *see* Valerie Gutmann Koch, *PGTandMe: Social Networking-Based Genetic Testing and the Evolving Research Model*, 22 HEALTH MATRIX 33, 36–37 (2012) (providing examples of direct-to-consumer genomics companies).

50. *See* Letter from Alberto Gutierrez, Dir., Office of In Vitro Diagnostics & Radiological Health, FDA, to Ann Wojcicki, CEO, 23andMe, Inc. (Nov. 22, 2013), http://www.fda.gov/iceci/enforcementactions/warningletters/2013/ucm376296.htm [http://perma.cc/G6D8-AJ28]. The FDA recently issued a Warning Letter to 23andMe and other DTS genomics companies making disease and health-related claims without proper FDA approval:

> [Y]ou are marketing the 23andMe Saliva Collection Kit and Personal Genome Service (PGS) without marketing clearance or approval in violation of the Federal Food, Drug and Cosmetic Act (the FD&C Act). This product is a device within the meaning of section 201(h) of the FD&C Act, 21 U.S.C. 321(h), because it is intended for use in the diagnosis of disease or other conditions or in the cure, mitigation, treatment, or prevention of disease, or is intended to affect the structure or function of the body.

Id. (citing 21 U.S.C. § 321(h) (2006)). After the warning, 23andMe has temporarily discontinued providing reports regarding health related information, but FDA subsequently issued authorization to provide reports for one disease. *Status of Our Health-Related Genetic Reports*, 23ANDME, https://www.23andme.com/health/ [https://perma.cc/J3NE-C3EP] (last visited June 28, 2015) ("In February, 2015, 23andMe was granted authorization by the U.S. Food and Drug Administration (FDA) to market the Bloom syndrome carrier status report. This is an important first step in fulfilling our commitment to return genetic health reports to consumers. . . . At this time, we do not know which health reports might be available or when they might be available."). *See also*, *infra* note 51.

store ancestry information.[51]

Biotechnology companies such as Illumina, Life Technologies, and Roche developed several sequencing technology platforms, collectively termed next-generation sequencing technologies, which significantly increased the amount of sequence data output for human genetics.[52] Next-generation sequencing typically uses shorter read lengths,[53] but uses a substantially higher number of overlapping reads to assemble the sequences computationally.[54] This exponentially lowered the cost and time required to read large stretches of DNA, from approximately $20 per base pair in 1990[55] to less than $0.10 per million base pairs in 2014.[56] In 2012, the average cost of whole genome sequencing of a human-sized genome–three billion base pairs was less than $10,000,[57] making the $1,000 genome within reach.[58]

---

51. *Bring Your Ancestry to Life Through Your DNA*, 23ANDME, https://www.23andme.com/ancestry/ [https://perma.cc/3LJE-7KVL] (last visited Mar. 28, 2015). 23andMe since has submitted one disease-related marker test for FDA approval, and was granted an approval. Robert Hof, *Seven Months After FDA Slapdown*, *23andMe Returns with New Health Report Submission*, FORBES (June 20, 2014, 9:04 AM), http://www.forbes.com/sites/roberthof/2014/06/20/seven-months-after-fda-slapdown-23andme-returns-with-new-health-report-submission [http://perma.cc/6UW9-SV8S]; *Status of Our Health-Related Genetic Reports*, *supra* note 50. However, 23andMe currently only provides ancestry reports and uninterpreted raw genetic data without health related information. *Id.*

52. Grada & Weinbrecht, *supra* note 14, at 1–2; *see* Elaine R. Mardis, *The Impact of Next-Generation Sequencing Technology on Genetics*, 24 TRENDS GENETICS 133, 133–35 (2008) (surveying Roche's and Illumina's next-generation instruments).

53. Read length is the length of DNA that can be read at once. Rob Carlson, *How Competition Improves DNA Sequencing*, SYNTHESIS (Apr. 23, 2013, 2:36 PM), http://www.synthesis.cc/2013/04/how-competition-improves-reading-dna.html [http://perma.cc/S8SL-LUBX].

54. *See* Grada & Weinbrecht, *supra* note 14, at 1, 2.

55. Rob Carlson, *Time for New DNA Synthesis and Sequencing Cost Curves*, SYNTHESIS (Feb. 12, 2014, 2:15 PM), http://www.synthesis.cc/2014/02/time-for-new-cost-curves-2014.html [http://perma.cc/UJ5U-G9PJ].

56. *DNA Sequencing Costs*, *supra* note 43.

57. This estimate accounts for additional direct and indirect costs and increased coverage requirements for long reads. *Id.* *See also The Human Genome Project Completion: Frequently Asked Questions, supra* note 14.

58. Church, *supra* note 14, at 47–48. "$1,000 genome" became a symbolic goal in DNA sequencing technology. *Id.* The goal is to make DNA sequencing so affordable that individuals and their doctors can easily use sequence information to understand variations and make healthcare decisions. *See id.* Currently, direct-to-consumer whole genome sequencing is available from one company, Gene by Gene, at a price of $10,395. GENE BY GENE, *supra* note 2.

Several large-scale public genomics projects emerged, backed by the increasing availability of large scale data and the growing popularity of direct-to-consumer genomics.[59] As more genomic data become available, the research community seeks access to the already sequenced genomes, to maximize the utility of existing data.[60] Increasing genome-scale DNA data accessibility can augment the statistical power necessary to understand the link between genetic variations and phenotypes.[61]

Public genomics projects and initiatives are international collaborative efforts to collect genome-scale information and the phenotypic information needed to elucidate the complex interplay between the genetic and environmental components of phenotypes.[62] These initiatives allow comparison of sequencing results completed around the world and maximize the obtained data's utility by allowing other researchers to mine the genomic data for additional connections to physical traits.[63] The International HapMap project, one of the earlier international consortia, publicly released genome-scale haplotype maps from SNP genotyping of different populations throughout the world.[64] The 1000 Genomes project, built upon the International HapMap project, aims to make whole genome sequencing data from about 2,500 individuals from twenty-five populations publicly

---

59. *See supra* note 4.

60. *See* Church, *supra* note 14. A comparison of many genotypes and phenotypes is necessary to understand how certain genes function. NAT'L HUMAN GENOME RESEARCH INST., *supra* note 13. The focus of genetics has moved from small scale approaches to a larger scale "systems" approach, making experiments much more expensive to perform. *See* Church, *supra* note 14; NAT'L HUMAN GENOME RESEARCH INST., *supra* note 13. Since whole genome sequencing and other techniques generate data that can be mined for additional purposes, making this data accessible to a wider population can help maximize the utility of the data, especially in the era of much more competitive and diminishing science research funding. *See*, *e.g.*, Conley et al., *supra* note 15; Pak C. Sham & Shaun M. Purcell, *Statistical Power and Significance Testing in Large-Scale Genetic Studies*, 15 NATURE REVIEWS GENETICS 335 *passim* (2014); *Frequently Asked Questions: Data Sharing*, NAT'L INSTS. OF HEALTH, http://grants.nih.gov/grants/policy/data_sharing/data_sharing_faqs.htm [http://perma.cc/H4G5-CFHZ] (last revised Feb. 16, 2004).

61. *See* Sham & Purcell, *supra* note 60. Typically, for complex genetic traits, a large sample size is required to achieve a statistically significant association between a genotype and a phenotype. *See id.* Making previously obtained data available allows scientists to use the data to increase sample size and obtain the statistical power necessary for significant associations. *See id.* Public and open-sourced genomics projects contributed to a significant increase in the availability of genomic and phenotypic data necessary for statistical analysis in human genetics. Conley et al., *supra* note 15; Lander, *supra* note 47.

62. *See* Conley et al., *supra* note 15, at 330–35.

63. *See id.*; Jeanne Erdmann, *As Personal Genomes Join Big Data Will Privacy and Access Shrink?*, 20 CHEMISTRY & BIOLOGY 1, 1–2 (2013); Sham & Purcell, *supra* note 60; *Frequently Asked Questions: Data Sharing*, *supra* note 60.

64. *About the HapMap*, INT'L HAPMAP PROJECT, http://hapmap.ncbi.nlm.nih.gov/thehapmap.html.en [http://perma.cc/PPS3-DB5E] (last visited Apr. 12, 2015).

available for the researchers worldwide to use.[65] The result of its pilot project are available as a public data set.[66] The Personal Genome Project (PGP) is a project that requests participants to make their genomic data openly available.[67]

Popularity of direct-to-consumer genomics also led to the launch of crowd-sourced genomics projects.[68] Often called participatory genomics, these projects use Internet databases to identify research populations, recruit participants, and collect genomic data.[69] One such project, openSNP,

---

65. 1000 GENOMES, *supra* note 4. *See also About the 1000 Genomes Project*, *supra* note 18.

66. *See* 1000 GENOMES, *supra* note 4. The data set is available via Amazon Web Services. *1000 Genomes Project and AWS*, AMAZON WEB SERVICES, http://aws.amazon.com/1000genomes/ [http://perma.cc/3G6B-4UGV] (last visited July 25, 2015).

67. PERS. GENOME PROJECT: HARV., *supra* note 4. The PGP accepts both whole genome sequencing and SNP-based data. *See Data & Samples*, PERS. GENOME PROJECT: HARV., http://www.personalgenomes.org/harvard/data [http://perma.cc/3Y2P-EV48] (last visited Apr. 13, 2015).

68. *See, e.g.*, Firas Khatib et al., *Crystal Structure of a Monomeric Retroviral Protease Solved by Protein Folding Game Players*, 18 NATURE STRUCTURAL & MOLECULAR BIOLOGY 1175, 1175 (2011); Robert M. Plenge et al., *Crowdsourcing Genetic Prediction of Clinical Utility in the Rheumatoid Arthritis Responder Challe*nge, 45 NATURE GENETICS 468, 468 (2013); Melanie Swan, *Crowd-sourced Health Research Studies: An Important Emerging Complement to Clinical Trials in the Public Health Research Ecosystem*, 14 J. MED. INTERNET RES. e46 *passim* (2012); *Is Crowdsourcing the Future of Scientific Research?,* MICH. ST. U. (May 15, 2014), http://msutoday.msu.edu/news/2014/is-crowdsourcing-the-future-of-scientific-research/ [http://perma.cc/SP6R-RVPL]. "Crowd-sourced studies" refers to studies that use "a large, often varied or undefined group or population to undertake a defined task." Dan Vorhaus, *Crowd-Sourcing vs. Open-Sourcing in Consumer Genomics*, GENOMICS L. REP. (Aug. 25, 2009), http://www.genomicslawreport.com/index.php/2009/08/25/crowd-sourcing-vs-open-sourcing-in-consumer-genomics/ [http://perma.cc/Z4K3-3EJ2]. For genomic research, crowd-sourcing may involve "using web-driven or other distributed modes of interaction to identify research populations, recruit participants and, ultimately, collect the data necessary to produce meaningful scientific research." *Id.*

69. For example, consumers who have data from direct-to-consumer services can upload their data to these projects. Vorhaus, *supra* note 68. For other examples of crowd-sourced genomics or medical projects, see DIYGENOMICS, http://www.diygenomics.org/ [http://perma.cc/57KD-4M74] (last visited July 25, 2015); GENOMERA, http://genomera.com/ [http://perma.cc/HS32-KKUH] (last visited July 25, 2015); OPENSNP, https://opensnp.org/ [https://perma.cc/9EJH-QPCM] (last visited July 25, 2015); PATIENTS LIKEME, http://www.patientslikeme.com/ [http://perma.cc/RHC6-8DVS] (last visited July 25, 2015); QUANTIFIED SELF, http://quantifiedself.com/ [http://perma.cc/V3VG-F69D] (last visited July 25, 2015). Some participatory genomics or medical projects have already produced results. *See, e.g.*, Adam A. Margolin et al., *Systematic Analysis of Challenge-Driven Improvements in Molecular Prognostic Models for Breast Cancer*, 5 SCI. TRANSLATIONAL MED. 181re1, 1–2 (2013) (reporting that a crowd-sourced genomic data

encourages direct-to-consumer genomics consumers to publish their test results for both scientific research and to obtain additional information.[70] For example, participants can find others with similar genetic variations and access scientific literature about their genotypes.[71]

DNA-based ancestry tracing is also popular.[72] Some of the DNA-based genealogy companies provide family finder databases that can assist in locating relatives and building a family tree.[73] Once considered a hobby for a select group, genealogy became a fast-growing industry, aided by direct-to-consumer genomics and public genealogy databases.[74] Searchable public DNA-based genealogy databases provide a significant expansion of public genealogical information, in addition to the numerous non-DNA based genealogy sites readily accessible by the public.[75]

III. PRIVACY ISSUES IN PUBLIC AND PARTICIPATORY GENOMICS

A. Re-identification: Cracking Anonymized Data

Even though most public genomics projects exclude identifiable information from genomic data, the risk of privacy breach by re-identification

---

analysis challenge by Sage Bionetworks resulted in a more accurate prognostic model for breast cancer outcomes.)

70. *See, e.g.*, Bastian Greshake et al., *openSNP–A Crowd-sourced Web Resource for Personal Genomics*, 9 PLOS ONE e89204 (2014), http://www.plosone.org/article/fetchObject.action?uri=info:doi/10.1371/journal.pone.0089204&representation=PDF[http://perma.cc/XB58-36JV]; OPENSNP, *supra* note 69.

71. Greshake et al., *supra* note 70; OPENSNP, *supra* note 69. In this sense, openSNP is both an open-source and crowd-sourced project. *Id.*; Vorhaus, *supra* note 68.

72. *See infra* note 74.

73. ANCESTRYDNA, http://dna.ancestry.com/ [http://perma.cc/5NQY-YFNN] (last visited July 27, 2015); FAM. TREE DNA, https://www.familytreedna.com/ [https://perma.cc/5S2V-FWVW] (last visited July 27, 2015); 23ANDME, *supra* note 49.

74. Alan Farnham, *Who's Your Daddy? Genealogy Becomes $1.6B Hobby*, ABC NEWS (Oct. 24, 2012), http://abcnews.go.com/Business/genealogy-hot-hobby-worth-16b-mormons/story?id=17544242 [http://perma.cc/FH5B-6P83]; Gregory Rodriguez, *How Genealogy Became Almost as Popular as Porn*, TIME.COM (May 30, 2014), http://time.com/133811/how-genealogy-became-almost-as-popular-as-porn/ [http://perma.cc/5U2P-9N5X]. Although certain cultures or religious groups have considered genealogy to be important in the past, genealogy was typically limited to those groups or some hobbyists before the advent of the internet in the 1990s. Rodriguez, *supra.* The current popularity of genealogy is exemplified by the fact that Ancestry.com, a popular genealogy service website, had over 2 million paid subscribers and approximately a billion dollars in revenue for 2012. Farnham, *supra.*

75. *See*, *e.g.*, *infra* note 85. *See also Top 100 Genealogy Websites for 2014*, GENEALOGYINTIME MAG., http://www.genealogyintime.com/articles/top-100-genealogy-websites-of-2014-page02.html [http://perma.cc/L5WS-DMPZ] (last visited July 27, 2015) (listing the top 100 most frequently visited genealogy websites in 2014).

still remains.[76] Re-identification studies revealed the privacy risk of public medical information.[77] In 1997, Latanya Sweeney, then a computer science graduate student, was able to identify the health record of the then Massachusetts governor William Weld.[78] Health records publicly released by Massachusetts contained the birthdate, sex, and zip code of the patients.[79] Although the governor was a well-known figure with a publicized hospitalization, which likely made re-identification easier,[80] Sweeney and others' subsequent research showed that an individual can be uniquely identified with relatively few items of information.[81] Re-identification research highlighted the privacy risk of public information containing personally identifying information, influencing the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule.[82]

Genome scientists at MIT recently published a prominent example of re-identification from publicly available genomic data.[83] They cross-referenced Y chromosome data from public genomics projects to public, searchable surname-based genealogy databases that also allow users to enter Y chromosome data.[84] They uniquely identified the surnames of some of the individuals, and in five out of the ten genomes with the most complete Y chromosome data, they successfully identified the individuals

76. Russ B. Altman et al., *Data Re-identification: Societal Safeguards*, 339 SCIENCE 1032, 1032–33 (2013). *See also supra* note 9.

77. *See*, *e.g.*, Gymrek et al., *supra* note 10; Jonathan Shaw, *Exposed: The Erosion of Privacy in the Internet Era*, HARV. MAG., Sept.-Oct. 2009, at 39–40; Latanya Sweeney, *k-anonymity: A Model for Protecting Privacy*, 10 INT'L J. UNCERTAINTY, FUZZINESS & KNOWLEDGE-BASED SYSTEMS 557, 557 (2002).

78. Sweeney, *supra* note 77, at 559.

79. *Id*. at 558–59.

80. Daniel C. Barth-Jones, The "Re-identification" of Governor William Weld's Medical Information: A Critical Re-examination of Health Data Identification Risks and Privacy Protections, Then and Now (July 24, 2012) (unpublished manuscript), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2076397 [http://perma.cc/RWH4-2BFJ].

81. *See*, *e.g.*, Malin & Sweeney, *supra* note 9.

82. Ohm, *supra* note 9, at 1737; Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, 45 C.F.R. §§ 160, 164 (2012). Sweeney's research influenced the HIPAA Privacy Rule, to limit birth dates only to years and ZIP code only to the first three digits, or the first two digits for ZIP codes with populations 20,000 or less. Ohm, *supra* note 9, at 1737; 45 C.F.R. § 164.514(b)(2).

83. Gymrek et al., *supra* note 10.

84. Gymrek et al., *supra* note 10, at 321–22. The scientists used publicly available whole genome sequencing data to infer the genotype of short repeated sequences in the Y chromosome that can serve as a hereditary fingerprint. *Id.* at 321.

and their families.[85] This study demonstrated that identification of an individual based on purely public information is possible with relative ease.[86] De-identification of genomic information by removing HIPAA identifiers does not guarantee the removal of all identifiable information because information inherent in the DNA can be retrieved by relatively simple processing.[87] In addition, the amount of data an adversary[88] can potentially access increases as people voluntarily make increasingly more personal information available online.[89]

B. Risk of Privacy Breach for Family and Relatives

Because genes are inherited, there is an additional privacy risk to close relatives of individuals with publicly available genomic information.[90] In 2013, scientists briefly published online the genome of Henrietta Lacks,

---

85. Gymrek et al., *supra* note 10, at 323. They used Y chromosome genotyping data available on the genealogy websites. *Id.*; YSEARCH, http://www.ysearch.org/ [http://perma.cc/5EGT-UQ46] (last visited July 27, 2015) (providing free public search based on Y chromosome genotyping data by FamilyTreeDNA).

86. Gymrek et al., *supra* note 10. Each complete pedigree identification took 3 to 7 hours by a trained individual. *Id.* at 323. Although the success rate was not high and this specific technique can only be used to identify males, the study demonstrated the proof-of-principle that re-identification is easily possible for persons without a highly public profile. *Id.*

87. Mark A. Rothstein, *Is Deidentification Sufficient to Protect Health Privacy in Research?*, 10 AM. J. BIOETHICS 3, 5–6 (2010).

88. In data science and data security, an adversary is "an individual, group, organization, or government that conducts or has the intent to conduct detrimental activities." *Explore Terms: A Glossary of Common Cybersecurity Terminology*, NAT'L INITIATIVE FOR CYBERSECURITY CAREERS & STUD., DEP'T HOMELAND SECURITY, http://niccs.us-cert.gov/glossary#adversary [http://perma.cc/A3CT-YR94] (last visited Jan. 17, 2015). In this context, detrimental activities can refer to an attack on the privacy of persons whose data is stored electronically. Tim Matthews, *Anatomy of a Data Breach*, INT'L ASS'N OF PRIVACY PROF'LS (2011), https://www.privacyassociation.org/media/presentations/12Summit/S12_Anatomy_of_a_Data_Breach_PPT.pdf [https://perma.cc/P7BW-GXYV]. For example, an adversary can cause a security breach by accessing data that they should not have access to. *Id.* In other cases, the adversary can cause a privacy breach by re-identifying publicly available information. Ohm, *supra* note 9, at 1707–08. Data security scientists often taken on the role as adversaries to examine the risk of these breaches. *See*, *e.g.*, Arvind Narayanan & Vitaly Shmatikov, *Robust De-Anonymization of Large Sparse Datasets*, *in* PROCEEDINGS OF THE 2008 IEEE SYMPOSIUM ON SECURITY & PRIVACY 111 (2008); Gymrek et al., *supra* note 10.

89. Erlich & Narayanan, *supra* note 22, at 411. For example, people voluntarily submit and make available their personal information in many social networking sites. *Id.*

90. *See, e.g.*, Mathias Humbert et al., *Addressing the Concerns of the Lacks Family: Quantification of Kin Genomic Privacy*, *in* PROCEEDINGS OF 2013 ACM SIGSAC CONFERENCE ON COMPUTER & COMMUNICATIONS SECURITY 1141 (2013). For example, significant parts of a person's genome can be deduced from a close relative's genome. Christopher A. Cassa et al., *My Sister's Keeper?: Genomic Research and the Identifiability of Siblings*, 1 BMC MED. GENOMICS 32, 32 (2008).

the woman from whom the widely used HeLa cell lines[91] originated, without the consent of her relatives.[92] The widely publicized case highlighted another important aspect of privacy in public genomics: a public genome can reveal information about the participant's relatives.[93] Although the scientists took the data off-line soon thereafter, at least fifteen people had already downloaded it and one was able to upload it onto a website called SNPedia to obtain a literature summary report about Henrietta Lacks and her family.[94]

Similar re-identification issues can also often surface in the form of family issues, such as paternity.[95] For example, in 2005, a 15-year-old boy was able to identify his anonymous sperm donor father by genotyping of his own Y chromosome and accessing paid online databases of birthplaces and birthdates.[96] The boy identified his biological father without access to the father's genotype, through publicly available genotypes of his potential biological male relatives.[97]

---

91. HeLa cells are one of the most widely used isolated cultured human cell lines in biological research. Rebecca Skloot, *The Immortal Life of Henrietta Lacks*, *the Sequel*, N.Y. TIMES (May 23, 2013), http://www.nytimes.com/2013/03/24/opinion/sunday/the-immortal-life-of-henrietta-lacks-the-sequel.html [http://perma.cc/V5JF-F2CC]. The name comes from the first two letters of Henrietta Lacks' first and last names. Wynne Parry, *Controversial 'Hela' Cells: Use Restricted Under New Plan*, LIVESCIENCE (Aug. 07, 2013, 12:57 PM) http://www.livescience.com/38728-hela-cells-restricted-new-nih-plan.html [http://perma.cc/3M6E-FVVY].

92. Skloot, *supra* note 91. Ms. Lacks died in 1951 and was unable to personally provide consent. *Id*. Although some of her relatives became known to the public in the recent years, they did not provide consent to publishing Ms. Lacks's genome. *Id.*

93. *Id.*

94. *Id.* SNPedia is a wiki website that collects information about different SNP variants and information related to the SNPs. SNPEDIA, http://www.snpedia.com/index.php/SNPedia [http://perma.cc/DD2A-TGJT] (last modified June 10, 2015). Uploading data onto SNPedia allows the generation of a literature summary report, called Promethease that generates a scientific literature report associated with the collection of SNPs that were entered. *Promethease*, SNPEDIA, http://www.snpedia.com/index.php/Promethease [http://perma.cc/WRA2-JE5A] (last modified Aug. 18, 2015); Skloot, *supra* note 91.

95. *See*, *e.g.*, Motluk, *supra* note 10 (describing the steps a 15-year-old boy took to identify his anonymous sperm donor father); George Doe, *With Genetic Testing*, *I Gave My Parents the Gift of Divorce*, VOX (Sept. 9, 2014, 7:50 AM), http://www.vox.com/2014/9/9/5975653/with-genetic-testing-i-gave-my-parents-the-gift-of-divorce-23andme [http://perma.cc/C2DE-2RGL] (recounting a story of a man who, by searching the 23 andMe close relative finder program using his SNP genotyping data, identified that he had a half-brother that no one knew about, which eventually led to his parents' divorce).

96. Motluk, *supra* note 10.

97. Motluk, *supra* note 10. The boy found two potential brothers who did not know each other but had the same last name, through the Y chromosome genotyping. *Id.* The

In addition, specific disease risks of an individual can be calculated based on the risks of relatives with public genomic information.[98] In another study, scientists were able to estimate the decrease in genetic privacy for relatives of individuals who shared their data on openSNP.[99] They cross-referenced Facebook to identify the relatives of individuals who contributed to openSNP to theoretically estimate the risk of each relative having genotypes associated with Alzheimer's disease.[100]

The HeLa genome controversy, successful paternity tracing, and the disease risk calculations of relatives reveal potential privacy risks to relatives of public genomics participants.[101] With more genomic and other personal information available to the public, adversaries may be able to assess disease risks of the participants' relatives, regardless of their own participation in public genomics, as in the case of Krypton and Xenon presented in the beginning of the Comment.[102]

C. Consent for Future Studies and Discoveries

Genomics advanced rapidly in the past twenty years, and developments and discoveries remain ongoing.[103] Due to the evolving nature of future discoveries, the issue of consent for future studies or developments also poses a problem for public genomics.[104] For example, additional future

---

boy's mother knew the donor's birthdate and place of birth. *Id.* By accessing a database that contained the list of persons born on that date in that city, the boy was able to identify his donor from the last name shared with the potential brothers. *Id.* The boy contacted his donor father within 10 days. *Id.*

98. *See* Humbert *supra* note 90. This is an example of how Krypton's risks for X-ALD could be calculated based on Xenon's information that is publicly available. *See supra* Part I.

99. Humbert *supra* note 90. See *supra* note 70, at 6, for a discussion of openSNP.

100. Humbert *supra* note 90, at 1141, 1149. The scientists estimated the level of uncertainty for each of the relatives having two SNPs that are known to significantly increase an individual's probability of having Alzheimer's by the age of eighty, based on the published SNP genotyping data. *Id.* at 1150.

101. Humbert *supra* note 90.

102. *See supra* Part I.

103. *See* PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 19. For example, the relationship between a gene and a specific mutation to the expressed physical characteristics (phenotype) is only clearly elucidated for some genes, such as for mutations involved diseases that are inherited in clear Mendelian manner. *See id.* Rare variants that were not picked up by SNP-based GWAS are being uncovered by whole genome sequencing, but the implication of these variants on diseases are not all known. Elizabeth T. Cirulli & David B. Goldstein, *Uncovering the Roles of Rare Variants in Common Disease Through Whole Genome Sequencing*, 11 NATURE REVIEWS GENETICS 415, 415 (2010). Having a large number of available whole genome sequencing data, will facilitate discoveries of rare variants and their effects on diseases. *See id.* at 415–16.

104. *See* Conley et al., *supra* note 15, at 329; Jeantine E. Lunshof et al., *From Genetic Privacy to Open Consent*, 9 NATURE REVIEWS GENETICS 406, 408–09 (2008).

studies may show a link between initially neutral data and negative outcomes.[105] Because genomic information is permanent, future negative implications cannot be detached from the genomic information once it is discovered.[106] This problem is worse in public genomics because publicly released data cannot be retracted.[107] The initial public release is typically voluntary and presumably with informed consent.[108] However, scientists cannot fully *inform* the participants of all future developments, risks and implications.[109]

Another complication that may arise from public genomic information is consent for the use of the information in additional future studies.[110] The participants initially contribute the genomic data for a specific purpose such as data gathering for a public genomics project or a GWAS for a particular disease.[111] To maximize the utility of the gathered information, the data may be used in future studies.[112] Regulatory agencies have not agreed on rules to address the consent issue for these additional

105. *See* Euan A. Ashley et al., *Genetics and Cardiovascular Disease: A Policy Statement from the American Heart Association*, 126 CIRCULATION 142, 149 (2012). For example, someone who participated in a public genomics project may have a mutation that is not currently associated with any diseases. *See id.* In time, scientists may discover that this mutation is associated with high risk of a disease and could be considered a negative factor in obtaining life insurance. *See id.* In this case, the subject's initial consent to the study may not have been completely "informed." *See id.*; PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 48.

106. *See* Erman Ayday et al., *The Chills and Thrills of Whole Genome Sequencing*, ARXIV:1306.1264v4, at 5–6 (2013), http://arxiv.org/pdf/1306.1264v5.pdf [http://perma.cc/YB7F-FL53]. This problem is less evident in other human subject studies, because the genomic data contains uniquely identifying information and a significant proportion of unassociated data is released without fully understanding the implication. *See id.* Another issue is that because genomic information is permanent and cannot be changed like other information such as bank account numbers and passwords, the information is not retractable. *See id.*

107. *See* Brett A. Williams & Leslie E. Wolf, *Biobanking*, *Consent*, *and Certificates of Confidentiality: Does the ANPRM Muddy the Water?*, 41 J.L. MED. & ETHICS 440, 448–49 (2013).

108. *See* Conley et al., *supra* note 15, at 351.

109. *See id.*

110. *See* Charles Safran et al., *Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper*, 14 J. AM. MED. INFORMATICS ASS'N 1, 2 (2006); *See* Conley et al., *supra* note 15, at 354–56.

111. *See* Conley et al., *supra* note 15, at 354–56. For a brief explanation of GWAS, see *supra* note 48.

112. *See* Jane Kaye et al., *From Patients to Partners: Participant-Centric Initiatives in Biomedical Research*, 13 NATURE REVIEWS GENETICS 371, 372 (2012) (explaining that the data can be used in a GWAS for different diseases).

studies.[113] Under the current regulation of the Common Rule and HIPAA, research with de-identified information does not require updated consent or notice.[114] Consent issues for publicly available genomic information must be reevaluated, as the public typically does not distinguish between identifiable and de-identified data, but does want to control the use of the information.[115] Obtaining updated, informed consent is difficult and expensive, and current requirements do not cover publicly available de-identified information, so a new mechanism to easily obtain additional or renewed consent is needed.[116]

D. Difficulties in Uniform Guidelines and International Enforcement

Another obstacle in considering the privacy issues of publicly available genomic information is that the projects and consortia are typically of international scale, and there is no single law or regulation that governs these projects or potential privacy breaches.[117] Each of the projects or consortia sets its own guidelines, and relies on the researchers using the data to abide by the data use guidelines.[118] Countries also have different privacy standards regarding health-related information.[119] Even if the researchers abide by the guidelines, once the data is publicly available, it becomes much more difficult to control or retract.[120]

IV. CURRENT LEGISLATION, REGULATION, AND LIMITATIONS

There are multiple laws and regulations potentially affecting public genomics and privacy.[121] As seen with other rapidly developing technological

---

113. *See* Conley et al., *supra* note 15, at 355; *See* PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 75.

114. *See* Rothstein, *supra* note 87, at 8; see *infra* Part IV.A for a more detailed discussion of HIPAA, and Part IV.C for a more detailed discussion of the Common Rule.

115. *See* Donald J. Willison et al., *Patients' Consent Preferences for Research Uses of Information in Electronic Medical Records: Interview and Survey Data*, 326 BMJ 373, 374 (2003).

116. *See* Conley et al., *supra* note 15, at 351–52.

117. GLOBAL ALLIANCE FOR GENOMICS & HEALTH, CREATING A GLOBAL ALLIANCE TO ENABLE RESPONSIBLE SHARING OF GENOMIC AND CLINICAL DATA 12 (2013), http://genomicsandhealth.org/files/public/White%20Paper%20June%203%20final_0.pdf [http://perma.cc/W73N-3SFR] [hereinafter RESPONSIBLE SHARING]; *see* PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 65.

118. RESPONSIBLE SHARING, *supra* note 117, at 25.

119. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 65.

120. Ayday et al., *supra* note 106, at 5.

121. *See generally* PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 60–69. The scope of the legislations and regulations does not encompass all of the issues of public genomics. *Id.* at 68–69.

areas, laws and regulations regarding genomic privacy can lag behind the speed of development of the technology and currently do not sufficiently protect participants from potential misuse of information.[122]

A. Health Insurance Portability and Accountability Act of 1996

Health Insurance Portability and Accountability Act (HIPAA) is one of the earliest pieces of legislation encompassing medical information privacy.[123] Pursuant to the authority of Title II, HIPAA sets forth policies, procedures, and guidelines for maintaining the privacy and security of personally identifiable health information.[124] The HIPAA-mandated Privacy Rule[125] governs whether a covered entity, such as a healthcare provider or a health plan, can or cannot disclose patient-identifiable information, such as name, address and social security number.[126]

Under HIPAA, two methods can be used to achieve de-identification of medical information.[127] First, an expert can determine that the risk of re-identification of individual is "very small," rendering the information unidentifiable.[128] Second, the information can become "de-identified" by

122. Conley et al., *supra* note 15, at 338–40.

123. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 61–62.

124. Health Insurance Portability and Accountability Act (HIPAA) of 1996, Pub. L. 104-191, 110 Stat. 1936, 2029 (codified as amended in scattered sections of 18 U.S.C., 26 U.S.C., 29 U.S.C., & 42 U.S.C.). HIPAA was enacted August 21, 1996. *Id.*

125. Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, 45 C.F.R. §§ 164.502(a), 164.514(b) (2014). In 2013, the Privacy Rule was updated to integrate changes under "the [Health Information Technology for Economic and Clinical Health (HITECH)] Act, enacted as part of the American Recovery and Reinvestment Act of 2009, and the Genetic Information Nondiscrimination Act of 2008 (GINA)." Press Release, U.S. Dep't of Health & Human Servs., New Rule Protects Patient Privacy, Secures Health Information (Jan. 17, 2013), http://www.hhs.gov/news/press/2013pres/01/20130117b.html [http://perma.cc/H8P9-FNL6].

126. 45 C.F.R. § 164.501.

127. *Id.* § 164.514(b) (2014).

128. *Id.* ("[An expert is] a person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable: (i) applying such principles and methods, determines that the risk is *very small* that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and (ii) documents the methods and results of the analysis that justify such determination.") (emphasis added).

removing HIPAA identifiers.[129] Furthermore, a covered entity can disclose information only if "the recipient signs a data use agreement indicating that the information will be used only for limited purposes."[130] In *Baser v. Department of Veterans Affairs*, the U.S. District Court for the Eastern District of Michigan denied the Department of Veterans Affairs' motion for summary judgment for refusing to provide information under the Freedom of Information Act, even though the Department of Veterans Affairs provided expert opinion analyzing the risk of re-identification when linked with other publicly available or commercially available databases.[131] The court stated that under HIPAA, the information is "de-identified" if it uses either one of the two methods, "expert determination" or removal of the eighteen HIPAA identifiers, but does not require both methods.[132]

Courts have reviewed the risk of re-identification of personal medical information in a handful of cases under the HIPAA Privacy Rule.[133] In the limited number of cases from different jurisdictions, courts differ on their assessment of re-identification risk in medical information for public release or for discovery purposes.[134] In some of the cases, courts acknowledged the risk of re-identification and the invasion of privacy.[135] In *Northwestern Memorial Hospital v. Ashcroft*, the Seventh Circuit Court concluded that the hospital's production of forty-five subpoenaed medical records would breach the privacy interests of the patients.[136] The court balanced the benefits and burdens of producing the subpoenaed record, including the risk of re-identification.[137] The court stated that even with the HIPAA identifiers removed, the patients' acquaintances or "skillful

---

129. *Id.* §§ 160.103, 164.514. HIPAA identifiers include names; address; dates; phone numbers; fax numbers; email addresses; social security numbers; medical record numbers; health plan beneficiary; numbers; account numbers; certificate/license numbers; vehicle identifiers; device identifiers and serial; numbers; web URLs; internet protocol (IP) addresses; biometric identifiers, including; finger and voice prints; full face photographic images; and any comparable images; any other unique identifying; number, characteristic, or code, (with certain exceptions). *Id.* §§ 160.103, 164.514(b) (2014).

130. Rothstein, *supra* note 87, at 4 (citing 45 C.F.R. §164.514(e)(4)).

131. Baser v. Dep't of Veterans Affairs, No. 13-CV-12591, 2014 WL 4897290, at *4, *7 (E.D. Mich. Sept. 30, 2014).

132. *Id.* at *4–5.

133. *See, e.g.*, Nw. Mem'l Hosp. v. Ashcroft, 362 F.3d 923, 924–26 (7th Cir. 2004); *Baser*, 2014 WL 4897290, at *4–5; Havemann v. Astrue, No. ELH-10-1498, 2012 WL 4378143 (D. Md. Sept. 24, 2012); Steinberg v. CVS Caremark Corp., 899 F. Supp. 2d 331, 338–39 (E.D. Pa. 2012); Cohan v. Ayabe, 322 P.3d 948, 957–59 (Haw. 2014).

134. *See* cases cited *supra* note 133.

135. *Nw. Mem'l Hosp.*, 362 F.3d at 929; *Havemann*, 2012 WL 4378143, at *1; *Cohan*, 322 P.3d at 950.

136. *Nw. Mem'l Hosp.*, 362 F.3d at 929.

137. *Id.*

Googlers" can re-identify the patients from the released data and can invade the patient's privacy.[138] Similarly, in *Havemann v. Astrue*, the U.S. District Court in Maryland granted summary judgment in favor of the Social Security Administration, who produced de-identified medical records under the Freedom of Information Act.[139] The plaintiff sought injunctive relief for the release of the full records, but the court reasoned that further release of information can lead to re-identification and violation of the privacy of individuals who are included in the record.[140]

In contrast, other courts dismissed the re-identification risk and considered the information to be sufficiently de-identified if the data is free of HIPAA identifiers.[141] In *Steinberg v. CVS Caremark Corporation*, the court rejected the plaintiff's argument that information de-identified according to HIPAA standards can be re-identified.[142] The court rejected the argument because the plaintiff only provided an academic journal article explaining the general risk of re-identification, but not any actual expert analysis of the data from the case.[143]

In general, few courts have determined that the risk of re-identification of medical data is significant, and many consider that removal of the eighteen HIPAA identifiers is sufficient for privacy protection.[144] However, the approach requires a revisit in the context of genomic information, as genomic data contains inherently identifiable information that is permanent.[145]

B. Genetic Information Nondiscrimination Act of 2008

Unlike HIPAA, which provides some privacy protection, the Genetic Information Nondiscrimination Act (GINA) provides protection against discrimination based on genetic information in health insurance and employment.[146] Under GINA, health insurers cannot: (1) use genetic

138. *Id.*
139. Havemann v. Astrue, No. ELH-10-1498, 2012 WL 4378143, at *7–9 (D. Md. Sept. 24, 2012).
140. *Id.*
141. Baser v. Dep't of Veterans Affairs, No. 13-CV-12591, 2014 WL 4897290, at *5 (E.D. Mich. Sept. 30, 2014); Steinberg v. CVS Caremark Corp., 899 F. Supp. 2d 331, 336–37 (E.D. Pa. 2012).
142. *Steinberg*, 899 F. Supp. 2d at 339.
143. *Id*.
144. *See id.* at 337; *Baser*, 2014 WL 4897290, at *5.
145. Ayday et al., *supra* note 106, at 5.
146. Genetic Information Nondiscrimination Act (GINA) of 2008, Pub. L. No. 110-233, 122 Stat. 881 (codified as amended in scattered sections of 29 U.S.C. & 42 U.S.C.);

information to determine coverage, eligibility, or premiums; (2) request or require genetic testing or genetic information for underwriting decisions; and (3) obtain genetic information for underwriting purposes.[147] Employers with more than fifteen employees cannot "fail or refuse to hire . . . discharge . . . or, otherwise to discriminate against any employee with respect to the compensation, terms, conditions, or privileges of employment" because of an employee's genetic information.[148]

In many of the actions brought under violation of GINA, the definition of "genetic information" was contested.[149] Although numerous employment lawsuits have used violation of GINA as a basis of a claim, many courts have concluded that plaintiffs failed to show a valid claim under GINA.[150] For example, in *Dumas v. Hurley Medical Center*, the U.S. District Court for the Eastern District of Michigan concluded that the plaintiff did not provide sufficient basis for a claim under GINA because the allegation for termination was based on the plaintiff's disclosure to the employer that she suffered physical and mental disabilities.[151] The court stated that the complaint did not allege any use or misuse of *genetic* information.[152]

GINA also extends coverage to genetic information of family members in addition to genetic information of individuals.[153] In particular, the

---

PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 67; Louise Slaughter, Essay*, Genetic Information Non-discrimination Act*, 50 HARV. J. ON LEGIS. 41, 42 (2013).

147. 29 U.S.C. § 1182 (2012).

148. 42 U.S.C. § 2000ff (2012).

149. *See, e.g.*, Conner-Goodgame v. Wells Fargo Bank, N.A., No. 2:12-cv-03426-IPJ, 2013 WL 5428448, at *10–11 (N.D. Ala. Sept. 26, 2013); Bell v. PSS World Med., Inc., No. 3:12-cv-381-J-99MMH-JRK, 2012 WL 6761660, at *2–3 (M.D. Fla. Dec. 7, 2012); Graves v. Brookfield Suites Hotel & Convention Ctr., No. 11-CV-01060, 2012 WL 3941774, at *1 (E.D. Wis. Sept. 10, 2012); Leone v. N. Jersey Orthopaedic Specialists, P.A., No. 11-3957 (ES), 2012 WL 1535198, at*5 (D.N.J. Apr. 27, 2012); Culbreth v. Wash. Metro. Area Transit Auth., No. RWT 10cv3321, 2012 WL 959385, at *3–4 (D. Md. Mar. 19, 2012); Dumas v. Hurley Med. Ctr., 837 F. Supp. 2d 655, 666 (E.D. Mich. 2011). GINA defines "genetic information" as "individual's genetic tests . . . the genetic tests of family members . . . the manifestation of a disease or disorder in family members of such individual." 42 U.S.C. § 2000ff(4). GINA also includes genetic information from participation in genetic research. *Id.*

150. *See, e.g.*, *Dumas*, 837 F. Supp. 2d at 666. Other courts have decided in a similar manner, dismissing claims under GINA because the complaints did not state any basis for discrimination based on genetic information. *See, e.g.*, *Conner-Goodgame*, 2013 WL 5428448 at *11; Williams v. Wells, No. 4:12-cv-02434-RBH, 2013 WL 4042037, at *1–2, *5 (D.S.C. Aug. 8, 2013); *Bell*, 2012 WL 6761660 at *3–4; *Graves*, 2012 WL 3941774 at *1; *Leone*, 2012 WL 1535198 at *6; *Culbreth*, 2012 U.S. Dist. LEXIS 37335 at *4.

151. *Dumas*, 837 F. Supp. 2d at 659, 666–67.

152. *Id.* at 666.

153. GINA defines family member as "a dependent . . . and [] any other individual who is a first-degree, second-degree, third-degree, or fourth-degree relative of such individual or of an individual described [as a dependent]." Genetic Information

manifestation of a disease or disorder by a family member is considered genetic information under GINA.[154] In *Bronsdon v. City of Naples*, the U.S. District Court for the Middle District of Florida explained, "plaintiff alleges that defendant's decision to deny benefits was based on his family's medical history . . . discrimination based on family medical history is prohibited under GINA, even if the individual has a manifested condition."[155] However, family medical history is only critical as genetic information if it is relevant in determining the individual's risk for a genetic disease.[156] In *Poore v. Peterbilt of Bristol*, the U.S. District Court for the Western District of Virginia concluded that the plaintiff's wife's diagnosis of multiple sclerosis did not provide sufficient basis for a claim under violation of GINA.[157] The court stated that the plaintiff's wife's diagnosis was not genetic information because his wife's diagnosis is not relevant to determine his genetic risks.[158] In *Lee v. City of Moraine Fire Department*, the U.S. District Court for the Southern District of Ohio considered "information about whether an employee's 'primary relative' has a history of prostate cancer" to be genetic information under GINA.[159]

Although GINA provides protection against discrimination based on genetic information, it does not cover all instances of discrimination.[160] For example, it does not cover discrimination in disability insurance, life insurance, or long-term care.[161] For public genomic information, GINA

Nondiscrimination Act (GINA) of 2008, Pub. L. No. 110-233, § 122 Stat. 885 (codified at 42 U.S.C. § 2000ff(3) (2012)).

154. 42 U.S.C. § 2000ff(4). Such information would include family disease history. *Id.*

155. Bronsdon v. City of Naples, No. 2:13-cv-778-FtM-29CM, 2014 U.S. Dist. LEXIS 70502, at *8 (M.D. Fla. May 22, 2014).

156. Maxwell v. Verde Valley Ambulance Co., No. CV-13-08044-PCT-BSB, 2014 WL 4470512, at *16–17 (D. Ariz. Sept. 11, 2014) (citing Poore v. Peterbilt of Bristol, L.L.C., 852 F. Supp. 2d 727 (W.D. Va. 2012)).

157. *Poore*, 852 F. Supp. 2d at 731.

158. *Id.* at 730–31. "[S]uch information is taken into account only with respect to the individual in which such disease or disorder occurs and not as genetic information with respect to any other individual." *Id.* (quoting H.R. REP. NO. 110-28, pt. 1, at 36 (2007)).

159. Lee v. City of Moraine Fire Dep't, No. 3:13cv00222, 2014 WL 1775621, at *5 (S.D. Ohio May 2, 2014).

160. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 67.

161. Brianna E. Kostecka, *GINA Will Protect You, Just Not From Death: The Genetic Information Nondiscrimination Act and Its Failure To Include Life Insurance within Its Protections*, 34 SETON HALL LEGIS. J. 93, 95 (2009); Angela L. Morrison, *A Research Revolution: Genetic Testing Consumers Become Research (and Privacy) Guinea*

does not regulate privacy concerns such as access or data security, but protects from discrimination in employment or health insurance contexts.[162]

C. The Common Rule: Regulation of Consent for Human Subjects

In addition to anti-discrimination and privacy laws, another aspect of regulation regarding public genomics concerns the rules for human subject research and consent.[163] The Common Rule,[164] a federal regulation governing federally funded human subject research in the United States, requires an independent institutional review board (IRB) to review and approve procedures and informed consent for human subject research.[165] Informed consent includes describing the procedure, explaining the procedure's risks and benefits, and providing the participants information such as the right to withdraw from the study and the degree of confidentiality in the research record.[166] However, previously available genomic data

---

Pigs, 9 J. ON TELECOMM. & HIGH TECH. L. 573, 583–84 (2011); Kimberly Shoenbill et al., *Genetic Data and Electronic Health Records: A Discussion of Ethical, Logistical and Technological Considerations*, 21 J. AM. MED. INFORMATICS ASS'N 171, 174 (2014); Slaughter, *supra* note 146, at 54. In the hypothetical from the Introduction of this Comment, Krypton was denied disability insurance based on his genetic information, and therefore may not have a cause of action against the insurance company under GINA. *See supra* Part I.

162. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 66–67.

163. *Id.* at 63–64.

164. The "Common Rule" was published in 1991 and codified by fifteen Federal departments and agencies as separate regulations. *Federal Policy for the Protection of Human Subjects ('Common Rule')*, U.S. DEP'T HEALTH & HUM. SERVS., http://www.hhs.gov/ohrp/humansubjects/commonrule/ [http://perma.cc/3L5Z-TNU8] (last visited July 27, 2015). The Department of Health and Human Services codified its rule in Protection of Human Subjects, 45 C.F.R. § 46. *Id.*; Stanley G. Korenman, *Common Rule*, TEACHING RESPONSIBLE CONDUCT RES. HUMANS (RCRH), http://ori.hhs.gov/education/products/ucla/chapter2/page04b.htm [http://perma.cc/3ZF3-LT9U] (last visited July 27, 2015).

165. Protection of Human Subjects, 45 C.F.R. § 46 (2014). At least five professionals of various backgrounds, including at least one scientist and at least one non-scientist, are required for the institutional review board (IRB). *Id.* § 46.107. The IRB must also include at least one member who is not affiliated with the institution and not an immediate family member of someone affiliated with the institution that is reviewed. *Id.* For IRB approval, the Common Rule sets forth criteria including: 1) minimizing risks to subjects; 2) evaluation of risk to subject compared to anticipated benefits; 3) equitable selection of subjects; 4) obtaining informed consent for subjects; 5) documented informed consent; 6) provisions to ensure subject safety; and 7) provisions to protect privacy of subjects and confidentiality of data. *Id.* § 46.111.

166. *Id.* § 46.116. One court has tested the reach of the Common Rule in the context of ownership of biological samples used in human subject research. Washington Univ. v. Catalona, 437 F. Supp. 2d 985, 990–91 (E.D. Mo. 2006). In *Washington Univ. v. Catalona*, the university filed a declaratory judgment action against the doctor who directed the research and the research participant who provided the sample to establish ownership of

free of identifying information can be used for additional research purposes without further IRB review or additional consent.[167]

This highlights some limitations of the Common Rule for public genomics projects. For example, in some public genomics projects, prior consent likely did not state the newly discovered risk of re-identification.[168] Furthermore, genomic information itself inherently contains re-identifiable data, and the lack of consent requirement for de-identified information may pose additional risks.[169] The de-identification requirements of the Common Rule are even lower than under HIPAA.[170] In addition, the Common Rule only applies to federally funded research, so many public genomics or open-source genomics projects are not covered.[171]

To address some of the issues regarding data re-identification and consent, the Department of Health and Human Services (DHHS) published an Advanced Notice of Proposed Rulemaking concerning whether certain types of genomic data should be considered identifiable.[172] For example, DHHS recognizes the risks of re-identification, and proposes to make the

---

the sample. *Id.* at 987. The District Court for the Eastern District of Missouri decided that the university obtained proper consent under the Common Rule and retains ownership of the biological materials. *Id.* at 991, 1002.

167. 45 C.F.R. § 46.101(b)(4) (exempting "[r]esearch involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects"); PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 63–64; *OHRP Guidance on Research Involving Coded Private Information or Biological Specimens*, OFF. FOR HUM. RES. PROTECTIONS, U.S. DEP'T HEALTH & HUM. SERVS. (Oct. 16, 2008), http://www.hhs.gov/ohrp/policy/cdebiol.html [http://perma.cc/SM6L-2VLZ] [hereinafter *OHRP Guidance*].

168. Conley, *supra* note 15, at 329; Lunshof et al., *supra* note 104, at 408.

169. Rothstein, *supra* note 87, at 6.

170. *See OHRP Guidance*, *supra* note 167. For example, under the Common Rule guidelines, information is considered "[not] . . . individually identifiable when they cannot be linked to specific individuals by the investigator(s) either directly or through coding system." *Id.* The Common Rule considers any coded information, where individual identity is converted into a code, or a number, de-identified. *Id.*; Rothstein, *supra* note 87, at 5.

171. COMM. ON HEALTH RESEARCH AND THE PRIVACY OF HEALTH INFO.: THE HIPAA PRIVACY RULE, INST. OF MED. OF THE NAT'L ACADS., BEYOND THE HIPAA PRIVACY RULE: ENHANCING PRIVACY, IMPROVING HEALTH THROUGH RESEARCH 126 (Sharyl J. Nass et al. eds., 2009).

172. Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators, 76 Fed. Reg. 44,512, 44,524 (proposed July 26, 2011) (to be codified at 21 C.F.R. pt. 50, 56 and 45 C.F.R. pt. 46, 160, 164).

Common Rule de-identification standard at least equal to the HIPAA standard, if not stricter.[173] Several different interest groups have presented opposing views dealing with whether genomic data contains information that is inherently identifiable, but DHHS has not released new guidelines.[174]

D. State Laws Regarding Genetic Information

Many states have additional anti-discrimination or privacy laws that offer varying scope protection.[175] Some of the state laws were enacted before GINA, and some serve to fill the gaps in protection under GINA.[176] GINA provides minimum protection as a federal law, but allows states to provide additional safeguards.[177]

Nineteen states have additional protection from discrimination beyond protection in employment and health insurance provided by GINA, such

---

173. *Id.* at 44,524–25.

> [R]apidly evolving advances in technology coupled with the increasing volume of data readily available may soon allow identification of an individual from data that is currently considered deidentified . . . . We are considering adopting the HIPAA standards for purposes of the Common Rule regarding what constitutes individually identifiable information . . . . [I]t might be advisable to evaluate the set of identifiers that must be removed for a data set to be considered "de-identified" under both human subjects regulations and the HIPAA Privacy Rule . . . [W]e are considering categorizing all research involving the primary collection of biospecimens as well as storage and secondary analysis of existing biospecimens as research involving identifiable information . . . .

Id.

174. *See, e.g.*, American Anthropological Association, Comment Letter on Proposed Rule on Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators (Oct. 19, 2011), http://www.aaanet.org/issues/policy-advocacy/upload/Human-Subjects-Research.pdf [http://perma.cc/2A63-QKAG]; Biotechnology Industry Organization, Comment Letter on Proposed Rule on Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators (Oct. 26, 2011), https://www.bio.org/sites/default/files/BIO%20Common%20Rule%20ANPRM%20comments%20FINAL-10%2026%202011_0.pdf [https://perma.cc/5GBJ-42QB]; Consortium of Independent Review Boards, Comment Letter on Proposed Rule on Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators (Oct. 26, 2011), http://www.consortiumofirb.org/Comments_on_ANPRM_October_2011.pdf [http://perma.cc/CH7D-4PKN]; World Privacy Forum, Comment Letter on Proposed Rule on Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators (Oct. 18, 2011), http://www.worldprivacyforum.org/wp-content/uploads/2011/10/WPF_CommonRule_Oct182011fs.pdf [http://perma.cc/NAR5-Z4BM]; *see also* Williams & Wolf, *supra* note 107, at 446.

175. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 67.

176. *Id.*; Kostecka, *supra* note 161, at 101.

177. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 67.

as in life insurance, long-term care, or disability insurance.[178] For example, the California Genetic Information Nondiscrimination Act extends anti-discrimination protection for emergency medical services, housing, receipt of services, qualifications for licensing, and participation in any state-funded programs, providing one of the widest range of protection beyond federal GINA.[179]

178. The nineteen states with their respective genetic information protection statues are: Arizona: ARIZ. REV. STAT. ANN. §§ 20-448, 41-1463 (2011 & Supp. 2014); California: CAL. CIV. CODE § 51(b) (West Supp. 2015); CAL. GOV'T. CODE § 12926(i)(2) (West Supp. 2015); CAL. INS. CODE 10149.1 (West 2013); Colorado: COLO. REV. STAT. § 10-3-1104.7(1) (2014); Idaho: IDAHO CODE ANN. § 39-8301–8304 (West 2011); IDAHO CODE ANN. § 41-1313 (West 2010); Kansas: KAN. STAT. ANN. §§ 40-2259, 44-1009(a)(9) (2000 & Supp. 2014); Kentucky: KY. REV. STAT. ANN. § 304.12-085(2)–(3) (West 2012); Maine: ME. REV. STAT. ANN. tit. 24-A, § 2159-C (2015); ME. REV. STAT. ANN. tit. 5, § 19302 (2013); Maryland: MD. CODE ANN., INS. §§ 18-120, 27-909 (LexisNexis 2011); MD. CODE ANN., STATE GOV'T. § 20-606 (LexisNexis 2014); Massachusetts: MASS. GEN. LAWS ANN. ch. 151B, § 4 (West Supp. 2015); MASS. GEN. LAWS ANN. ch. 175, §§ 108H, 108I, 120E (West 2011); Minnesota: MINN. STAT. §§ 72A.139, 181.974, 375.1306 (2014); Missouri: MO. REV. STAT. § 375.1303 (West 2013); New Hampshire: N.H. REV. STAT. ANN. §§ 141-H:1 to H:4 (2014); New Jersey: N.J. STAT. ANN. §§ 17B:26-3.2, 17B:30-12 (West 2014); New Mexico: N.M. STAT. ANN. § 24-21-1 to -7 (West 2014); New York: N.Y. EXEC. LAW. § 296 (McKinney 2014); N.Y. CIV. RIGHTS LAW § 79-l (McKinney 2014); N.Y. INS. LAW § 2615 (McKinney 2014); Oregon: OR. REV. STAT. §§ 192.531, 659A.303 (2013); Texas: TEX. INS. CODE ANN. §§ 546.001–.051 (West 2014); TEX. LAB. CODE ANN. §§ 21.401–.402 (West 2014); Vermont: VT. STAT. ANN. tit. 18 §§ 9331–34 (2014); Wisconsin: WIS. STAT. §§ 111.32–.335, 631.89 (2014). *See* Anya E.R. Prince, *Comprehensive Protection of Genetic Information: Once Size Privacy or Property Models May Not Fit All*, 79 BROOKLYN L. REV. 175, 183–85 (2013). If Krypton were a resident of one of the states that provide protection for disability insurance, such as Arizona, Idaho, Kansas, Kentucky, Massachusetts, New Jersey, New York, Vermont or Wisconsin, he may have a valid state claim against the disability insurance company who discriminated against him in his disability insurance application. *See supra* Part I.

179. 2011 Cal Legis. Serv. 2888 (West). California Genetic Information Nondiscrimination Act was proposed as S.B. 559 by Alex Padilla in the 2011-2012 Regular Session and was codified as amended at CAL. BUS. & PROF. CODE § 23438; CAL. CIV. CODE § 51; CAL. EDUC. CODE § 32228; CAL. ELEC. CODE § 354.5; CAL. GOV'T CODE §§ 11135, 12920, 12921, 12926, 12926.1, 12930, 12931, 12935, 12940, 12944, 12955, 12955.8, 12956.1, 12956.2, 12993; CAL. PENAL CODE § 868.8; CAL. REV. & TAX. CODE § 17269, 24343.2; CAL. WELF. & INST. CODE § 4900. *Id. See also* Jennifer K. Wagner, *A New Law To Raise GINA's Floor in California*, GENOMICS L. REP. (Dec. 7, 2011), http://www.genomics lawreport.com/index.php/2011/12/07/a-new-law-to-raise-ginas-floor-in-california/ [http://perma.cc/6BVQ-5MG6]. California's protection is one of the first strong and comprehensive protections against discrimination. Prince, *supra* note 178, at 211.

States typically focus on either a property regime or a privacy regime in the establishment of genetic information protection legislation.[180] States that have a property regime for protection focus on the genetic material itself, such as property interest of the DNA sample, and cover limited genetic information.[181] Under the property regime, the statutes provide the genetic information providers, such as participants of a genomics study, property rights to their genetic information.[182] For example, Alaska Statutes section 18.13.010(a)(2), with the broadest protection, provides that the DNA samples and the results of a DNA analysis are exclusive property of the person providing the samples.[183] The Alaska statute provides a private right of action against those who surreptitiously collect a DNA sample or those who disclose DNA testing results.[184] These laws tend to focus on the consent requirements for gathering DNA samples, and to a limited extent, genetic information.[185]

Other states have constitutional provisions regarding privacy or separate privacy protection laws.[186] Ten states have constitutional provisions that

---

180. The five states that create a property regime in protecting genetic information with their respective genetic information protection statues are: Alaska: ALASKA STAT. § 18.13 (2014); Colorado: COLO. REV. STAT. § 10-3-1104.7(1) (2014); Florida: FLA. STAT. §§ 760.40(2)(a), 627.6561 (2014); Georgia: GA. CODE ANN. § 33-54-1 (2014); Louisiana: LA. STAT. ANN. § 22:1023(E) (2014). *See* Prince, *supra* note 178, at 183–185.

181. Prince, *supra* note 178, at 195.

182. Jana M. Belflower, Note, *Keeping Pace With Progress: A Proposal for Florida's Genetic Testing Statute*, 42 STETSON L. REV. 249, 264–65 (2012).

183. ALASKA STAT. § 18.13.010(a)(2) (2014). The Colorado, Georgia, and Louisiana statutes are limited to insurance contexts. COLO. REV. STAT. § 10-3-1104.6 (2014); GA. CODE ANN. § 33-54-1 (2013); LA. STAT. ANN. § 22:1023 (2014). The Florida statute is not limited to insurance contexts, and it only applies to genetic analysis results and not the physical sample itself. FLA. STAT. § 760.40 (2014); Belflower, *supra* note 182, at 265 n.104. The District Court of Appeal of Florida has tested the consent requirement for genetic analysis of the Florida statute in *Doe v. Suntrust Bank*, 34 So. 3d 133 (Fla. Dist. Ct. App. 2010) where the plaintiff, the decedent's trust, sued the known children of the decedent to submit to DNA testing in order to determine whether the alleged children were related to the known children. *Suntrust Bank*, 32 So. 3d at 135–36. The court quashed the trial court order to produce DNA analysis samples, stating that the purpose of the statute is to protect individuals from unconsented DNA analysis, and that the plaintiff did not provide sufficient "good cause" for the request. *Id.* at 138–41 (stating that the section 760.40 of the Florida Statutes "criminalizes performing DNA analysis or disclosing the results without obtaining the informed consent of the person tested.").

184. ALASKA STAT. § 18.13.020 (2014) ("A person may bring a civil action against a person who collects a DNA sample from the person, performs a DNA analysis on a sample, retains a DNA sample or the results of a DNA analysis, or discloses the results of a DNA analysis in violation of this chapter.").

185. Prince, *supra* note 178, at 195–96.

186. *See* Prince, *supra* note 178, at 184; *Privacy Protections in State Constitutions*, NAT'L CONF. ST. LEGISLATURES (Dec. 12, 2014), http://www.ncsl.org/research/telecommunications-and-information-technology/privacy-protections-in-state-constitutions.aspx [http://perma.cc/2P3S-5X8N].

relate to the right of privacy.[187] In addition to provisions that mirror the Fourth Amendment protections regarding search and seizure, the state privacy provisions also have specific references to privacy, and provide an expanded privacy protection.[188] In Hawaii, the state supreme court evaluated the state's constitutional right of privacy in the medical information context.[189] In *Cohan v. Ayabe*, the Hawaii Supreme Court granted a mandamus relief to prevent the release of confidential health information outside of the specific litigation, reasoning that although the

187. The ten states with their respective state constitution provisions regarding privacy protection are: Alaska: ALASKA CONST. art. I, § 22 ("The right of the people to privacy is recognized and shall not be infringed."); Arizona: ARIZ. CONST. art. II, § 8 ("No person shall be disturbed in his private affairs, or his home invaded, without authority of law."); California: CAL. CONST. art. I, § 1 ("[A]mong [the inalienable rights of people] are those of enjoying and defending life and liberty; acquiring, possessing, and protecting property; and pursuing and obtaining safety, happiness."); Florida: FLA. CONST. art. I, § 23 ("Every natural person has the right to be let alone and free from governmental intrusion into the person's private life except as otherwise provided herein."); Florida: FLA. CONST. art. I, § 12 ("The right of the people to be secure in their persons, houses, papers and effects against unreasonable searches and seizures, and against the unreasonable interception of private communications by any means, shall not be violated."); Hawaii: HAW. CONST. art. I, § 6 ("The right of the people to privacy is recognized and shall not be infringed without the showing of a compelling state interest."); Hawaii: HAW. CONST. art. I, § 7 ("The right of the people to be secure in their persons, houses, papers and effects against unreasonable searches, seizures and invasions of privacy shall not be violated[.]"); Illinois: ILL. CONST. art. I, § 6 ("The people shall have the right to be secure in their persons, houses, papers and other possessions against unreasonable searches, seizures, invasions of privacy or interceptions of communications by eavesdropping devices or other means."); Louisiana: LA. CONST. art. I, § 5 ("Every person shall be secure in his person, property, communications, houses, papers, and effects against unreasonable searches, seizures, or invasions of privacy."); Montana: MONT. CONST. art. II, § 10 ("The right of individual privacy is essential to the well-being of a free society and shall not be infringed without the showing of a compelling state interest."); South Carolina: S.C. CONST. art. I, § 10 ("The right of the people to be secure in their persons, houses, papers, and effects against unreasonable searches and seizures and unreasonable invasions of privacy shall not be violated . . . ."); Washington: WASH. CONST. art. I, § 7 ("No person shall be disturbed in his private affairs, or his home invaded, without authority of law."). *See Privacy Protections in State Constitutions*, *supra* note 186. In addition, Missouri voters approved a state constitutional amendment that provides protection for unreasonable search and seizure of electronic data, making the law the first of its kind in the nation. Jason C. Gavejian, *Missouri Constitutional Amendment Protects Electronic Privacy*, WORKPLACE PRIVACY, DATA MGMT. & SECURITY REP. (Aug. 8, 2014), http://www.workplaceprivacyreport.com/2014/08/articles/workplace-investigations/missouri-constitutional-amendment-protects-electronic-privacy/ [http://perma.cc/NJ3K-FAVB].

188. *See Privacy Protections in State Constitutions*, *supra* note 186.

189. HAW. CONST. art. I, § 6; Cohan v. Ayabe, 322 P.3d 948, 965 (Haw. 2014).

information was de-identified, there was a risk of re-identification that would amount to an invasion of privacy under Article I, Section 6 of Hawaii's constitution.[190] In other cases, state constitutional privacy provisions were tested in the context of criminal law DNA data banks or court-ordered paternity testing.[191] However, courts have not evaluated privacy protections provided by the state constitution in the context of re-identification of public genomic information.[192]

State legislation focusing on privacy rights of genetic information vary widely in the scope of coverage.[193] Ten states have general privacy protection laws regarding an individual's right to privacy in genetic

190. HAW. CONST. art. I, § 6; *Cohan*, 322 P.3d at 965 (citing Nw. Mem'l Hosp. v. Ashcroft, 362 F.3d 923, 933 (7th Cir. 2004)).

191. *See*, *e.g.*, Nason v. State, 102 P.3d 962 (Alaska Ct. App. 2004) (upholding the constitutionality of sections 44.41.035(b) and 11.56.760(a) of the Alaska DNA collection statutes); People v. Buza, 180 Cal. Rptr. 3d 753 (Ct. App. 2014) (striking down the Forensic Identification Data Base and Data Bank Act of 1998, section 295 of the California Penal Code, as being invalid under article I, section 13 and 1 of the California Constitution because it intrudes the arrestee's privacy), *cert. granted*, 342 P.3d 415 (Cal. 2015) (mem.); Cnty. of San Diego v. Mason, 147 Cal. Rptr. 3d 135 (Ct. App. 2012) (holding that the trial court's order for DNA paternity testing did not violate privacy provided under article 1, section 1 of the California Constitution).

192. LexisAdvance searches performed on January 18, 2015, with the following search terms resulted in no cases: genom! and (privacy and constitution!) and (re-identif! or reidentif!); genom! and (privacy and constitution!) and (de-identif! or deidentif!); (genom! and privacy) and (re-identif! or reidentif!); (genom! and privacy) and (de-identif! or deidentif!). The search term "(constitution! and privacy) and (genom! or genet!) and (re-identif! or reidentif!)" resulted in three cases, but none were relevant to publicly available genetic information. Higgins v. Tex. Dep't of Health Servs., 801 F. Supp. 2d 541 (W.D. Tex. 2011); People v. Boyer, 133 P.3d 581 (Cal. 2006); Meyers v. Zoning Bd. of Appeals, No. CV 950535547, 1997 WL 325816 (Conn. Super. Ct. June 3, 1997). In *Higgins v. Texas Department of Health Services*, a class of parents whose infant children's blood samples were taken for screening under section 33.011 of the Annotated Texas Health and Safety Code sued the Texas Department and Health Services, stating concerns of re-distribution and re-identification as one of the reasons. *Higgins*, 801 F. Supp. 2d at 544–47. The Federal District Court for the Western District of Texas held that the plaintiffs only had a speculative injury and were unable to substantiate their concerns about re-distribution or re-identification; therefore, they lacked standing. *Id.* at 553. Most litigation regarding genetic data involve challenges to forensic DNA databases under state privacy provisions or the Fourth Amendment of the constitution. *See, e.g.*, cases cited *supra* note 191; U.S. v. Pool, 621 F.3d 1213 (9th Cir. 2010) (determining that the government's interest in identification using DNA data outweighed defendant's privacy interests), *vacated*, 659 F.3d 761 (9th Cir. 2011) (en banc); Banks v. U.S., 490 F.3d 1178 (10th Cir. 2007) (deciding that enforcement DNA Analysis Backlog Elimination Act of 2000 did not violate defendant's Fourth Amendment rights); U.S. v. Kincade, 379 F.3d 813 (9th Cir. 2004) (deciding that DNA Analysis Backlog Elimination Act of 2000, which requires certain convicted individuals to provide biological samples for DNA analysis, did not violate defendant's Forth Amendment rights).

193. *See* Prince, *supra* note 178, at 198.

information.[194] These laws establish rules on collection, access, retention and disclosure of genetic information, and generally require informed consent to collect, access, retain, or disclose the genetic information.[195] For example, New Jersey's Genetic Privacy Act forbids collecting, retaining, or disclosing genetic information without authorization from the individual providing the genetic information.[196] Delaware requires physical genetic samples to be destroyed promptly after obtaining genetic information unless necessary for criminal investigation, authorized by court order, authorized by the individual, or necessary for anonymous research purposes where the identity of the subject is not released.[197] New York's statute considers the privacy risks to the relatives in addition to the risks posed to the individuals themselves.[198] Some states that provide privacy or property-based protection of genetic information include

194. The ten states with their respective state constitution provisions regarding privacy protection are: Delaware: DEL. CODE ANN. tit.16 § 1202(a) (2012) (establishing a privacy interest); DEL. CODE ANN. tit.18 § 2317 (2012) (regulating insurance); DEL. CODE ANN. tit. 19 § 711 (2012) (regulating employment); Illinois: 410 ILL. COMP. STAT. 513/1–/35 (2013); Iowa: IOWA CODE ANN. § 729.6 (West 2010); Minnesota: MINN. STAT. § 13.386(3)(4)(ii) (2010) (establishing a privacy interest); MINN. STAT. § 72A.139 (2010) (regulating health and life insurance); MINN. STAT. § 181.974 (2010) (regulating employment); New Hampshire: N.H. REV. STAT. ANN. §§ 141-H:1–:4 (LexisNexis 2009); New Jersey: N.J. STAT. ANN. § 10:5-43 to -48 (West 2010) (establishing a privacy interest); N.J. STAT. ANN. § 17B:26-3.2 (West 2010) (regulating genetic information in health insurance); N.J. STAT. ANN. § 17B:30-12 (West 2010) (regulating life and disability insurance); New Mexico: N.M. STAT. ANN. § 24-21-1 to -7 (2005); New York: N.Y. EXEC. LAW. § 296 (McKinney 2014) (regulating employment); N.Y. CIV. RIGHTS LAW § 79-l (McKinney 2014) (establishing a privacy right); N.Y. INS. LAW § 2615 (McKinney 2014) (regulating health, life, long-term care, and disability insurances); Oregon: OR. REV. STAT. § 192.531 (2012) (establishing a privacy right); OR. REV. STAT. § 659A.303 (2012) (regulating employment); OR. REV. STAT. § 746.135 (2012); South Dakota: S.D. CODIFIED LAWS § 34-14-22 (2003) (establishing a privacy right); S.D. CODIFIED LAWS § 58-1-25 (2003) (regulating health insurance); S.D. CODIFIED LAWS § 60-2-20 (2003) (regulating employment). *See also* Prince, *supra* note 178, at 198.

195. *See* Prince, *supra* note 178, at 198. Interestingly, the Minnesota statute provides that written consent for dissemination is only valid for one year at most. MINN. STAT. § 13.386(3)(4)(ii) (West 2010).

196. N.J. STAT. ANN. § 10:5-45 (West 2010).

197. DEL. CODE ANN. tit. § 1203(b) (2012).

198. N.Y. INS. LAW § 2615 (McKinney 2014) ("No person who lawfully possesses information derived from a genetic test who may be genetically related to the tested individual; nor shall any inferences be drawn, used, or communicated regarding the possible genetic status of the non-consenting individual.").

penalties, such as fines or imprisonment, for violations of the statute.[199] Unfortunately, these states statutes generally provide exceptions for anonymized genetic research, and therefore would not apply to de-identified public genomic data.[200]

E. National Institutes of Health Genomic Data Sharing Policy

The National Institutes of Health (NIH) recently issued the NIH Genomic Data Sharing Policy, which applies to larger NIH-funded genomic studies with 100 or more subjects.[201] The new policy states that genomic data must be de-identified by the removal of the eighteen HIPAA identifiers.[202] Researchers may release the data through either unrestricted access or controlled access methods.[203] The risk of re-identification must be explicitly conveyed to the subjects, particularly for open access data repositories.[204]

The NIH policy also sets forth data release policies and the responsibilities of the users accessing the publicly available genomic data.[205] Users seeking access to the controlled data must submit a request, and upon

---

199. *See* ALASKA STAT. § 18.13.020 (2014) ("In addition to the actual damages suffered by the person, a person violating this chapter shall be liable to the person for damages in the amount of $5,000 or, if the violation resulted in profit or monetary gain to the violator, $100,000."); CAL. INS. CODE §10149.1 (West 2013) ("Any person who negligently discloses results of a test for a genetic characteristic . . . shall be assessed a civil penalty in an amount not to exceed one thousand dollars ($1,000) plus court costs[.]"); DEL. CODE ANN. 16 § 1208 (2014) ("Any person who willfully obtains or discloses genetic information in violation of this subchapter shall be punished by a fine not less than $5,000 nor more than $50,000."); MO. REV. STAT. § 374.049 (2013) ("An order to impose a civil penalty or forfeiture . . . One thousand dollars per each level two violation, up to an aggregate civil penalty or forfeiture of fifty thousand dollars per annum for multiple violations"); OR. REV. STAT. § 192.541 (2013) (setting forth different levels of fine if greater than actual damages, depending on the intent of the violator); VT. STAT. ANN. tit. 18 § 9335 (2014) ("Any person who intentionally violates section 9333 or subsection 9334(a) of this chapter shall be imprisoned not more than one year or fined not more than $10,000.00, or both.").

200. *See* Prince, *supra* note 178, at 206–07.

201. Final NIH Genomic Data Sharing Policy, 79 Fed. Reg. 51,345, 51,345 (Aug. 28, 2014); U.S. DEP'T OF HEALTH & HUMAN SERVS., SUPPLEMENTAL INFORMATION TO THE NIH GENOMIC DATA SHARING POLICY (2014), http://gds.nih.gov/pdf/supplemental_info_GDS_Policy.pdf [http://perma.cc/S88K-4PVN].

202. Final NIH Genomic Data Sharing Policy, 79 Fed. Reg. at 51,345. See *supra* note 129 for a list of HIPAA identifiers.

203. Final NIH Genomic Data Sharing Policy, 79 Fed. Reg. at 51,351. For a more detailed discussion on access control, see *infra* Part V.C.1.

204. Final NIH Genomic Data Sharing Policy, 79 Fed. Reg. at 51,348.

205. *Id.* at 51,352. *See also Genomic Data User Code of Conduct*, NAT'L INSTS. OF HEALTH (Apr. 2, 2010), http://gds.nih.gov/pdf/Genomic_Data_User_Code_of_Conduct.pdf [http://perma.cc/YP63-K3W4] (establishing a code for researchers using dbGaP).

grant, they can only use the data for approved research.[206] Approved users cannot attempt re-identification or provide the data to unauthorized users.[207] Overall, this policy acknowledges the risk of re-identification and emphasizes the requirement of informed consent that includes disclosure of access levels and risks to the subjects.[208]

F. International Guidelines

Many of the public genomics projects are international, so current legislation in the United States does not apply to all data from public genomics projects.[209] Some countries or jurisdictions have general privacy laws which encompass genetic information.[210] For example, the European Union's Data Protection Directive offers privacy protection for sensitive data regulated by data commissions, such as health-related data.[211] The directive covers any type of "personal data" that can be "directly or

206. Final NIH Genomic Data Sharing Policy, 79 Fed. Reg. at 51,352. *See also* Paltoo et al., *supra* note 25, at 935–36 (describing the access process for dbGaP). Access requests are reviewed by NIH Data Access Committees "composed of senior Federal employees with appropriate scientific, bioethics, and human subjects' research expertise." *NIH Data Access Committees and Chairs*, NAT'L INSTS. OF HEALTH, http://gds.nih.gov/04po2_1DAC.html [http://perma.cc/A42X-R6SF] (last visited Aug. 3, 2015). The policy additionally recommends that investigators obtain a Certificate of Confidentiality as a safeguard to prevent forced disclosure of the accessed genomic information. Final NIH Genomic Data Sharing Policy, 79 Fed. Reg. at 51,351–52 (Aug. 28, 2014). Certificates of Confidentiality are issued by NIH to protect identifiable research information, including genomic and clinical data, from forced disclosure in court or other proceedings. *Certificates of Confidentiality (CoC) Kiosk*, NAT'L INSTS. OF HEALTH, http://grants.nih.gov/grants/policy/coc/index.htm [http://perma.cc/ZSL8-ZMLY] (last updated July 24, 2015). For a discussion of the limits of the Certificate of Confidentiality, see Brown & Lowenberg, *supra* note 29, at 93–97.

207. Final NIH Genomic Data Sharing Policy, 79 Fed. Reg. at 51,352.

208. *See id.* at 51,348, 51,351.

209. RESPONSIBLE SHARING, *supra* note 117, at 12.

210. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 65. For example, the European Union has a Directive that provides strict overarching privacy protection. Council Directive 95/46, art. 1–2, 1995 O.J. (L 281) 31, 38 (EC).

211. Council Directive 95/46, art. 2, 1995 O.J. (L 281) 38 (EC). The European Union has strict privacy protection laws, and personal data can only be obtained "under strict conditions, for a legitimate purpose." *Protection of Personal Data*, EUR. COMMISSION, http://ec.europa.eu/justice/data-protection/ [http://perma.cc/6CUS-45GU] (last updated Aug. 27, 2015). "Member States shall prohibit the processing of *personal data* revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life." Council Directive 95/46, art. 8(1), 1995 O.J. (L 281) 40 (EC) (emphasis added).

indirectly" linked to a person, by any administrator, including public and private institutions.[212] However, the directive excludes public data, and is criticized as overreaching because all data that is potentially re-identifiable, regardless of the presence of other protective measures, can be regulated under this directive.[213] Other countries specifically have laws regulating genetic information and research, including consent, and anonymization.[214] General international guidelines and principles, including the Nuremberg Code for human subject consent,[215] the United Nations International Declaration on Human Genetic Data and Universal Declaration on the Human Genome and Human Rights[216] serve as general principles in the research community. Additionally, the international research community is working to achieve self-governing guidelines for genomic data sharing and data security.[217] Although these are not legally binding regulations, the scientific research and medical communities are actively discussing

---

212. Ohm, *supra* note 9, at 1738.

213. Council Directive 95/46, art. 8(2)(e), 8(3), 1995 O.J. (L 281) 38 (EC). The Directive excludes data processing that is

> manifestly made public by the data subject or is necessary for the establishment, exercise or defence of legal claims . . . required for the purposes of preventive medicine, medical diagnosis, the provision of care or treatment or the management of health-care services, and where those data are processed by a health professional . . . .

Id.; Ohm, *supra* note 9, at 1738, 1741.

214. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 65. For example, Chile has a law that regulates genetic research and prohibits discrimination on basis of genetic heritage. The law additionally sets forth the confidentiality of genetic information and anonymization of genetic data. Law No. 20120, Septiembre 22, 2006, DIARIO OFICIAL [D.O.] (Chile). Estonia, France, Spain, Germany and Israel also have similar laws. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 65, 133, n.120.

215. 2 U.S. GOV'T PRINTING OFFICE, TRIALS OF WAR CRIMINALS BEFORE THE NUREMBERG MILITARY TRIBUNALS UNDER CONTROL COUNCIL LAW NO. 10, at 181–82 (1949), http://www.loc.gov/rr/frd/Military_Law/pdf/NT_war-criminals_Vol-II.pdf [http://perma.cc/S7XD-UZ4K].

216. Educational, Scientific, and Cultural Organization Res. 2003/22, International Declaration on Human Genetic Data, (Oct. 16, 2003), http://unesdoc.unesco.org/images/0013/001331/133171e.pdf#page=45 [http://perma.cc/F7C4-K68P]; Educational, Scientific, and Cultural Organization Res. 1997/16, Universal Declaration on the Human Genome and Human Rights (Nov. 11, 1997), http://portal.unesco.org/en/ev.php-URL_ID=13177&URL_DO=DO_TOPIC&URL_SECTION=201.html [http://perma.cc/53WB-D9JP].

217. RESPONSIBLE SHARING, *supra* note 117, at 1, 6; GLOBAL ALLIANCE FOR GENOMICS & HEALTH, *Security Infrastructure: Standards and Implementation Practices for Protecting the Privacy and Security of Shared Genomic and Clinical Data*, 6–12 (2014), http://genomicsandhealth.org/files/public/(Sept26-CLEAN)DRAFT_SecurityFramework.pdf [http://perma.cc/J7R8-LTYF].

and collaborating to build self-governing guidelines for genomic data to minimize potential privacy breach.[218]

V. MULTIMERIC SOLUTION TO APPROACH PRIVACY ISSUES

To address the emerging privacy issues from re-identification of publicly available genomic data, a multi-pronged solution must be implemented due to the complexity of privacy issues and the numerous stakeholders involved.[219] The solution to prevent re-identification and minimize the risks to family and relatives should focus on strengthening protection by banning malicious re-identification and providing broader anti-discrimination protection. The issue of consent should be addressed by updates to the Common Rule and through participant education. Although legislation alone cannot solve the multitude of potential issues,[220] it will provide a solid foundation to prevent privacy breaches for participants who have risked their privacy for the benefit of scientific progress. Complementary self-governance approaches by the research community, such as access control and data standardization, are required to address some of the difficulties in providing a basic level of protection globally.[221]

---

218. *See* Altman, *supra* note 76, at 1033; P3G Consortium et al., *supra* note 25, at 1–4; GLOBAL ALLIANCE FOR GENOMICS & HEALTH, http://genomicsandhealth.org/ [http://perma.cc/KZ2R-S2H8] (last visited July 25, 2015).

219. RESPONSIBLE SHARING, *supra* note 117, at 6–12 (listing potential stakeholders, including research participants, researchers, data custodians, clinicians, the public, regulatory agency, organizer of public consortia, organizers of crowd-sourcing projects, and private direct-to-consumer providers.)

220. *See*, PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 4–11. In particular, legislation sometimes lags behind the technological developments. *See*, *e.g.*, John Burn-Murdoch, *Data Protection Law Is in Danger of Lagging Behind Technological Change*, GUARDIAN (Apr. 12, 2013, 7:25 AM), http://www.theguardian.com/news/datablog/2013/apr/12/data-protection-law-lagging-behind-technology [http://perma.cc/68QC-2XWB]; *State of Federal Privacy and Data Security Law: Lagging Behind the Times?*: *Hearing Before the Subcomm. On Oversight of Gov't Mgmt., the Fed. Workforce, and D.C. of the S. Comm. On Homeland Sec. & Governmental Affairs*, 112th Cong. 19–21 (2012) (statement of Peter Swire, C. William O'Neill Professor of Law, Ohio State Univ.), http://www.gpo.gov/fdsys/pkg/CHRG-112shrg76066/pdf/CHRG-112shrg76066.pdf [http://perma.cc/5W4Q-UVQC].

221. *See*, *e.g.*, RESPONSIBLE SHARING, *supra* note 117, at 9–10.

A. Addressing Re-identification and Effects on Relatives

1. Ban on Malicious Re-identification

Some privacy experts have proposed a proactive ban on re-identification which would establish various penalties for re-identification except in cases of approved re-identification research.[222] For example, privacy policy consultant Robert Gellman proposed a legislation that "establishes standards for behavior and proposes civil and criminal penalties for violations."[223] In this proposal, "data recipients would also be required to maintain technical, administrative, and other safeguards against re-identification."[224] Gellman proposes a "legislative-based contractual solution" where the data discloser and the data recipient voluntarily enter into a contract that defines the responsibilities of each party, and offers remedies if one of the parties injures the data subject.[225] In exchange for the responsibilities, the data recipients "would benefit [from] being able to offer potential disclosers more assurance that a data transfer will not create liabilities."[226] Regardless of whether criminal punishment is appropriate, imposing penalties would help deter malicious adversaries from re-identification efforts.[227]

In addition, the ban should clearly provide the participants a cause of action and a remedy if the participant suffers a consequence of malicious re-identification,[228] such as in Krypton's denial of insurance in the hypothetical. As current medical privacy protection under HIPAA does not provide a

---

222. Barth-Jones, *supra* note 80, at 14; Robert Gellman, *Deidentification Dilemma: A Legislative and Contractual Proposal*, 21 FORDHAM INTELL. PROP., MEDIA & ENT. L.J. 33 (2010); Ohm, *supra* note 9, at 1758, n.120.

223. Gellman, *supra* note 222, at 48, 59 (proposing a civil penalty for persons who fail to report a breach of data, and a fine or imprisonment for those who willfully re-identifies or sells re-identified information).

224. *Id.* at 49.

225. A data discloser would be someone who initially has the data, such as those running the public genomics projects. *Id.* at 48. A data recipient would be someone who is looking to have access to the data, such as a scientist seeking to obtain the existing data for future analysis. *Id.* A data subject would be someone whose information is contained in the data, such as participants of public genomics projects. *Id.* at 35.

226. *Id.* at 49.

227. *Id.* at 53. Penalties can include civil fines similar to those set forth in state statutes for genetic privacy violations. *Id.*

228. *See id.* at 53. Generally, HIPAA does not provide a private cause of action for cases of privacy violation. Acara v. Banks, 470 F.3d 569, 570 (5th Cir. 2006); Steinberg v. CVS Caremark Corp., 899 F. Supp. 2d 331, 337 (E.D. Pa. 2012); Carol Gerner & Fred Smith, *HIPAA Does Not Provide Right of Action in Privacy Violation Cases*, HEALTHCARE L. NEWSL. (Apr. 2007), http://www.sedgwicklaw.com/Publications/detail.aspx?pub=4680 [http://perma.cc/5NMW-6PCE]. Plaintiffs often bring actions under violation of related state laws. *See*, *e.g.*, *Steinberg*, 899 F. Supp. 2d at 337–39; Acosta v. Byrum, 638 S.E.2d 246, 253 (N.C. Ct. App. 2006).

cause of action for participants,[229] a right of action should be provided for privacy violations under HIPAA.[230]

The re-identification ban should encompass public information, because re-identification concerns are most prominent for such information.[231] Public information does not require a security breach to obtain, but adversaries still can re-identify the information for privacy breach or discrimination.[232] This type of ban would be broader than Gellman's contractual approach between the data discloser and data recipient because any member of the public can be a data recipient without a contract.[233] Privacy scholar Daniel Barth-Jones suggests "[p]rohibiting of the re-identification, or attempted re-identification, of individuals and their relatives, family or household members . . . establish[ing] civil and criminal penalties for any unauthorized re-identification of de-identified data."[234] To protect participants who provided their information to public genomics projects, the re-identification ban should also apply to information that is already public.

A re-identification ban may be difficult to enforce when the data is already publicly available and because detecting acts of re-identification is difficult.[235] However, stricter penalties and providing a right of action and clear remedies for citizens whose privacy was violated by re-identification would bolster deterrence and help overcome the difficulties in enforcement.[236] In addition, if clear access control and audit processes are implemented,[237] re-identification efforts may be easier to detect and a ban may be easier to enforce.[238]

A restriction on re-identification, however, should not prohibit re-identification research.[239] Part of the solution requires an accurate risk

229. *See supra* note 228.
230. *See, e.g.*, Ohm, *supra* note 9, at 1758.
231. *See id.* at 1717.
232. *See*, *e.g.*, PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 14–16; Rothstein, *supra* note 87, at 8; Ohm, *supra* note 9, at 1717.
233. *See generally* Gellman, *supra* note 222.
234. Barth-Jones, *supra* note 80, at 14.
235. Ohm, *supra* note 9, at 1758.
236. *Id.*
237. *See supra* Part V.C a for a more detailed discussion on access control.
238. Ohm, *supra* note 9, at 1758.
239. Salil Vadhan et al., Comment Letter on Proposed Rule on Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators (Oct. 26, 2011), http://privacytools.seas.harvard.edu/files/privacytools/files/commonruleanprm.pdf [http://perma.cc/CQV5-2MZL].

assessment of re-identification.[240] Analogous to financial privacy or social network data privacy, there are risks and benefits associated with the use of the technology and making information public.[241] The bulk of the risk assessment comes from re-identification research, and therefore re-identification for research purposes should be an exception to the ban.[242] Risk assessment should also include determining actual realistic probabilities that a non-expert can be successful in re-identification.[243] Updated risk assessments are also necessary as new technologies develop and as more information, both genomic data and other personal information, becomes public.[244] Accurate risk assessment and risk management can reduce fears of privacy violations and help encourage participation in public genomics projects.[245]

Therefore, the legislature should prioritize prohibiting malicious re-identification as a solution to protect participant privacy, while allowing data scientists to accurately determine additional re-identification risks. The re-identification ban should also apply to public information, and should be complemented by other legislative approaches, such as broader anti-discrimination protection.

---

240. Jean E. McEwen et al., *Evolving Approaches to the Ethical Management of Genomic Data*, 29 TRENDS GENETICS 375, 378–79 (2013); C. Heeney et al., *Assessing the Privacy Risks of Data Sharing in Genomics*, 14 PUB. HEALTH GENOMICS 17, 17 (2011), http://www.karger.com/Article/Pdf/294150 [http://perma.cc/8HEB-JW2U].

241. Bruce R. Korf, *Genomic Privacy in the Information Age*, 59 CLINICAL CHEMISTRY 1148, 1149 (2013). Despite the risks of identity thieves and publicized security breaches, credit cards are still used and personal information is voluntarily published in social networking websites and every day. *Id.* Therefore, with accurate risk assessment and protective measures, encouragement of participation in public genomics is possible. *Id.* In fact, it is still much easier and effective to hack into financial information rather than health related or genomic information, so the incentives for health information hacking or re-identification may be smaller. Ohm, *supra* note 9, at 1767.

242. Erlich & Narayanan, *supra* note 22, at 420. Careful risk assessment based on risks and implications of actual cases of re-identification is necessary. *Id.* Barth-Jones suggests a re-identification ban that allows "Institutional Review Board (IRB) approved reidentification research to be conducted, but would ban any re-identification attempts conducted without essential human subjects research protections." Barth-Jones, *supra* note 80, at 14.

243. Erlich & Narayanan, *supra* note 22, at 420.

244. *See* David W. Craig et al., *Assessing and Managing Risk When Sharing Aggregate Genetic Variant Data*, 12 NATURE REVIEWS GENETICS 730, 730 (2011); Korf, *supra* note 241, at 1149; Ohm, *supra* note 9, at 1705; J.M. Oliver et al., *Balancing the Risks and Benefits of Genomic Data Sharing: Genome Research Participants' Perspectives*, 15 PUB. HEALTH GENOMICS 106 (2012).

245. *See* Erlich & Narayanan, *supra* note 22, at 419–20.

2. Broader Anti-Discrimination Legislation

GINA and some state laws provide protection against certain types of discrimination based on genetic data.[246] However, the protection excludes certain information, such as manifested symptoms.[247] This protection is also limited to employment and health insurance, and does not apply to other situations such as life insurance, disability and long-term care.[248] Although GINA provides a good minimum federal protection against discrimination based on genetic information, it fails to cover other potentially discriminatory situations, and the existence and scope of complementary state laws vary widely.[249] GINA should expand to fill the gaps for anti-discrimination, such as in cases of life insurance, disability and long-term care.[250] GINA should also specifically provide patients with protection against discrimination using re-identified data.[251]

3. Data Privacy Protection Legislation

GINA does not address privacy issues, but the HIPAA Privacy Rule provides some measure of privacy protection for personally identifiable information.[252] In addition to HIPAA, general privacy laws, such as those provided by state constitutional privacy provisions or statutes, can also be a means to provide fundamental privacy protection.[253] These laws may bolster the protection against re-identification threats,[254] in addition to

246. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 77–80.

247. Jessica L. Roberts, *Preempting Discrimination: Lessons from the Genetic Information Nondiscrimination Act*, 63 VAND. L. REV. 439, 455–56 (2010). Under GINA, a test that detects a condition that has manifested itself (a test to determine symptomatic persons) is not a "genetic test." *Id.* (citing 29 U.S.C. § 1191b(d)(7)(B) (Supp. II 2009)).

248. Shoenbill et al., *supra* note 161, at 174; Slaughter, *supra* note 146, at 54.

249. *See supra* Part IV., pts. 1–5.

250. *See, e.g.*, PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 77; Morrison, *supra* note 161, at 584; Shoenbill et al., *supra* note 161, at 174.

251. *See, e.g.*, PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 77–80; Morrison, *supra* note 161, at 597–98; Shoenbill et al., *supra* note 161, at 177.

252. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 66–67.

253. *See supra* Part IV.D.

254. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 4–10.

serving as a ban on malicious re-identification and a strong anti-discrimination measure.

There are two main limitations with HIPAA regarding public genomics. First, although the HIPAA list of identifiers includes "biometric identifiers,"[255] HIPAA does not state whether this list includes genomic information.[256] Furthermore, the entities covered in HIPAA do not include many entities that manage public genomics projects, such as academic institutions, federal agencies and scientific consortia.[257] Although they are not entities covered by HIPAA, they must comply with the privacy protection requirements of the Common Rule.[258] Currently, the de-identification requirements for the Common Rule are less strict than for HIPAA.[259] The proposed updates to the Common Rule advocates for distinguishing identifiable and de-identified information and adopting the HIPAA standards for de-identification.[260] For consistency with HIPAA and participant privacy protection, the Common Rule should follow HIPAA standards for privacy, and factor in the potential risks of re-identification and future research.[261]

---

255. U.S. DEP'T OF HEALTH & HUMAN SERVS., GUIDANCE REGARDING METHODS FOR DE-IDENTIFICATION OF PROTECTED HEALTH INFORMATION IN ACCORDANCE WITH THE HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT (HIPAA) PRIVACY RULE 23 (2012), http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/hhs_deid_guidance.pdf [http://perma.cc/8FGS-WUWK] (explaining that biometric identifiers include fingerprints and voiceprints).

256. INST. OF MED. OF THE NAT'L ACADS., *supra* note 171, at 180, 190. In a proposed update to the Common Rule regarding human subject research, the Department of Health and Human Services contemplated including biospecimen, from which genomic DNA can be extracted and sequenced, to be re-identifiable information. *See supra* Part IV.C; Human Subject Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators, 76 Fed. Reg. 44,512, 44,524 (July 26, 2011).

257. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 63–64.

258. *OHRP Guidance*, *supra* note 167. Academic institutions and consortia that are governmentally funded are subject to the Common Rule for human subject research. *Id.*; *see supra* Part IV.C.

259. *OHRP Guidance*, *supra* note 167. See *supra* Part IV.C for a comparison of HIPAA and Common Rule Standards.

260. Human Subject Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators, 76 Fed. Reg. 44,512, 44,525 (July 26, 2011).

261. *Id.*; Prince, *supra* note 178, at 206–07.

B. Addressing the Consent Issue

1. Updates to the Common Rule

Additional consent for future projects must be addressed separately from the re-identification ban.[262] The Common Rule, which currently governs federally funded human research, does not require IRB review or additional consent for studies using data that are stripped of HIPAA identifiers.[263] The Common Rule should be updated to address future uses and re-identification concerns for public genomics. For example, if the initial consent did not include information about privacy risks or when it can be used for different studies, a new consent should be obtained when possible.[264] In the past, some program participants provided blanket consent, consent without a limitation on the scope and duration,[265] but this type of consent is not fully informational because the participants are not informed of future developments or uses.[266] Some programs employ opt-out models, which advise participants that they will be involved in a future research study and to notify the program administrators only if they do not wish to participate.[267] However, although it helps increase participation, opt-out consent is often not fully informing, and patients prefer the opt-in option.[268]

---

262. Protection of Human Subjects, 45 C.F.R. § 46 (2014); Conley et al., *supra* note 15, at 351–53; Korenman, *supra* note 164.

263. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 63–64; *OHRP Guidance*, *supra* note 167.

264. International Declaration on Human Genetic Data, *supra* note 216 (*"*Human genetic data, human proteomic data and the biological samples collected for one of the purposes . . . should not be used for a different purpose that is incompatible with the original consent, unless the prior, free, informed and express consent of the person concerned is obtained . . . .); Lunshof et al., *supra* note 104. However, an exception should be provided for cases where obtaining renewed consent is unduly burdensome, such as when the participant is since deceased. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 91.

265. Lunshof et al., *supra* note 104, at 408.

266. *Id.*

267. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 92.

268. *See* Margaret F.A. Otlowski, *Tackling Legal Challenges Posed by Population Biobanks: Reconceptualising Consent Requirements*, 20 MED. L. REV. 191, 213 (2012).

After the publication of successful re-identification efforts, the HapMap Project and the 1000 Genomes project underwent re-consent procedures.[269] The re-consent procedures included removal of the participants' year of birth from the data and informing the participants of the inability to guarantee privacy.[270] Although obtaining re-consent for public and de-identified information may be costly,[271] updated consent should be obtained if the participants can be reached.[272]

2. Open Consent

Open consent is the Personal Genome Project (PGP)'s solution to address evolving consent issue.[273] Some of the other public genomics projects' consent forms include information about re-identification and other future risks.[274] The PGP is significantly more upfront about the privacy risks and does not guarantee privacy to the participants.[275] Instead

---

269. Laura L. Rodriguez et al., *The Complexities of Genomic Identifiability*, 339 SCIENCE 275, 275 (2013).

270. *Id.* at 276.

271. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 91; Lunshof et al., *supra* note 104, at 408. Obtaining re-consent also poses additional risk of privacy breach. Lunshof et al., *supra* note 104, at 408. For example, to contact the subject again for additional consent, the database must retain contact information about the subject, which can be vulnerable to adversaries aiming to re-identify, because the contact information is stored within the database. *See id.*; Morrison, *supra* note 161, at 601; Otlowski, *supra* note 268.

272. RESPONSIBLE SHARING, *supra* note 117, at 14.

273. McEwen et al., *supra* note 240, at 378–79.

274. *See*, *e.g.*, *Consent to Participate*, 1000 GENOMES, http://www.1000genomes.org/sites/1000genomes.org/files/docs/Informed%20Consent%20Form%20Template.pdf [http://perma.cc/EU4S-DM5G] (last visited Aug. 3, 2015). The 1000 Genomes participant consent form warns participants that "[a]s technology advances, there may be new ways of linking information back to you that we cannot foresee now. Also, we cannot always foresee the results of research, so new risks may come up in the future that we cannot predict now." *Id.* It asks the participants to weigh the benefits of discovery from genome research against potential risks themselves. *Id.* In contrast, the Coriell Personalized Medicine Collaborative states that participants have more control, and does not discuss currently unknown future risks. *CPMC FAQs*, CORIELL PERSONALIZED MED. COLLABORATIVE, http://cpmc1.coriell.org/about-the-cpmc-study/cpmc-faqs [http://perma.cc/RHF9-UMTL] (last visited Aug. 3, 2015) ("CPMC participants control access to their CPMC information through the secure web portal. Your results will not be shared with anyone, for example, your doctors, family, friends, employer, insurance company or anyone else, without your permission.").

275. *See* Madeleine P. Ball et al., *Harvard Personal Genome Project: Lessons from Participatory Public Research*, 6 GENOME MED. 10, 10–11 (2014), http://www.genomemedicine.com/content/pdf/gm527.pdf [http://perma.cc/HF9H-QV92]; Church, *supra* note 14, at 53; Conley et al., *supra* note 15, at 229–30. The Personal Genome Project (PGP) takes the stance that privacy cannot be guaranteed, and "abstains from any assurance to participants of privacy or anonymity." Ball et al., *supra*, at 10. Some data scientists

of relying on privacy guarantees, the PGP relies on "information altruists"[276] to bear the risks of privacy breaches.[277] The PGP focuses on rigorous privacy education and active participation.[278] After enrollment and data generation, interaction with the participants continues through the website, allowing participants to review and add data, and build an ongoing relationship with the researchers.[279] In turn, the PGP grants the participants more control over how much information they will make public.[280] However, once the participant decides to make the information public, it is available without any access control for the data users.[281]

As exemplified in the PGP, the open consent approach can be successfully implemented to address the need for additional or updated consent for public genomics projects.[282] In particular, it can be applied to public

successfully identified PGP participants, who had only some of the information available online at the PGP website, by name. Latanya Sweeney et al., *Identifying Participants in the Personal Genome Project by Name*, ARXIV:1304.7605 (Apr. 29, 2013), http://arxiv.org/abs/1304.7605 [http://perma.cc/7U28-RT9G].

276. "Information altruists" are those who understand and fully accept the privacy risk yet make their information publicly available. Conley et al., *supra* note 15, at 354. These "information altruists" must pass a rigorous enrollment exam, which tests how well informed the potential participants are about the study and its risks, including the privacy risks to the subjects themselves and their relatives. Lunshof et al., *supra* note 104, at 409. Participants must answer all questions correctly. Ball et al., *supra* note 275, at 2. However, simply relying on information altruism without providing other risk-mitigation measures may also reduce participation because not everyone is comfortable with freely providing their information without assurances. McEwen et al., *supra* note 240, at 378–79. Another criticism of the information altruism approach is that having a subset of population selected as "information altruists" may bias the data. Isaac S. Kohane & Russ B. Altman, *Health-Information Altruists—A Potentially Critical Resource*, 353 NEW ENG. J. MED. 2074, 2075 (2005).

277. Ball et al., *supra* note 275, at 2; Conley et al., *supra* note 15, at 354.

278. Ball et al., *supra* note 275, at 6; Conley et al., *supra* note 15, at 354.

279. Ball et al., *supra* note 275, at 6.

280. *Id.* at 4–5. Some participants have identified themselves, such as the founder George Church, and others have published stories and articles about their own genome. *See*, *e.g.*, John Lauerman, *Harvard Mapping My DNA Turns Scary as Threatening Gene Emerges*, BLOOMBERG BUS. (Feb. 14, 2012, 9:01 PM), http://www.bloomberg.com/news/2012-02-15/harvard-mapping-my-dna-turns-scary-as-threatening-gene-emerges.html [http://perma.cc/FL6H-JYCR]; Steven Pinker, *My Genome*, *My Self*, N.Y. TIMES (Jan. 7, 2009), http://www.nytimes.com/2009/01/11/magazine/11Genome-t.html?pagewanted=all [http://perma.cc/FHX5-8CM5].

281. Ball et al., *supra* note 275, at 6.

282. *Id.* at 1; Conley et al., *supra* note 15, at 354. However, the open consent approach may not be feasible or suitable for all public genomics projects. McEwen et al., *supra* note 240, at 379. The PGP's approach sets the bar of participation higher and reduces the number of potential participants, due to the large number of hurdles to qualify.

driven crowd-sourced projects, where transparency about privacy risks and participant education is critical.[283] Open consent and participant-centric initiatives can be adopted in other projects for uses such as participant education and updating consent for future studies or future developments.[284] For example, Portable Legal Consent, an interactive informed consent document and software that can be used for dynamic genomic projects, was built based on PGP's approach, and can easily be adopted for other participatory genomics projects.[285]

3. Participant-centric Initiatives

As a part of an effort to standardize genomic data and develop systems for security, storage, and access of genomic data, groups such as the Global Alliance for Genomics and Health[286] are also developing tools for participant-centric initiatives.[287] Participant-centric initiatives refers to "tools, programs and projects that empower participants to engage in the research process using information technology," and includes easy interfaces that participants can access to provide consent, obtain data, and

---

Id. at 377. One study showed that PGP was able to enroll only close to 50% of the requesters, with approximately 30% of the requesters not completing or failing the enrollment exam. Ball et al., *supra* note 275, at 2. Additionally, this may limit participation because many participants desire some type of access control. Amy L. McGuire et al., *To Share or Not To Share: A Randomized Trial of Consent for Data Sharing in Genome Research*, 13 GENETICS IN MED. 948, 948 (2011), http://www.nature.com/gim/journal/v13/n11/pdf/gim2011159a.pdf [http://perma.cc/PZT3-YJYR].

283. Ball et al., *supra* note 275, at 6; Heeney et al., *supra* note 240, at 17; Muhammad Naveed et al., *Privacy and Security in the Genomic Era*, ARXIV:1405.1891, 26 (June 17, 2015), http://arxiv.org/abs/1405.1891 [http://perma.cc/HN2A-97JU]. Although some crowd-sourced projects do address the issue of consent, consent and education is not as extensive as projects that more closely monitor subject participation and consent, such as the PGP, because crowd-sourced projects are open to a significantly wider population and participation is less controlled. *See* Greshake et al., *supra* note 70.

284. Ball et al., *supra* note 275, at 43; Conley et al., *supra* note 15, at 4; Gutmann Koch, *supra* note 49, at 9. See *infra* Part V.B.3, for a discussion on participant-centric initiatives.

285. *Portable Consent*, SAGE BIONETWORKS, http://sagebase.org/category/portable-consent/ [http://perma.cc/7QED-DLGD] (last visited Aug. 3, 2015).

286. The Global Alliance for Genomics and Health is an international coalition of research institutions, healthcare providers, funding agencies, disease advocacy groups, and biotech and information technology companies involved in the study of genomics and genomic medicine. *Frequently Asked Questions*, GLOBAL ALLIANCE FOR GENOMICS & HEALTH, http://genomicsandhealth.org/about-the-global-alliance/frequently-asked-questions/ [http://perma.cc/7WD9-LQSV] (last visited Apr. 16, 2015). The Global Alliance aims to maximize the potential of genomic data by promoting effective and responsible genomic data sharing and establishing standards to "enable the responsible, voluntary, and secure sharing of genomic and clinical data." *Id.*

287. RESPONSIBLE SHARING, *supra* note 117, at 25.

customize data access control.[288] Some public genomics or other public health projects have successfully used these interfaces and platforms.[289] Developing participant-centric initiatives requires a concerted effort by the scientific community.[290] Current efforts include developing common customizable platforms, such as the Platform for Engaging Everyone Responsibly for clinical information and biological specimen,[291] and standardizing sequencing data formats, to be compatible with many sequencing platforms and cloud storage and to easily develop user interfaces.[292] User interfaces can form a social network-like structure, be used to obtain updated consent as necessary, and facilitate the interaction between the participants and the scientists accessing the data.[293] Particularly for crowd-sourced projects where risk and benefit education is not as extensive as other public genomics projects,[294] these interfaces can be adopted for better education and information.[295] This type of approach should be used for international projects where the local regulation, participant attitude, and awareness differ because it can be used to customize data release to each participant's comfort level.[296]

C. Addressing Uniform Guidelines: Data Standardization and Management

1. Access Control and Standardization

Restricting data access can be one way to minimize privacy breach by screening out potential adversaries and misusers, without significant costs to the researchers that want to access the data for legitimate research

288. *Id.*
289. Kaye et al., *supra* note 112, at 372. For examples of public genomics or other public health projects, such as PGP and PatientsLikeMe, see *supra* Part II.B.
290. RESPONSIBLE SHARING, *supra* note 117, at 25.
291. *Platform for Engaging Everyone Responsibly*, GENETIC ALLIANCE, http://geneticalliance.org/programs/biotrust/peer [http://perma.cc/A49M-FWUR] (last visited July 25, 2015).
292. RESPONSIBLE SHARING, *supra* note 117, at 25. See *infra* Part V.C.1 for further discussion on data standardization efforts.
293. RESPONSIBLE SHARING, *supra* note 117, at 10, 16.
294. *See*, *e.g.*, Greshake et al., *supra* note 70. See *supra* Part II.B. for further discussion on participatory genomics projects.
295. Kaye et al., *supra* note 112, at 374.
296. RESPONSIBLE SHARING, *supra* note 117, at 12.

reasons.[297] Access control measures essentially make public genomics less *public* by screening and controlling users who can use, own or download the data.[298] This approach can be a complementary solution provided by the research community to balance data availability and privacy protection.[299]

Reasonable access control is preferred by some research participants as a means to mitigate privacy.[300] One study of subject consent for data sharing in the United States found that after a debriefing of three data release options–public release, restricted access, or no data release–approximately 53% of the subjects chose public release, 33% chose restricted access, and 14% chose no release.[301] This study shows that at least a third of the participants prefer some type of access control for the data.[302] Attitudes towards data sharing also differ in different countries and demographic groups.[303] Generally, there is a willingness to participate in public genomics, but with some reservations about privacy concerns regarding personal data.[304] Access control may provide relief to some of these reservations of to encourage participation.[305] Additionally, a recent

297. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 80–81.

298. Conley et al., *supra* note 15, at 338–40.

299. *See, e.g.*, PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 80–81; Erlich & Narayanan, *supra* note 22, at 416–17; *dbGaP*, NAT'L CTR. FOR BIOTECHNOLOGY INFO., http://www.ncbi.nlm.nih.gov/gap [http://perma.cc/E5SA-5BAD] (last visited July 25, 2015).

300. *See* McGuire et al., *supra* note 282, at 948; Naveed et al., *supra* note 283, at 26–27.

301. McGuire et al., *supra* note 282, at 952.

302. *See id.*

303. *Id.* at 950–53; Amy L. McGuire et al., *DNA Data Sharing: Research Participants' Perspectives*, 10 GENETICS IN MED. 46, 46–50 (2008); RESPONSIBLE SHARING, *supra* note 117, at 7.

304. RESPONSIBLE SHARING, *supra* note 117, at 7.

305. *See* PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 80–81; Kaye et al., *supra* note 112, at 373–75. A criticism of access control is that it provides unnecessary hindrance to research. Conley et al., *supra* note 15, at 339–40. Open source data advocates argue that providing open access is necessary to "maximize the potential for discovery," as access control can serve as a barrier to more scientists analyzing the data and making discoveries based on the data. *See, e.g.*, Church, *supra* note 14; Greshake et al., *supra* note 70. They argue that restricting access to data from public genomics projects makes these projects not "public," and does not help "democratize[] genomics research." Conley et al., *supra* note 15, at 339–40. For example, the level of access for the open access projects, such as the HapMap and 1000 Genomes projects, is substantially higher than dbGaP, a controlled access database containing a related dataset. Rodriguez et al., *supra* note 269, at 276. Another criticism is that controlling access provides a false sense of security because genomic data cannot be completely de-identified and access control cannot guarantee privacy. Ball et al., *supra* note 275, at 1; Conley et al., *supra* note 15, at 339–40; Lunshof et al., *supra* note 104, at 408. As discussed above in Part V.B.2, the Personal Genome Project (PGP) does not guarantee privacy for the

survey of genomics and genetics experts showed that a majority of the experts are willing to bear monetary and time costs to protect participant privacy.[306] Furthermore, only 7% of the experts believed that "privacy enhancing technologies are a nuisance in the case of genetics,"[307] indicating that access control can be a reasonable solution that to provide privacy protection for the participants.[308]

One example of the access control approach is taken by the database of Genotypes and Phenotypes (dbGaP) hosted at the National Center for Biotechnology Information.[309] A user requesting access to dbGaP must apply, state their specific uses for the information, and receive approval before accessing the data.[310] A study of dbGaP user access published in 2013 identified no privacy breaches.[311] After the publication of some of the re-identification studies, institutions such as the National Human Genome Research Institute, the Wellcome Trust Sanger Institute,[312] and

participants, because it cannot be guaranteed. Ball et al., *supra* note 275, at 1; Conley et al., *supra* note 15, at 339–40; Lunshof et al., *supra* note 104, at 408.

306. Naveed et al., *supra* note 283, at 13. However, the majority of experts indicated that they would not trade accuracy of the data for privacy protection. *Id.*

307. *Id.* at 14.

308. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 80–81.

309. NAT'L CTR. FOR BIOTECHNOLOGY INFO., *supra* note 299.

310. *Id.*

311. Erin M. Ramos et al., Commentary, *A Mechanism for Controlled Access to GWAS Data: Experience of the GAIN Data Access Committee*, 92 AM. J. HUM. GENETICS 479, 484–85 (2013), http://www.sciencedirect.com/science/article/pii/S0002929713000803 [http://perma.cc/7M4W-F5EX]. However, there were eight technical data management incidents, which were violations of the terms of agreement. *Id.*

312. The Wellcome Trust, founded by Sir Henry Wellcome, is a charitable foundation based in London, England, that provides research funding for biomedical research, including genomic research. *About Us*, WELLCOME TR., http://www.wellcome.ac.uk/About-us/index.htm [http://perma.cc/39VL-TQR7] (last visited Jan. 17, 2015). The Wellcome Trust also established the Wellcome Trust Sanger Institute in Cambridge, England, one of the most prominent genomics research institutes in the world and a participant in the Human Genome Project (HGP). *About Us*, WELLCOME TR. SANGER INST., http://www.sanger.ac.uk/about/ [http://perma.cc/67DF-RHJ5] (last modified Apr. 19, 2013).

the Broad Institute[313] started restricting public access to pooled genomic data.[314]

These measures must also be accompanied by policies that require users to not distribute the data or use it for purposes other than research.[315] For example, to be approved for access to dbGaP, users must agree to several conditions, including: (1) storing the data in a secure location, (2) not distributing the data, (3) not attempting re-identification, and (4) reporting use and any adverse events.[316] The recently released NIH Data Sharing Policy sets forth policies and the responsibilities of the users who access the available data.[317] Although this specific policy only encompasses NIH-funded studies,[318] similar research governance policies should be in place for other control access databases and projects that are not NIH-funded.[319]

Another layer of access control can involve access auditing.[320] A "trust-but-verify" query audit allows monitoring of any adverse events to deter malicious users.[321] This approach requires a standardized interface for the analysis of genomic data.[322] Private healthcare providers have used

313. The Broad Institute, located in Cambridge, Massachusetts, is a leading biomedical research institute focusing on genomics and genomic medicine. *Areas of Focus*, BROAD INST., https://www.broadinstitute.org/what-broad/areas-focus/areas-focus [https://perma.cc/8XCT-QRSJ] (last visited July 25, 2015). It was founded by Eli and Edythe L. Broad and is affiliated with Harvard University and Massachusetts Institute of Technology. *History & Leadership*, BROAD INST., https://www.broadinstitute.org/history-leadership/history-leadership [https://perma.cc/Z893-2QFZ] (last visited July 25, 2015). The Broad Institute is also one of the leaders in the effort to make genomic data securely accessible. *Our Approach*, BROAD INST., https://www.broadinstitute.org/what-broad/our-approach/our-approach [https://perma.cc/FL4U-UVPN] (last visited July 25, 2015).

314. Misha Angrist, *Eyes Wide Open: The Personal Genome Project*, *Citizen Science and Veracity in Informed Consent*, 6 PERSP. MED. 691, 692 n.6 (2009); Naveed et al., *supra* note 283, at 26.

315. *See, e.g.*, NAT'L INSTS. OF HEALTH, *supra* note 205.

316. *dbGaP Approved User Code of Conduct*, NAT'L CTR. FOR BIOTECHNOLOGY INFO., https://dbgap.ncbi.nlm.nih.gov/aa/Code_of_Conduct.html [https://perma.cc/G4TS-XAGL] (last visited Apr. 14, 2015).

317. Final NIH Genomic Data Sharing Policy, 79 Fed. Reg. 51,345, 51,352 (Aug. 28, 2014). *See also Genomic Data User Code of Conduct*, *supra* note 205. For a discussion of the NIH Genomic Data Sharing Policy, see *supra* Part IV.E.

318. *Genomic Data User Code of Conduct*, *supra* note 205.

319. *See* Knoppers, *supra* note 25, at 1–4.

320. Erlich & Narayanan, *supra* note 22, at 417.

321. *Id.* "Trust-but-verify" trusts the users to make queries to the data without downloading it, records the queries, and verifies that the users did not execute any malicious queries. *Id.*

322. *Id.* Currently, genomic formats are not under uniform standards, which makes auditing efforts more difficult. *Id.*; RESPONSIBLE SHARING, *supra* note 117, at 19–20.

computational methods to audit for suspicious access to records.[323] Similar approaches can be adopted to audit user access in genomic databases with controlled access, and across different data formats and platforms if data formats are standardized.[324]

To implement data management and security measures more effectively, efforts to standardize data formats and platforms should be a priority for projects where participants and researchers are global.[325] Some leaders of the genomic research community are working to establish standard formats for genomic data and analysis tools, to facilitate effective and secure genomic data sharing.[326] Setting uniform data standards will allow for more secure storage and easier access control.[327] These standards will also provide a more consistent framework across platforms for data analysis for the researchers.[328] In addition, data standards and platforms can provide participants easier access to and more control over their data, and allow interaction with researchers and other communities through tools built on the platforms.[329]

---

323. Aziz A. Boxwala et al., *Using Statistical and Machine Learning to Help Institutions Detect Suspicious Access To Electronic Health Records*, 18 J. AM. MED. INFORMATICS ASS'N 498, 503 (2011). The method uses a statistical model and electronic health record access logs to identify rare suspicious access events, such as accessing more than 200 records in a day, accessing a coworker's record or accessing a neighbor's record. *Id.*

324. Erlich & Narayanan, *supra* note 22, at 417; RESPONSIBLE SHARING, *supra* note 117, at 19–20.

325. RESPONSIBLE SHARING, *supra* note 117, at 10.

326. Some of the scientific community is working to establish standardization of genomic data. *See*, *e.g.*, Erlich & Narayanan, *supra* note 22, at 417; RESPONSIBLE SHARING, *supra* note 117, at 10; *SAMtools*, SOURCEFORGE.NET, http://samtools.sourceforge.net/ [http://perma.cc/7TBZ-L9MW] (last modified Sept. 9, 2012); *Welcome to VCFtools*, VCFTOOLS, https://vcftools.github.io/index.html [https://perma.cc/4AUC-9QDR] (last visited Aug. 3, 2015).

327. RESPONSIBLE SHARING, *supra* note 117, at 9–10.

328. Some genomic databases are currently available on cloud services such as Amazon Web Services. *See supra* note 12. Standardization of data formats will facilitate access and analysis across different sequencing and analysis platforms. RESPONSIBLE SHARING, *supra* note 117, at 10.

329. RESPONSIBLE SHARING, *supra* note 117, at 10, 13. For example, standardized data and analysis tools will greatly facilitate development of tools for updated consent based on participant-centric initiatives. *See supra* Part V.B.1.

2. *Added Security Measures: Encryption and Anonymization*

Data custodians can use encryption, the process of making data unreadable except to those who possess the encryption key,[330] to mitigate privacy issues for genomic information.[331] One example of encryption techniques involves encrypting and storing genomic information on the cloud.[332] It only allows users to make computational queries and receive interpreted results from the cloud instead of downloading the raw data set to make those queries locally, essentially restricting access to the entire raw data.[333] A genomic database for biological relative search effectively employed this approach, allowing users to input their genomic information and identify relatives, but not providing access the individual relative's data.[334]

In addition to encryption techniques and basic de-identification by stripping the data of HIPAA identifiers, computer scientists and statisticians have developed other anonymization techniques to mitigate privacy breaches.[335] One method is k-anonymity, which ensures that no record is unique in that dataset.[336] Another technique is differential privacy, which

---

330. Encryption techniques, including security certificates, are widely used in various websites to ensure safe communication and data transfer between computers. *What Is Encryption?*, SURVEILLANCE SELF-DEFENSE, A PROJECT OF THE ELECTRONIC FRONTIER FOUNDATION, https://ssd.eff.org/en/module/what-encryption [https://perma.cc/A28Y-MV5Z] (last updated Nov. 3, 2014). Similar encryption techniques can be used for databases containing genomic information. Erlich & Narayanan, *supra* note 22, at 418–19. Mathematicians and cryptographers are developing new algorithms to apply to different types of computational data, such as genomic data. *Id.*

331. Erlich & Narayanan, *supra* note 22, at 418–19; Lucila Ohno-Machado et al., *Genomes in the Cloud: Balancing Privacy Rights and the Public Good*, 2013 AMIA SUMMITS ON TRANSLATIONAL SCIENCE PROCEEDINGS 128 (2013). One criticism of encryption methods is that they substantially limit what a researcher can do with the data and requires constant participation from the data guardian, dramatically increasing the cost of the use of the data. Ohm, *supra* note 9, at 1756. Furthermore, even the best systems for encryption and cloud computation are prone to bugs. *Id.* at 1757; Naveed et al., *supra* note 283. However, as increasingly more data becomes available online in cloud systems, a cryptographic solution can be an additional safeguard to these databases. Erlich & Narayanan, *supra* note 22, at 418–19.

332. Erlich & Narayanan, *supra* note 22, at 418–19.

333. *Id.* at 418 (providing a query of risk calculations for a specific risk factor as an example). Users cannot "possess" the data by downloading to their own computers. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 75.

334. *See* Dan He et al., *Identifying Genetic Relatives Without Compromising Privacy*, 24 GENOME RES. 664, 664–65 (2014).

335. Erlich & Narayanan, *supra* note 22, at 417–18; Ohm, *supra* note 9, at 1755–56.

336. Sweeney, *supra* note 77, at 557. K-anonymity takes out some of the identifiers in the data set so that at least k-1 records have the same combination of identifiers, preventing unique identification of one record. Erlich & Narayanan, *supra* note 22, at 417–18; Sweeney, *supra* note 77, at 557. However, if the k value is set too high, it may reduce the use of the data by leaving out critical information. Erlich & Narayanan, *supra* note 22, at 417–18.

adds controlled noise to the dataset before release, essentially making it unclear whether a record was in the original data set or in the noise set.[337] Although these methods alone cannot solve the issue of privacy breaches in public genomic data, they can provide added layers of security to make malicious breaches and re-identification more difficult.[338]

VI. CONCLUSION

Recent advancement in sequencing technologies and internet-based data sharing has borne the exciting fruits of public and participatory genomics.[339] The increase of publicly available genomic data greatly facilitated advances in human genetics, but at a cost of increased privacy concerns.[340] Recent re-identification studies have highlighted the privacy risks and a need for stronger privacy protection for those who contribute their genomic data to these projects.[341] There must be a balance between data availability for scientific advancement and participant privacy protection.[342] Complete privacy guarantees are difficult because unique personal information is inherent and permanent in the genomic data.[343] Due to the complexity of the problem, there is no silver bullet, but only a multimeric solution.[344]

The most critical component of the solution is a legislative ban of malicious re-identification and strengthened anti-discrimination protection.[345] Although some measures are already in place, most legislation excludes de-identified public information.[346] These legislative protections must

337. Noise refers to random data that does not contain meaningful information. *See* Kato Mivule, *Utilizing Noise Addition for Data Privacy*, *an Overview*, ARXIV:1309.3958 (2013), http://arxiv.org/ftp/arxiv/papers/1309/1309.3958.pdf [http://perma.cc/D3EQ-3ZLH]; Erlich & Narayanan, *supra* note 22, at 417–18. Differential privacy has been used in some GWAS studies, although the amount of noise that has to be added has so far made it impractical. Erlich & Narayanan, *supra* note 22, at 418–19.

338. Erlich & Narayanan, *supra* note 22, at 417–18; Ohm, *supra* note 9, at 1755–56.

339. *See supra* Part II.B.

340. *See supra* Part III.

341. *See supra* Part III.A.

342. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 101.

343. *See*, *e.g.*, Ball et al., *supra* note 275, at 1.

344. *See*, *e.g.*, PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 100–01.

345. *See supra* Part V.A.

346. Rothstein, *supra* note 87, at 8.

expand to consider re-identification risks, to protect the participants from privacy breaches, and to offer adequate remedies.[347]

To address the issues of consent, the Common Rule should be updated, factoring in the re-identification risks.[348] Additionally, implementing various access, education, and consent tools can encourage more informed participation.[349]

A complementary component of the multimeric solution is the adoption of standards by the research community.[350] Data standardization is critical not only in storing, managing and analyzing data across different platforms and projects, but is also important in facilitating patient participation, access management, implementation of protective measures and interactive consent, and can aid in providing standard privacy protection for international collaborative projects.[351] Reasonable mitigation of re-identification risks with this multimeric approach will encourage participation and data contribution in public and participatory genomics projects, and foster a more profound understanding of human genetics and evolution.[352]

---

347. PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 68–69.

348. *See, e.g.*, Human Subject Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators, 76 Fed. Reg. 44,512, 44,524 (July 26, 2011); *supra* Part V.B.1.

349. *See*, *e.g.*, Ball et al., *supra* note 275, at 1; Kaye et al., *supra* note 112, at 372; Lunshof et al., *supra* note 104, at 408; *Portable Consent* , *supra* note 285.

350. *See, e.g.*, RESPONSIBLE SHARING, *supra* note 117.

351. *See supra* Part V.C.1.

352. *See* PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *supra* note 1, at 52–54; McGuire et al., *supra* note 282, at 948.